



**FACULTAD DE INGENIERÍA, ARQUITECTURA Y
URBANISMO**

ESCUELA PROFESIONAL DE INGENIERÍA CIVIL

TESIS

**APLICACIÓN DE REDES NEURONALES
ARTIFICIALES PARA LA VALORIZACIÓN DE LOS
DEPARTAMENTOS HABITACIONALES DE LA
CIUDAD DE LIMA EN 2020**

**PARA OPTAR EL TÍTULO PROFESIONAL DE:
INGENIERO CIVIL**

Autor:

**Bach. Chávez Barturén Daniel Armando
(Orcid: 0000-0002-7968-0656)**

Asesor:

**Dr. Muñoz Pérez Sócrates Pedro
(Orcid: 0000-0003-3182-8735)**

Línea de Investigación:

Infraestructura, Tecnología y Medio Ambiente

**Pimentel – Perú
2022**

**APLICACIÓN DE REDES NEURONALES ARTIFICIALES
PARA LA VALORIZACIÓN DE DEPARTAMENTOS
HABITACIONALES DE LA CIUDAD DE LIMA EN 2020**

Aprobación del jurado

**Dr. Muñoz Pérez Sócrates Pedro
ASESOR**

**Mg. Marín Bardales Noé Humberto
PRESIDENTE**

**Mg. Villegas Granados Luis
Mariano
SECRETARIO**

**Mg. Idrogo Pérez César Antonio
VOCAL**

Dedicatoria

A Zeus, a mis abuelos, a mis padres, a mi
hermana, a todos los seres que tuve el placer de
conocer y sobre todo al universo que siempre me
ha dirigido a la verdadera libertad.

Agradecimientos

Agradezco al universo por la vida que me dio y que he sabido disfrutar, a mis abuelos por tanto amor entregado hacia mí, a mis padres, Lucy y David por la valentía de criarme con tanta comprensión.

A mi hermana, Karin. Nunca terminare de agradecerte por todo lo que has hecho por mí, por mis padres y por Zeus.

A mi gran amigo, Frank Valderrama, con quien desde segundo año de carrera no solo realizábamos las exposiciones más breves que he visto, además, llegábamos a debatir formulas establecidas debido a la variedad de biografía estudiada.

Debo agradecer a tantas personas que influenciaron en mi vida. Seres que me enseñaron a amar, odiar, reír y llorar. Estoy totalmente agradecidos con cada uno de ellos, pues han logrado forjarme mejor de lo que alguna vez soñé llegar a ser como persona.

Deje al final a mi mejor amigo Zeus, pues es el ser que más he logrado amar y el que más me ha podido amar. Siempre recordaré cada momento a tu lado. Espero el momento de volverte a ver algún día, sigue brillando mi ratón. Algún día espero llegar a tener la misma fuerza que tu tuviste en vida, te amo.

APLICACIÓN DE REDES NEURONALES ARTIFICIALES PARA LA VALORIZACIÓN DE LOS DEPARTAMENTOS HABITACIONALES DE LA CIUDAD DE LIMA EN 2020

APPLICATION OF ARTIFICIAL NEURAL NETWORKS FOR THE VALORIZATION OF THE HOUSING DEPARTMENTS OF THE CITY OF LIMA IN 2020

Daniel Armando Chávez Barturen¹

Resumen

La determinación del valor de departamentos habitacionales en la ciudad de Lima presenta muchas dificultades debido a los métodos de tasación que son costosos, requieren mucho tiempo y se caracterizan por su inflexibilidad, esto requiere metodologías actuales que simplifiquen sustancialmente el problema. Desde un punto de vista técnico, la aplicación de Redes Neuronales Artificiales podría ser una fuente objetiva para avaluar departamentos habitacionales, el tener valores más cercanos a las condiciones de mercado, aumentará la seguridad en la toma de decisiones en diferentes escenarios. Para la obtención de la base de datos de trabajo se recurre a los datos disgregados de ventas de departamentos en Lima, brindados por el Banco Central de Reserva del Perú. Esta base de datos fue purificada y procesada para el desarrollo del modelo de Redes Neuronales Artificiales con la optimización bayesiana. A pesar de llegar a tener un modelo con resultados satisfactorios se optó por mejorar dicho modelo aumentando el número de epochs y neuronales en las capas ocultas, obteniendo el mejor modelo como producto final.

Palabras clave: Bienes Raíces, Optimización Bayesiana, Perceptrón Multicapa, Redes Neuronales Artificiales

1 Adscrita a la Escuela Académica Profesional de Ingeniería Civil, Pregrado, Universidad Señor de Sipán, Pimentel, Perú, email: cbarturendanie@crece.uss.edu.pe

ORCID: <https://orcid.org/0000-0002-7968-0656>

Abstract

The determination of the value of housing apartments in the city of Lima presents many difficulties due to the valuation methods that are expensive, require a lot of time and are characterized by their inflexibility, this requires current methodologies that substantially simplify the problem. From a technical point of view, the application of Artificial Neural Networks could be an objective source to assess housing apartments, having values closer to market conditions will increase the security in decision-making in different scenarios. To obtain the work database, the disaggregated data of apartment sales in Lima, provided by the Central Reserve Bank of Peru, is used. This database was purified and processed for the development of the Artificial Neural Networks model with Bayesian optimization. Despite having a model with satisfactory results, it was decided to improve said model by increasing the number of epochs and neurons in the hidden layers, obtaining the best model as a final product.

Keywords: Artificial Neural Networks, Bayesian Optimization, Multilayer Perceptron, Real Estate

ÍNDICE

I. INTRODUCCIÓN	14
1.1. Realidad problemática	14
1.1.1.Nivel internacional.....	14
1.1.2.Nivel nacional y local.....	16
1.2. Antecedentes de estudio.....	18
1.2.1.Nivel internacional.....	18
1.2.2.A nivel nacional	23
1.2.3.A nivel local	23
1.3. Teorías relacionadas al tema.....	23
1.3.1.Valuación de departamentos habitacionales.....	23
1.3.2.Inteligencia Artificial.....	25
1.3.3.Redes Neuronales Artificiales (RNA)	29
1.3.4.Redes neuronales artificiales: Componentes	35
1.3.5.Redes neuronales artificiales: Desarrollo de modelos.....	41
1.3.6.Redes neuronales artificiales: Optimización de modelos	44
1.3.7.Redes neuronales artificiales: Arquitectura	44
1.4. Formulación del problema.....	48
1.5. Justificación e importancia del estudio.....	48
1.6. Hipótesis	49
1.7. Objetivos	49
1.7.1.Objetivo general	49
1.7.2.Objetivos específicos	49
II. MATERIAL Y MÉTODO	51

2.1. Tipo y diseño de investigación	51
2.1.1.Tipo de investigación.....	51
2.1.2.Diseño de investigación	51
2.2. Población y muestra.....	52
2.2.1.Población	52
2.2.2.Muestra	52
2.3. Variables y operacionalización.....	52
2.4. Técnicas e instrumentos de recolección de datos, validez y confiabilidad	56
2.4.1.Técnicas de recolección de datos	56
2.4.2.Instrumentos de recolección de datos.....	56
2.4.3.Procedimientos para la recolección de datos	56
2.5. Procedimiento de análisis de datos.....	59
2.5.1.Entradas y salidas del modelo	59
2.5.2.División del conjunto de datos.....	59
2.5.3.Determinación de la arquitectura del modelo	59
2.6. Criterios éticos	60
2.7. Criterios de rigor científico.....	60
III. RESULTADOS	62
3.1. Resultados en tablas y figuras	62
3.1.1.Exploración del conjunto de datos	62
3.1.2.Entradas y salidas del modelo	65
3.1.3.División del conjunto de datos.....	69
3.1.4.Desarrollo del modelo de Red Neuronal Artificial para evaluar departamentos habitacionales en la ciudad de Lima.	72

3.1.5.Mejorando el modelo J.A.R.V.I.S.	76
3.1.6.Resumen de los modelos de Redes Neuronales Artificiales desarrollados.....	80
3.2. Discusión de resultados	85
3.2.1.Aplicación de modelos de redes neuronales artificiales para evaluar departamentos habitacionales en la ciudad de Lima.....	85
3.2.2.Avaluar departamentos habitacionales en la ciudad de Lima y los IVDH utilizadas en la red neuronal artificial.....	86
3.2.3.El número de neuronas y la precisión	86
3.2.4.El tiempo de entrenamiento afecta en la precisión.....	87
IV. CONCLUSIONES.....	89
4.1. Conclusiones.....	89
4.2. Recomendaciones	90
ANEXOS	100
Anexo 1. Formato de la ficha de registro de datos brindados por el banco central de reserva del Perú	101
Anexo 2. Formato de la ficha de registro de datos utilizadas en las redes neuronales artificiales.....	102
Anexo 3. Formato de la ficha de registro de datos para los hiperparámetros asociados al modelo e hiperparámetros asociados al descenso del gradiente	103
Anexo 4. Tabla de registro de datos para el rendimiento de los modelos de redes neuronales artificiales	104
Anexo 5. Tabla de los 10 mejores modelos procesados.....	105
Anexo 6. Características del conjunto de datos	106

Anexo 7.	Rendimiento de los modelos de Redes Neuronales	
	Artificiales mejorados	108
Anexo 8.	Gráfica de comparación entre los valores predicho y los	
	valores reales.....	109

ÍNDICE DE FIGURAS

Figura 1 Estructura de una red neuronal biológica.....	30
Figura 2 Estructura de una red neuronal artificial.....	31
Figura 3 Perceptrón sin función de activación.....	32
Figura 4 Perceptrón con función de activación.....	33
Figura 5 Arquitectura general de un MLP.....	34
Figura 6 Función de activación Lineal.....	37
Figura 7 Función de activación Sigmoide.....	38
Figura 8 Función de activación ReLU.....	39
Figura 9 Función de activación Tangente hiperbólica.....	40
Figura 10 Validación cruzada: División del conjunto de entrenamiento en 3 particiones.....	43
Figura 11 Distribución del conjunto de datos según el año de venta del departamento.....	63
Figura 12 Distribución del conjunto de datos según el distrito del departamento.....	64
Figura 13 Diagrama de dispersión para el precio en dólares del inmueble y las variables de entrada.....	68
Figura 14 Error Absoluto Medio (MAE) en las pérdidas de entrenamiento y validación.....	75
Figura 15 Error Cuadrático Medio (MSE) en las pérdidas de entrenamiento y validación.....	75
Figura 16 Rendimiento de los modelos de Redes Neuronales Artificiales desarrollados para mejorar el J.A.R.V.I.S.....	77
Figura 17 Rendimientos de los modelos desarrollados para mejorar el J.A.R.V.I.S según el número de neuronas ocultas.....	78

Figura 18 Gráfica de la correlación de las predicciones de los modelos propuestos.....	79
---	----

INDICE DE TABLA

Tabla 1 Operacionalización de la variable independiente.....	53
Tabla 2 Operacionalización de la variable dependiente.....	55
Tabla 3 Entradas y salida del Modelo de Red Neuronal Artificial propuesto.....	66
Tabla 4 Estadística del conjunto de datos utilizados como entradas de la Red Neuronal Artificial.	69
Tabla 5 Estadística del conjunto de datos utilizados como salida de la Red Neuronal Artificial.	70
Tabla 6 Prueba U de Mann-Whitney para la igualdad de medias entre el conjunto de entrenamiento y conjunto de prueba para las variables de entrada	71
Tabla 7 Prueba U de Mann-Whitney para la igualdad de medias entre el conjunto de entrenamiento y conjunto de prueba para las variables de salida	72
Tabla 8 Mejores 10 Redes Neuronales Artificiales para evaluar departamentos habitacionales en la ciudad de Lima.....	73
Tabla 9 Estructura y rendimiento de los modelos de redes neuronales artificiales desarrollados.....	82
Tabla 10 Arquitectura e Hiperparámetros de los mejores modelos aplicando optimización bayesiana.....	83

CAPITULO I:
INTRODUCCION

I. INTRODUCCIÓN

1.1. Realidad problemática

1.1.1. Nivel internacional

La valuación del sector inmobiliario es algo muy relevante en la vida de un ciudadano. Estas valuaciones son importantes para diversos fines, por ella la Norma Internacional de Valuaciones surge como una guía de carácter transnacional, con el objetivo de generar confianza y credibilidad pública, dando a las valoradoras normas actualizadas. (International Valuation Standards Council, 2020)

La teoría de valuación de bienes raíces se basa en una hipótesis, donde una muestra aleatoria de valuación de una vivienda en una población normal coincida con el precio real. Esto rara vez se refleja debido a dos factores: el primer problema consiste en que los avalúos rara vez analizan las muestras con métodos basados en la lógica aleatoria y el segundo problema es debido a la atipicidad de las muestras, esto origina una invalidación sistemática de la distribución normal de la población (Antoniucci & Marella, 2017).

Dentro del mayor problema para realizar investigaciones en la valoración de bienes raíces es debido a la poca transparencia y facilidad con respecto a la información del mercado, pues siempre se muestran condiciones y aspectos que obligan a los investigadores a trabajar con datos pequeños o simples (Del Giudice et al., 2017).

Se investigó que en diferentes escenarios la obtención del valor de mercado de inmuebles urbanos es fundamental en la decisión de la vida de los ciudadanos. Es conocido que el gobierno se basa en este valor como base para cobrar impuestos, sirve como base para diferentes trámites ante entidades, tanto públicas como privadas. (Preciado Carrillo, 2015)

En busca de una manera sencilla de determinar el precio total de la vivienda en Córdoba, se propuso el uso de la metodología de Redes Neuronales Artificiales (RNA), ya que este posee una mayor capacidad para reconocer el proceso de información de los precios. Se indicó que este método se ha dado por un proceso de error y prueba, es este método el que indica superioridad con respecto a los métodos clásicos, pues la Redes Neuronales Artificiales no cuenta comuna rigidez en el análisis de datos, ya que existe un elevado grado de multicolinealidad en las variables explicativas del precio del bien. Este método ofrece la posibilidad de introducir las variables originales previa estandarización como inputs, a diferencia de los métodos clásicos que necesitan generar índices lo que origina una gran pérdida de información y problemas a la hora de arrojar los resultados. (Ceular et al., 2008)

Se han realizado muchos estudios comparando los sistemas de Inteligencia Artificial (IA) con los métodos tradicionales de avalúo. La mayoría de estas investigaciones calculan el porcentaje de error de un sistema de IA y de otro de regresión múltiple, añadiendo a ellas información del mercado, donde les brinda su precio de venta. Las metodologías de IA arrojan en las pruebas unos errores medio que se sitúan entre el 5 y el 10%, mientras que los de Regresión Múltiple se sitúan más entre el 10 y el 15%. De igual manera, la investigación muestra una mejor precisión del sistema de inteligencia artificial. (Mora-Esperanza, 2004)

A pesar de los colapsos del mercado inmobiliario han sido raros en Colombia, según la experiencia de la región a finales de la década de 1990 es un recordatorio útil de los efectos sistémicos que incluso un pequeño sector hipotecario puede tener en la economía. Además, como se vio en la reciente crisis de vivienda en los Estados Unidos, los problemas en un mercado pequeño pueden volverse sistémicos, especialmente en nuevos mercados con brechas de datos significativas. Los precios y rentas de la propiedad de vivienda en algunas áreas metropolitanas de la región generalmente no aparecen fuera de línea cuando se miden en relación con el ingreso per cápita del país, así como con los precios de alquiler, aunque

son aproximándose a niveles espumosos en algunos países (Colombia, Brasil y Perú). (Cubeddu et al., 2012)

1.1.2. Nivel nacional y local

La demanda potencial para la adquisición de viviendas asciende a 163 mil hogares en Lima, de las cuales el 16,9% tiene un conocimiento de crédito hipotecario; y un 72% tiene conocimiento de instituciones financieras. Finalmente, el 26.7 % de los encuestados optan por habitar en un departamento en edificio. (Fondo Mivivienda, 2018). En cuanto al desarrollo del Perú, para el segundo trimestre de 2020, el sector construcción disminuyó en un 67,2% como resultado de la disminución de obras públicas y privadas a causa de la pandemia. (Instituto Nacional de Estadística e Informática, 2020).

La inversión de proyectos inmobiliarios representa que el Perú y en especial Lima ha tenido un buen crecimiento, con solvencia económica y oportunidades laborales. Se ha observado que muchos usuarios que demandan la compra de departamentos y viviendas han utilizado plataformas digitales a través de un dispositivo móvil o PC. (Gómez, 2019).

Se realizó un estudio sobre la comparación de los factores determinantes en la tasación de inmuebles, con la finalidad de comparar los distintos precios de bienes inmuebles obtenidos mediante los distintos tipos de tasaciones solicitadas, tomando como muestra informes de tasaciones de diferentes peritos a nivel de todo el Perú. En Perú a pesar de que existen normativas que regulan los métodos o criterios técnicos de tasación, el país no se encuentra lo suficientemente desarrollado en el mercado inmobiliario y carece de información de datos confiables que faciliten la comparación óptima de bienes inmuebles. (Olaechea-cunza, 2019).

Según un estudio realizado en la ciudad de Lima el precio por metro cuadrado de los departamentos habitacionales en las áreas urbanas de Lima se ha incrementado a niveles sin precedentes, donde menciona que el precio por metro cuadrado en Lima surgió de US \$521 dólares de 2006 a US \$1838 dólares a fines de 2013. (Orrego, 2014)

Se indica que se debe tener cautela en la inversión de un inmueble, ya que esta es una gran decisión en la vida de una persona, puesto que ellos significaron el riesgo de los ahorros de toda su vida. En Lima, la demanda de departamentos habitacionales aumentó cerca de 1.2% mes al mes en los últimos 5 años. (Gestión Noticias, 2018)

Según un estudio realizado por una empresa inmobiliaria latinoamericana, nos dice que vivir en los distritos de Lima Top no significa que todos los inmuebles se cotizan igual. Hay una diferencia entre Barranco, Miraflores y San Isidro, por citar algunos, al igual sucede en Lince, Pueblo Libre o Jesús María. El 57% de los departamentos en venta son de 3 habitaciones, seguidos por el 25% de 2 habitaciones, el 11% de 4 habitaciones o más y el 8% de 1 habitación. El 57% de inmuebles en venta son de departamentos, seguido por el 19% de casa, el 12% de otro, el 9% y el 3% de local. La web-site Properati nos indica que existe una gran dificultad por la población para adquirir un departamento por la falta de confianza en los precios de los inmuebles por no tener en cuenta un respaldo de cotización, es esta web-site la única que cuenta con una aplicación para realizar una cotización de un predio por sus características. (Properati, 2019)

El impuesto predial está regulado por la Ley De Tributación Municipal, quien a su vez basa su impuesto de acuerdo al valor total del predio y los valores unitarios oficiales de la edificación, todas ellas regidas de acuerdo al Consejo Nacional de Tasaciones. El problema principal con este impuesto a nivel nacional es debido a que no existe una actualización catastral de ciudades que hayan crecido de manera descontrolada como se aprecia en la mayoría del país. Esto genera que el impuesto final a pagar sea un cálculo erróneo(Gobierno del Perú, 2017).

1.2. Antecedentes de estudio

1.2.1. Nivel internacional

Chaphalkar & Sandbhor (2019) realizó un estudio titulado: «Efecto de la muestra de entrenamiento y las características de la red en la predicción del valor de la propiedad real basada en redes neuronales», teniendo como problema el efecto de la muestra de entrenamiento y las características de la red en la predicción del valor de la propiedad real basada en la red neuronal, tomando como muestra más de 3000 instancias de venta que se registraron y tomaron como datos de entrada en la ciudad de Pune, India. Los resultados arrojaron un porcentaje variable de la tasa de predicción. Se comparan los resultados y arrojan que se obtiene una red robusta que aseguran un gran porcentaje variable en la tasa de predicción. El estudio concluye en que el proceso garantiza una mejor generalización que resulta en una mejor previsibilidad a la hora de predecir el valor de una propiedad en la ciudad de Pune.

Preciado Carrillo (2019) en su estudio titulado «URBAN REAL ESTATE VALUATION: Comparative multivariate regression model versus artificial neural networks for the city of Morelia, Michoacán, Mexico », realizó un estudio sobre la valoración de inmuebles urbanos, haciendo una comparación entre un modelo económico y una red neuronal del tipo multicapa back propagation con la finalidad de estimar y predecir el precio de casas sobre las diferentes variables del mercado inmobiliario, tomando como muestra 688 inmuebles de venta consumada y 297 inmuebles con precio de oferta con un factor de negociación del 5% en la ciudad de Morelia, México. Los resultados arrojaron lo siguiente: El método de red neuronal con 16 unidades en la capa oculta, supera a los métodos tradicionales y el error absoluto mostrado entre los métodos propuestos, arrojan a la Red Neuronal Artificial cómo el que menor error arrojó. El estudio llegó a la conclusión, que según la data brindan en el caso de Morelia, las redes neuronales explican mejor las variables explicativas del

precio de los inmuebles que los modelos de regresión tradicionales, planteándose como una alternativa práctica para la adquisición de inmuebles.

Epley (2017) en su estudio titulado «La necesidad de referir los modelos de valoración automática a la valoración», teniendo como problema la aplicación de modelos automáticos sin referencias en el campo de valoración, tomando como muestra datos de valorizaciones recopilada de diversas investigaciones con la finalidad de lograr un modelo de Red Neuronal Artificial del tipo perceptrón multicapa. En su investigación se desarrollaron modelos de valorización de manera automática, además se encontró la existencia de una desconexión entre las leyes estándares de tasación y la comunidad financiera para la valorización de inmuebles, por lo que es esencial utilizar varios criterios de consenso para la selección de las variables a utilizar en la Red Neuronal Artificial.

Horvath et al. (2021) en su estudio titulado «Derivación de tamaños de muestra adecuados para el modelado basado en Redes Neuronales Artificiales de tareas de valoración de bienes raíces mediante análisis de complejidad», teniendo como problema la generación de una base de datos adecuada para el modelado basado en Red Neuronal Artificial de tasaciones inmobiliarias en Alemania, tomando como muestra datos del comité de expertos de tasación del estado federal de Baja Sajonia. Los resultados arrojan lo siguiente: Agregar submercados inmobiliarios (base de datos de otras localidades) para generar una base de datos suficiente para el modelado de Red Neuronal Artificial es necesario como mínimo un lote de 5000 muestras, pero 9000 para llegar a una estimación confiable. Esto se aplica a la estimación de parámetros de Red Neuronal Artificial, que generalmente presenta modelos menos complejos, menores errores de prueba y menor varianza de estimación.

Kamire et al. (2021) en su estudio titulado «Real Property Value Prediction Capability Using Fuzzy Logic and ANFIS», teniendo como problema la comparación de los resultados de la predicción entre la lógica difusa y ANFIS (sistema de inferencia neuro difuso adaptativo), tomando

como muestra datos de la ciudad de Pune, India. Los resultados arrojan lo siguiente: la lógica difusa es eficaz cuando existe una relación entre las variables, pero el ANFIS predijo casos desconocidos con bastante eficacia. Ha sobreestimado algunos de los valores, pero, aun así, ha tratado de desarrollar alguna relación para el nuevo caso y ha predicho los resultados. El estudio llegó a la conclusión, que ANFIS es bastante eficaz para predecir conjuntos de datos no lineales y cuantitativos.

Rey Carmona & Núñez Tabales (2017), en su estudio titulado «Alternativas para la valoración de inmuebles urbanos», teniendo como objetivo recopilar todas las alternativas existentes para valoración de inmuebles urbanos. Los resultados arrojan lo siguiente: las redes neuronales artificiales y la metodología de Precios Hedónicos destacan sobre las otras opciones. La conclusión del estudio es, muchos estudios apoyan la metodología de redes neuronales artificiales, pero en casos especiales no llegan a ser mejores, por lo que se recomienda aplicar cada caso concreto para llegar a una solución más idónea.

Štubňová et al. (2020), en su estudio titulado «Estimación del precio del mercado inmobiliario residencial: comparación de redes neuronales artificiales y modelo de precios hedónicos», teniendo como objetivo realizar una comparación entre los modelos tradicionales (modelos hedónicos) y los modelos actualizados redes neuronales artificiales. Utilizando como muestra datos de propiedades residenciales vendidas en el mercado inmobiliario de la ciudad de Nitra, República Eslovaca. La conclusión de la investigación demuestra que las redes neuronales artificiales supera en precisión la estimación del precio a comparación del modelo hedónico.

Del Giudice et al. (2017), en su estudio titulado «Tasaciones de bienes raíces con enfoque bayesiano y método de Monte Carlo híbrido de cadena de Markov: una aplicación a un área urbana central de Nápoles», teniendo como objetivo es demostrar que el Método Monte Carlo Híbrido en Cadena de Markov obtiene mejores resultados que la metodología Redes Neuronales Artificiales, análisis de regresión múltiple tradicional y el método semi paramétrico de spline penalizado. Todos los métodos utilizaron la

misma muestra de Nápoles. Llegando a la conclusión que tan solo la metodología de MCHMCM y redes neuronales artificiales han logrado obtener un error porcentual promedio en su predicción de 6.61% y 7.33% respectivamente.

Zhou et al.(2018) en su estudio titulado «Redes neuronales artificiales y tasación masiva de inmuebles», teniendo como objetivo enfocarse en el desarrollo teórico de redes neuronales artificiales y la valoración masiva de bienes raíces. Utilizando en su revisión de literatura la investigación teórica de redes neuronales artificiales aplicadas en valoración de bienes raíces, la evolución de las redes neuronales artificiales en modelos y mejora de algoritmos informáticos, y en las investigaciones de redes neuronales artificiales aplicadas en los distintos campos de valoración de bienes raíces. Llegando a la conclusión que la metodología de redes neuronales artificiales para tasaciones masivas demuestra asemejarse en precisión y tiempo.

Purohit et al. (2021) en su estudio titulado «Estudio de Eficacia de Modelos de Pronóstico para Bienes Raíces Residenciales», teniendo como objetivo discutir la eficacia y popularidad de los modelos de pronóstico inmobiliario. Utilizando revisión de literatura, quedando claro que se han realizado muchas investigaciones sobre factores microeconómicos que utilizan variables como vecindario, servicios y factores locales para la investigación. Llegando a utilizar 16 de regresión múltiple, 4 de redes neuronales artificiales, 3 de historia tiempo y 4 de regresión de cresta. Llegando a la conclusión que los modelos de redes neuronales artificiales llegaron a tener un error más bajo de 8.92%, el más bajo a comparación de los otros métodos.

Çelik (2021) en su estudio titulado «Métodos tradicionales en la valoración de inmuebles», teniendo como objetivo realizar una comparación de valuación entre la zona residencial y la zona urbana en la provincia de Samsun, el distrito de Büyükkolpınar y la provincia de Gümüşhane, Merkez, distrito de Bağlarbaşı. El resultado de las investigaciones determinó un valor de coeficiente de correlación en las

zonas residenciales de 0.9231 y en las zonas industriales de 0.9156. determinando que las Redes neuronales artificiales superaron de gran manera a los métodos tradicionales que no sirvieron para una valoración en masa.

Sridhar & Sathyanathan (2022) en su estudio titulado «Modelado de precios de casas residenciales usando análisis de regresión», teniendo como objetivo utilizar el análisis factorial, el análisis de conexiones y el análisis de regresión lineal para modelar los precios de una vivienda residencial. El estudio se realizó en el barrio de Chengalpattu, determinando así los factores que explican los precios de las viviendas. Los resultados arrojaron que el error obtenido por el análisis factorial fue superior al 15%, el análisis de regresión lineal llegó hasta un 12% de error y la metodología de redes neuronales artificiales llegó a tener errores abajo del 10%.

Małgorzata et al. (2022) en su estudio titulado «La aceptación internacional del uso de herramientas modernas automatizadas es imprescindible para el desarrollo sostenible del mercado inmobiliario», teniendo como objetivo sistematizar los términos y las metodologías utilizadas en el ámbito de valoración inmobiliaria y análisis de mercado. La investigación se llevó a cabo utilizando el método de análisis crítico del estado de arte actual de la literatura, normas de valoración y opiniones de expertos en el mercado inmobiliario. Los resultados del estudio es que las entidades deben cambiar la percepción de las diferentes soluciones automatizadas que operan, pues estos métodos se basan en gran proporción a los métodos tradicionales que muchos están acostumbrados.

Lee & Jae (2021) en su estudio titulado «Predicción del índice de precios de la vivienda utilizando una red neuronal artificial», teniendo como objetivo aplicar redes neuronales artificiales para valorizar departamentos en Seúl, haciendo una comparación de los micro factores, los macro factores y la combinación de ambos. Teniendo como resultado una predicción del 87.5% en 2 años y en el último caso una predicción del 95.8%.

1.2.2. A nivel nacional

No existen investigaciones que apunten a los objetivos planteados en la presente investigación.

1.2.3. A nivel local

No existen investigaciones que apunten a los objetivos planteados en la presente investigación.

1.3. Teorías relacionadas al tema

1.3.1. Valuación de departamentos habitacionales

El consejo de las Normas Internacionales de Valuación define la Valuación como el acto o proceso de determinar un estimado de valor de un activo o pasivo. Nos dice que al hablar de la palabra “valor” es referirse a la conclusión del valuador sobre la cuantía estimada de un activo o pasivo. En otras palabras, el valor depende de la cuantía que el valuador le ponga a determinado objeto con el uso de tablas de valor establecidas. (International Valuation Standards Council, 2020)

1.3.1.1. Valor de Mercado o Venal de un inmueble (VM)

El valor de mercado es el precio justo con el cual se subasta un inmueble en un mercado razonable, sin tener a consideración el valor agregado de los trámites de adquisición, la comisión de publicidad o la comisión de la inmobiliaria. (Tinsa, 2016).

1.3.1.2. Bienes inmuebles

El consejo de las Normas Internacionales de Valuación reconoce cuatro tipos de bienes, pero vamos a tomar en cuenta tan sólo los bienes inmuebles que son terrenos y cualquier otra cosa que pertenezcan a esto,

ya sean aspectos naturales o agregados puestos por el ser humano. También considera a todo tipo de instalaciones puestos en el predio, estos inmuebles son denominados inmuebles urbanos. (International Valuation Standards Council, 2020).

1.3.1.3. Departamentos Habitacionales

Al hablar de departamentos habitacionales, hablamos sobre una parte de un edificio multifamiliar. Que, según El Reglamento Nacional de Edificaciones, nos dice que se considera a una parte de un edificio multifamiliar, donde se encuentren dos o más viviendas independientes con un terreno de propiedad común. (Gobierno del Perú, 2020).

1.3.1.4. Normativas y reglamento internacionales y Nacionales de Valuación

Estas normas van a servir como pautas para los atributos o variables de la Red Neuronal Artificial a tomar en cuenta para realizar la Valoración de los Departamentos Habitacionales.

- **Normas internacionales de Valuación (IVS):** (International Valuation Standards Council, 2020)
- **Uniform Standards of Professional Appraisal Practice (USPAP):** (The Appraisal Foundation, 2019)
- **Reglamento Nacional Tasaciones de Perú:** (Gobierno del Perú, 2016)
- **Reglamento del Centro de Peritaje Del Colegio de Ingenieros Del Perú:** (Colegio De Ingenieros Del Perú, 2018)

1.3.1.5. Indicadores de Valoración de Departamentos Habitacionales (IVDH)

Los IVDH son herramientas de evaluación según la tipología de vivienda, ya que según esta se deberá realizar estudios específicos para cada una de ellas, esto va a depender del tipo, localización y entorno inmediato del predio. (Abadi Abbo et al., 2009).

Para la elección de los IVDH se ha considerado la información recopilada del Banco Central de Reserva del Perú (BCRP), para ser precisos en su base de datos disgregada de “indicadores de precios de alquiler y ventas de departamentos”.

1.3.2. Inteligencia Artificial

La Inteligencia Artificial (IA) es una disciplina cuyo objetivo trata de simular el sistema intelectual humano en un sistema artificial. La IA se utiliza principalmente para identificación de patrones o tratamiento de datos, debido a que la inteligencia humana que tratan de imitar llega a comprender percepciones sensoriales, dentro de ellas la visión, audición y reconocimiento de patrones. (Benitez, 2014).

Según Goodfellow et al. (2016) la IA ayuda a resolver problemas que eran intelectualmente complejos para los humanos, pero simples para las máquinas con este sistema. Para que las máquinas puedan resolver estos problemas es necesario que estas obtengan su propio conocimiento, para que ellas puedan aprender patrones.

El aprendizaje automático es una rama de la IA cuyo objetivo es desarrollar técnicas que logren el aprendizaje de las computadoras mediante algoritmos de aprendizaje automático, los cuales son capaces de aprender a partir de una base de datos, en otras palabras, el programa informático mejora su rendimiento en una tarea asignada conforme mejora con la experiencia. Entre los principales trabajos que trata de resolver el aprendizaje automático es la clasificación, regresión y agrupamiento. (Godfrey & Gashler, 2015)

Cuando se descubre la tarea que se quiere resolver, se debe evaluar el rendimiento del algoritmo de aprendizaje automático, los cuales pueden clasificarse según la capacidad de experiencia que se les permite tener en el proceso de aprendizaje: aprendizaje supervisado y aprendizaje no supervisado.

El aprendizaje supervisado consiste en experimentar la base de datos con características establecidas y su conjunto de salida se compara con el conjunto de salidas que debería tener. Por otra parte, el aprendizaje no supervisado, es el algoritmo quien aprende de la base de datos, descubre y agrupo según sus patrones y características.

1.3.2.1. Capacidad del modelo

Según Goodfellow et al. (2016), el gran desafío que tiene el aprendizaje automático es que el algoritmo de aprendizaje no tan solo obtenga un buen rendimiento con los datos ya utilizados, sino también con datos no conocidos, debido a esto se trata de fragmentar el base de datos en una data para la fase de entrenamiento y otra para la fase de prueba. A esta habilidad de obtener un buen rendimiento con una data nunca antes vista se denomina generalización.

La finalidad del aprendizaje automático es reducir el error de generalización, para esto se trata de ingresar una nueva entrada con la cual comparar los datos predichos por el modelo y los datos reales (Chollet, 2018).

Para Goodfellow *et al.* (2016), el rendimiento de un modelo de aprendizaje automático está dado por su capacidad de:

- a. Reducir el error en la fase de entrenamiento.
- b. Reducir la diferencia entre el error en la fase de entrenamiento y el error en la fase de prueba (error de generalización).

El entrenamiento, la optimización y la generalización están relacionados; mientras menor sea el error de entrenamiento se estima que el error generalizado también va a decrecer (Chollet, 2018). Sin embargo, uno de los principales desafíos de realizar modelos de aprendizaje automático en la fase de entrenamiento es el *underfitting* y el *overfitting* (Goodfellow et al., 2016).

El *overfitting* aparece debido a la gran diferencia de error de entrenamiento y el error de generalización, esto ocurre debido a que el modelo es muy complejo para la cantidad de información ingresada, generando que el error de generalización no disminuya y comience a aumentar (Chollet, 2018).

El *underfitting* aparece debido a que el modelo no tiene la capacidad de reducir el error en la fase de entrenamiento, esto ocurre porque el modelo es muy simple, por lo que no es capaz de absorber los aspectos y variabilidad de los datos. Pero no es recomendable hacerlo muy complejo porque esto originaría que apareciese el *overfitting*. (Müller & Guido, 2020).

Para lograr de reducción y aumento del modelo es necesario controlar la aparición del *overfitting* o *underfitting* controlando su capacidad. Para el caso del *underfitting* ese necesario aumentar el número de neuronas ocultas o se aumenta el número de capas ocultas (Chollet, 2018). En el caso del *overfitting* es necesario disminuir el número de parámetros de aprendizaje como lo son: el número de neuronas en cada capa oculta y el número de capas ocultas. Estos factores como el número de capas ocultas y el número de neuronas en cada capa ocultas nos determinan la capacidad del modelo, pero se debe tener mucho cuidado debido a que estos factores pueden memorizar propiedades del conjunto de entrenamiento que no le sirven para el conjunto de prueba. (Goodfellow et al., 2016).

No existen parámetros ni fórmulas que nos determinen el número de capas y neuronas para obtener el mejor modelo, por lo que es necesario realizar iteraciones que se deben hacer para el conjunto de entrenamiento. Este procedimiento no se debe repetir con el conjunto de prueba, ya que estos datos nos van a servir para evaluar el rendimiento final del modelo. (Chollet, 2018)

Según Goodfellow *et al.* (2016), nos demuestra que existen diferentes estudios donde ingresar una mayor cantidad de datos para entrenar el modelo ayudan a disminuir el error de generalización. Esto

quiere decir que mientras más datos tengamos, se obtendrán modelos más complejos y sin sobre ajustarse. Todo esto es debido a la relación entre la complejidad del modelo y la variedad de datos.

1.3.2.2. Hiperparámetros

Los controladores del comportamiento de una red neuronal artificial son los hiperparámetros, estos se deben establecer antes de la optimización de la red. El problema más grande de los hiperparámetros es llamado Underfitting, que es una configuración difícil de optimizar, el cual inhibe la capacidad suficiente para generalizar. Debido a este problema se recomienda tener un lote de datos para el entrenamiento del problema, otro lote de datos para el conjunto de validación (utilizado para ajustar los hiperparámetros) y por último un lote que sólo se debe usar para el conjunto de prueba para finalizar el entrenamiento. (Goodfellow et al., 2016; Vijayashanthar et al., 2018)

1.3.2.3. Descenso del gradiente

Los modelos de aprendizaje automática necesitan un proceso de mejoramiento en su función de pérdida. Esta función consiste en optimizar (minimizar o maximizar) la función objetiva. (Goodfellow et al., 2016).

Para identificar el éxito de una red neuronal artificial es necesario utilizar la función de pérdida, para ello utilizan los algoritmos de aprendizaje que determinan cómo la red neuronal artificial actualizara los pesos en función de la pérdida. (Chollet, 2018). El error de aprendizaje no es más que la diferencia entre los resultados esperados y el resultado que la red neuronal artificial arroja. Para ello la red neuronal artificial modifica los pesos en función de su algoritmo de aprendizaje para que la red neuronal artificial aprenda cual es la función de salida que se desea (Casas Roma et al., 2019).

Se denomina descenso del gradiente al algoritmo de optimización iterativo de primer orden que nos permite encontrar el mínimo de una

función utilizando la derivada. Se trata del algoritmo más común para optimizar la función de pérdida de una red neuronal artificial. Aquellos algoritmos que utilizan la gradiente se denominan algoritmos de primer orden (Goodfellow et al., 2016).

El vector gradiente indica la dirección donde crece la función de pérdida, en otras palabras, el sentido donde crece el error. No obstante, lo que se busca es minimizar dicho error, por lo que mueve en dirección opuesta al gradiente. (Goodfellow et al., 2016).

La finalidad del descenso del gradiente es encontrar el valor mínimo de una función, que se denomina mínimo global, que en muchos casos es confundido con algún punto local que es un valor próximo al deseado. Se debe evitar llegar a uno de estos puntos locales, aunque no siempre es posible, por lo que puede aceptarse un mínimo local, siempre y cuando sea un valor significativamente bajo a la función de pérdida. (Goodfellow et al., 2016).

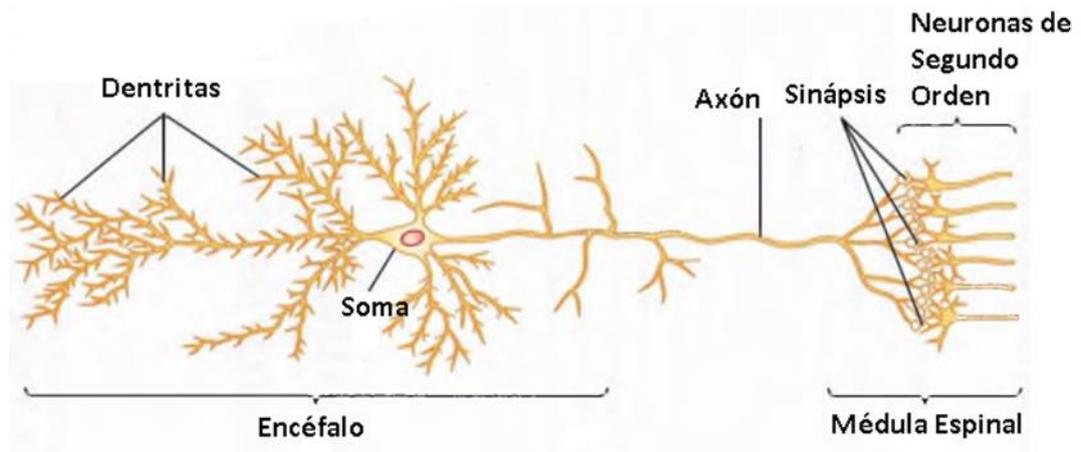
1.3.3. Redes Neuronales Artificiales (RNA)

Lo que llegamos a entender por red neuronal es: un sistema estructurado de procesamiento de información inspirado en el funcionamiento y aprendizaje realizado por las neuronas biológicas.

En cuanto a complejidad, las redes neuronales no tienen comparación, ni en su capacidad de realizar procesos de pensamiento o acciones de control. Esta red recibe millones de fragmentos de información que son absorbidos hasta generar respuestas en el sistema. El sistema nervioso central contiene más de 100.000 millones de neuronas. (Guyton y Hall, 2016).

Figura 1

Estructura de una red neuronal biológica.



Nota. De “Tratado de Fisiología Médica” por Z. Guyton y Hall, 2016, Decimotercera Edición.

Las Neuronas se identifican por tres partes:

- Cuerpo celular o Soma.
- Dendritas.
- Axón.

Las neuronas son células excitables especializadas en la recepción de estímulos y la conducción del impulso nervioso. Todas tienen un cuerpo celular o pericarion, que a partir de ellas se prolongan en su superficie las neuritas, las cuales son responsables de recibir la información y conducirla. La neurita tubular larga recibe el nombre de axón, el cual junto con la dendrita se conocen como las fibras nerviosas. (Snell, 2014)

Las RNA utilizan principalmente como herramienta predictora y analítica de variables dentro de la Big Data. Se basan en la utilización de neuronas como estructura de nodos, donde se produce el aprendizaje constante a medida que la red de datos es alimentada.

Según Escamilla-García et al. (2020) es un algoritmo basado en el concepto de una neurona biológica, puesto que los elementos de

procesamiento (neurona) del modelo computacional están interconectados con coeficientes (pesos) en cada una de las conexiones.

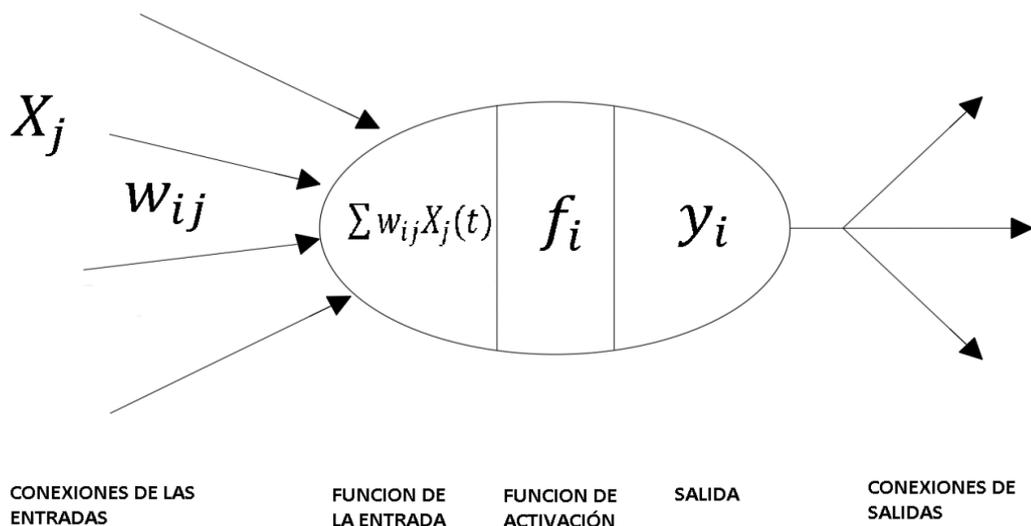
De acuerdo a Serrano et al. (2010), cada autor define de una manera diferente las RNA ya que no hay un concepto general, pero todas ellas tratan de simular el comportamiento biológico. Estos elementos tienen una similitud con el cerebro humano por la distribución de las operaciones a realizar en una serie de elementos básicos, que en otras palabras se conoce como neuronas, pues estos elementos se interconectan.

De acuerdo a Escamilla-García et al. (2020), una neurona biológica consta de tres partes y la neurona artificial tiene la misma estructura, como se aprecia en la **Figura 1**

Denominado también procesador elemental a un dispositivo simple de cálculo que, a partir de un vector de entradas $x(t)$ provenientes del exterior o de otras neuronas, proporciona una única respuesta o salida $y(t)$. Los elementos esenciales que constituyen la neurona (i) son los siguientes:

Figura 2

Estructura de una red neuronal artificial.



Nota. La neurona artificial tiene un conjunto de entrada que denotadas $X_j(t)$ y de pesos sinápticos w_{ij} . Ellas tienen una regla de propagación $h_i(t) =$

$\sum w_{ij}X_j(t)$. Que a su vez tienen una función de activación $y_i(t) = f_i(y_i(t - 1), h_i(t))$, que proporcionan una salida $y_i(t)$. De: Propia.

1.3.3.1. Perceptrón

Según Casas Roma *et al.* (2019), perceptrón es la unidad más básica de las RNA, esta unidad ayuda a resolver problemas sencillos y su operación es parecida a la regresión no lineal. Esta unidad está conformada por una neurona conectada a una capa de entrada con su respectiva capa de salida (Zhang *et al.*, 2018).

Según Wang *et al.* (2020), el perceptrón sin función de activación puede expresarse de la siguiente manera:

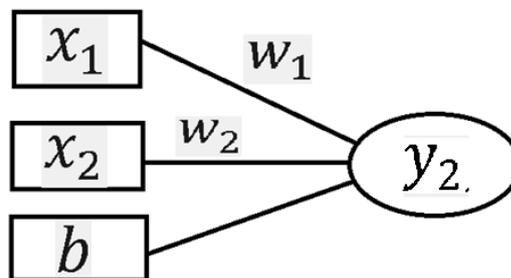
$$y_1 = w_1x_1 + w_2x_2 + b \quad (1.1)$$

$$y_1 = \sum_{i=1}^n w_i x_i + b \quad (1.2)$$

Tomando las ecuaciones (1.1) y (1.2), se representaría de forma gráfica como se muestra en la Figura 3.

Figura 3

Perceptrón sin función de activación



Nota. Adaptado de “La influencia de la función de activación en un modelo de red neuronal de convolución del reconocimiento de expresiones faciales” por Y. Wang *et al.*, 2020, *Applied Sciences*, 10(5), p. 3, (<http://dx.doi.org/10.3390/app10051897>). CC BY.

Wang *et al.* (2020) en su investigación señala que la ecuación anterior permite resolver tan solo problemas lineales, por lo que es necesario añadir una función de activación no lineal para permitir resolver problemas no lineales.

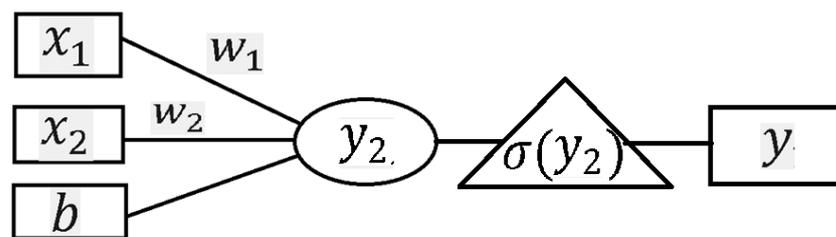
$$y_2 = w_1x_1 + w_2x_2 + b \quad (1.3)$$

$$y = \sigma(y_2) \quad (1.4)$$

Tomando en cuenta las ecuaciones (1.3) y (1.4), se representaría de forma gráfica como se muestra en la Figura 4.

Figura 4

Perceptrón con función de activación



Nota. Adaptado de “La influencia de la función de activación en un modelo de red neuronal de convolución del reconocimiento de expresiones faciales” por Y. Wang et al., 2020, Applied Sciences, 10(5), p. 3, (<http://dx.doi.org/10.3390/app10051897>). CC BY.

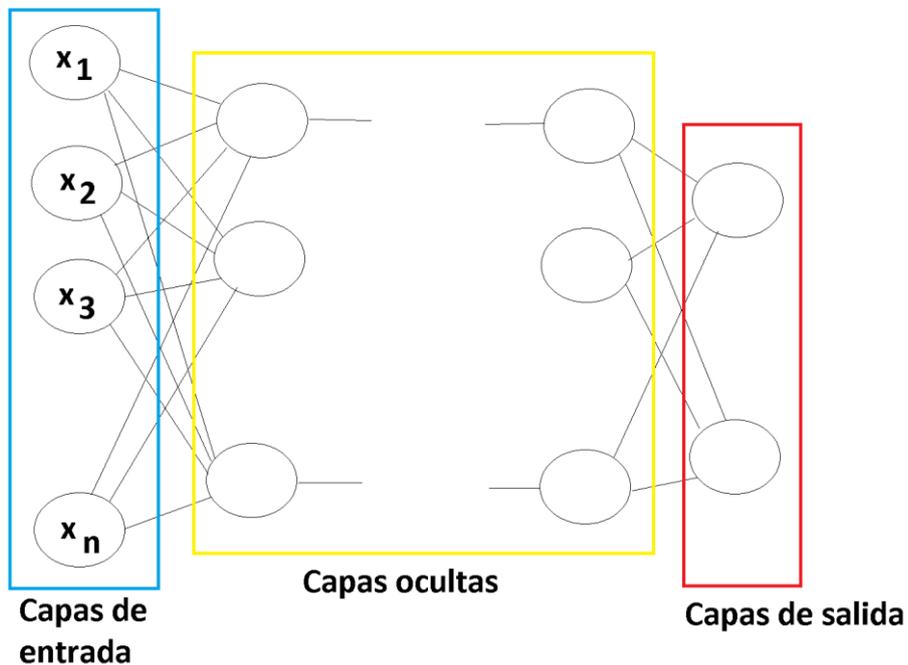
1.3.3.2. Perceptrón multicapa

El Perceptrón Multicapa (MLP), se distingue debido a que interconecta un conjunto de entradas con un conjunto de salida que a su vez está conectada a múltiples capas ocultas, está compuesto por tres partes importantes (Jassam Mohammed et al., 2020):

- a) Capas de entrada.
- b) Una o más capas ocultas.
- c) Una capa de salida.

Figura 5

Arquitectura general de un MLP



Nota. Adaptado de “Análisis de electroencefalografía basado en aprendizaje profundo: una revisión sistemática” por Y. Roy *et al.*, (2019), *Journal of Neural Engineering*, 16(5), p. 5, (<https://doi.org/10.1088/1741-2552/ab260c>). CC BY.

1.3.3.3. Método *Backpropagation* (BP)

Escamilla-García *et al.* (2020), es un entrenamiento que se le aplica a las RNA, para aplicar el aprendizaje supervisado.

Este método consiste en realizar la propagación hacia adelante (feedforward), que consiste en entrenar el modelo de RNA para obtener una salida. Luego ocurre la propagación hacia atrás (backpropagation), en otras palabras, se calcula el error cometido en la capa de salida, después ese mismo error se propaga hacia atrás hasta conseguir modificar los pesos sinápticos correspondientes. (Casas Roma *et al.*, 2019).

Saikia *et al.* (2020) indica una serie de pasos para entender el funcionamiento del BP:

1. Se inician los pesos de la red de forma aleatoria.
2. Se presenta el conjunto de entrenamiento de la red.
3. Con el fin de generar la salida se propaga la entrada.
4. Para calcular el error se realiza una comparación entre la salida esperada y la salida de la red.
5. Para minimizar el error es necesario ajustar los pesos.
6. Se repiten los pasos del 2 al 6 y el proceso se repite hasta que el error no disminuya más.

1.3.4. Redes neuronales artificiales: Componentes

1.3.4.1. La neurona

Es la unidad más básica de una RNA y es aquella que se encarga de aplicar funciones a los datos de entrada, que a su vez es recibida de otra neurona, lo que genera un nuevo valor de salida. estas unidades contienen sus pesos sinápticos y una función que se aplica para la conexión entre ellas. (Escamilla-García et al., 2020)

- Función suma ponderada.

$$z(x) = \sum_{j=1}^n x_j \cdot w_j^i \quad (1.5)$$

- Función máxima.

$$z(x) = \max(x_1 w_1^i, \dots, x_n w_n^i) \quad (1.6)$$

- Función mínima.

$$z(x) = \min(x_1 w_1^i, \dots, x_n w_n^i) \quad (1.7)$$

- Función lógica AND o OR.

$$z(x) = \min(x_1 w_1^i \wedge \dots \wedge x_n w_n^i) \quad (1.8)$$

$$z(x) = \min(x_1 w_1^i \vee \dots \vee x_n w_n^i) \quad (1.9)$$

1.3.4.2. Función de pérdida

La función de pérdida es la parte de la red que permitirá evaluar el éxito del modelo. Estas funciones se usan de la misma manera en la que se evalúan las funciones lineales. Existen algunas recomendaciones cuando se tratan de problemas de clasificación, regresión y predicción de secuencias (Chollet, 2018).

François Chollet (2017) indica las funciones de pérdida y su respectivo uso en las RNA:

- **Entropía Cruzada (*Binary cross-entropy*)**. Se recomienda el uso cuando la salida es una probabilidad y una clasificación de dos clases.
- **Entropía Cruzada Categórica (*Categorical cross-entropy*)**. Se recomienda el uso para clasificaciones con más de 2 clases.
- **Error Cuadrático Medio (*Mean squared error, MSE*)**. Utilizado en problemas de regresión. Para calcular el MSE se utiliza el cuadrado de la diferencia entre las predicciones y los valores reales.
- **Error Absoluto Medio (*Mean Absolute Error, MAE*)**. Utilizado para medir el error en la etapa de entrenamiento. Cuando se utiliza con la validación cruzada nos determina el número de epochs óptimo para el modelo de RNA.

1.3.4.3. Unidades de salida

La elección del tipo de unidad de salida depende del problema que queramos solucionar. Por consiguiente la función de activación y pérdida a utilizar por la neurona para calcular la salida de la RNA (Szandafá, 2021). En otras palabras, se elige la forma de representar la salida, en consecuencia, de ello se determina la forma que tendrá la función de

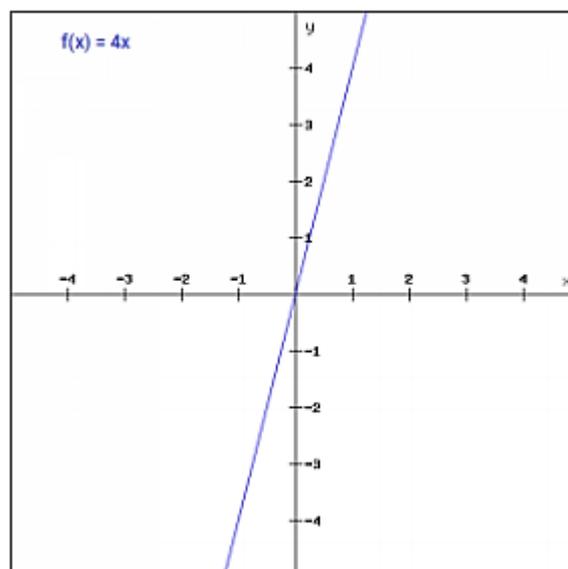
perdida. Todas las unidades usadas como salida de un modelo de RNA pueden utilizarse como unidades ocultas (Goodfellow et al., 2016).

Uno de los hiperparámetros usados en las RNA es la función de activación. Esta función afectara en el entrenamiento y el desempeño del modelo, que a su vez va a afectar el aprendizaje y generalización. Esta función determina la actividad de una neurona. (Nwankpa et al., 2018). El mayor aporte de la función de activación es que permite a el modelo de RNA ajustarse a las curvas y superficies no lineales.

- **Función Lineal.** A pesar que la gradiente de esta función no es cero, esta es directamente proporcional a su entrada. El principal problema de esta función es que la derivada de la función siempre es constante, lo cual genera que la última neurona permanezca lineal a los valores de las neuronas antecesoras, debido a esto no minimiza el error y son utilizadas en la capa de salida en problemas de regresión. (Szandala, 2021).

Figura 6

Función de activación Lineal

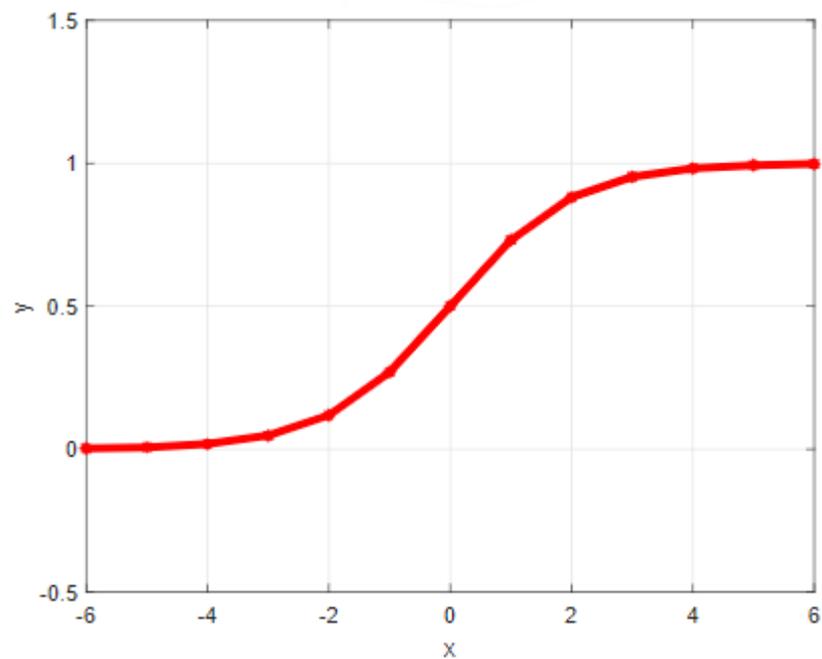


Nota. De “Activation Functions in Neural Networks” por S. Sharma et al., 2020, *International Journal of Engineering Applied Sciences and Technology*, 4(12), p. 312, (<https://doi.org/10.33564/IJEAST.2020.v04i12.054>). CC BY.

- **Función Sigmoide.** Es una función no línea que cuenta con la ventaja de estar acotada, dando resultados entre 0 y 1. Son utilizadas en clasificación binaria, regresión logística, entre otros. La desventaja de esta función es que mientras se acerca a los valores límites, la RNA aprende despacio o puede llegar a un punto en el que deja de aprender. Al no estar centrada en 0 complica el aprendizaje, todo debido a que los pesos no se pueden aumentar y disminuir a la vez. (Szandała, 2021).

Figura 7

Función de activación Sigmoide



Nota. De “La influencia de la función de activación en un modelo de red neuronal de convolución del reconocimiento de expresiones faciales” por Y. Wang *et al.*, 2020, *Applied Sciences*, 10(5), p. 5, (<http://dx.doi.org/10.3390/app10051897>). CC BY.

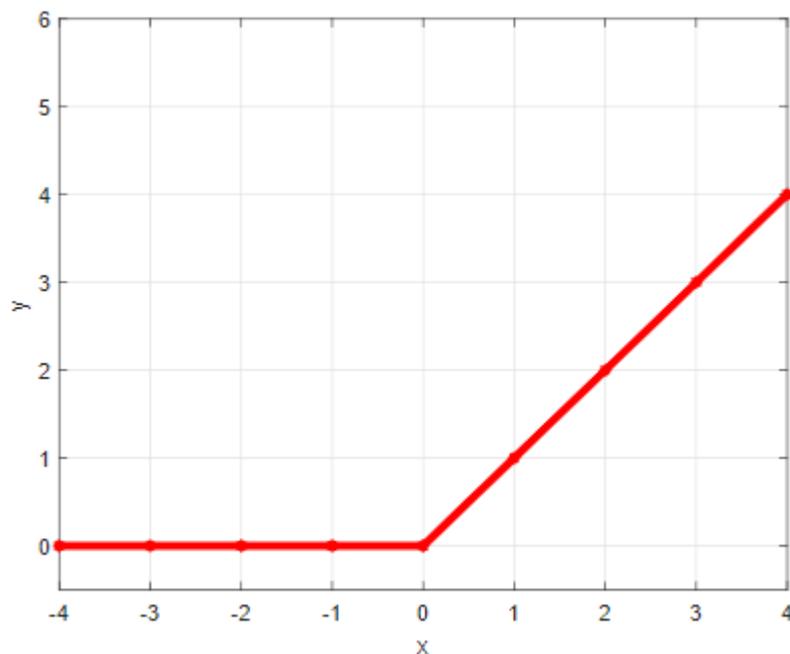
1.3.4.4. Unidades ocultas

A esta parte de la RNA se les llama funciones de las neuronas ocultas, las cuales toman a la función de entrada y lo transforman para que cada neurona haga más aproximaciones lineales. Lo complicado en esta parte es determinar la función que dará el mejor desempeño de la RNA, este proceso consiste en ensayo de prueba y error. (Goodfellow et al., 2016).

- a) Intuir la función de activación logre sea el óptimo para la neurona oculta.
 - b) Una vez elegida la función de activación se entrena el modelo.
 - c) Evaluar el rendimiento del modelo.
- **Función ReLU.** Las *Rectified Linear Units* son la función más utilizada en la actualidad, debido a la facilidad para optimizar tanto como una función lineal a pesar de la diferencia que muestra en la Figura 8. La desventaja de esta función es que propone valores negativos en 0, lo cual trae complicaciones en el aprendizaje, debido a que las neuronas no vuelven a aprender. (Goodfellow et al., 2016).

Figura 8

Función de activación ReLU

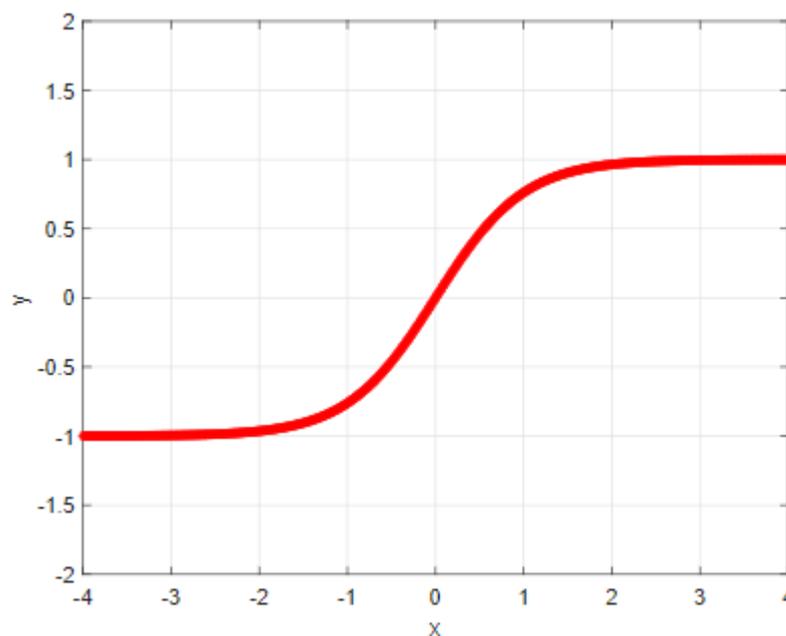


Nota. De “La influencia de la función de activación en un modelo de red neuronal de convolución del reconocimiento de expresiones faciales” por Y. Wang *et al.*, 2020, *Applied Sciences*, 10(5), p. 5, (<http://dx.doi.org/10.3390/app10051897>). CC BY.

- **Función Tangente hiperbólica.** Esta función presenta entradas negativas en su función, lo cual permitirá a la RNA aprender nuevamente cuando sus valores resulten negativos. La desventaja de esta función es debido a que va tomando un valor más pequeño donde le es imposible corregir y la RNA no llega a aprender (Goodfellow *et al.*, 2016).

Figura 9

Función de activación Tangente hiperbólica



Nota. De “La influencia de la función de activación en un modelo de red neuronal de convolución del reconocimiento de expresiones faciales” por Y. Wang *et al.*, 2020, *Applied Sciences*, 10(5), p. 5, (<http://dx.doi.org/10.3390/app10051897>). CC BY.

1.3.5. Redes neuronales artificiales: Desarrollo de modelos

A pesar del aumento del número de investigaciones en el campo de RNA no se ha podido proporcionar una evaluación estadística del rendimiento ni una justificación de las elecciones de la metodología del desarrollo de los modelos. La falta de evaluación de estos parámetros ha logrado que muchos investigadores desarrollen modelos de RNA por medio de ensayo y error, debido a esto es necesario enfocarse en conocer los factores que afectan al rendimiento del modelo, dentro de estos se encuentra las muestras del conjunto de datos, las variables de entrada y su debido procesamiento, así como la arquitectura y los hiperparámetros que ayudan a la optimización y validación del modelo. (Lathuiliere et al., 2020).

1.3.5.1. División del conjunto de datos

CHOLLET (2018), en sus investigaciones nos recomienda no utilizar el mismo conjunto de datos para entrenar como para evaluar debido a la filtración de información de los datos. Esto es debido a que la RNA va a lograr aprender los patrones del conjunto de datos, los cuales al momento de realizar la evaluación la RNA nos brindará el dato ingresado y no el procesado; que es la finalidad de la investigación. Debido a esto es necesario dividir el conjunto de datos en conjunto de datos para validación y conjunto de datos para prueba (conjunto de validación y conjunto de prueba respectivamente).

SINGH & SINGH (2016), en su investigación recomienda que utilicemos el 20% de los datos para la prueba, 80% de datos restantes para el entrenamiento y el 20% (de los datos restantes) para la validación. El conjunto de entrenamiento y el conjunto de validación son usados durante el entrenamiento, sin embargo, el conjunto de validación se usa para ajustar el modelo, es decir, encontrar el número de neuronas, etc. Una vez ajustado el modelo, se evalúa la RNA con el conjunto de prueba.

Algunos puntos a tener en cuenta a la hora de dividir los datos según François Chollet (2017) son:

- a. la representatividad de los datos.
- b. la flecha del tiempo.
- c. la redundancia de los datos.

Dentro de la representatividad de los datos, el autor se refiere a que la selección de conjunto de datos, ya sea para el entrenamiento, validación y prueba tengan la misma variedad de datos entre sí a pesar de que no se deben repetir, para esto se recomienda mezclar los datos al azar y luego dividirlos.

Las RNA que siguen una continuidad en el tiempo no deben mezclar sus datos de forma aleatoria, para esto se recomienda que la data de los valores del conjunto de entrenamiento sea anterior a los datos del conjunto de validación.

Cuando se trabaja con datos recopilados del mundo real, aparecen los casos donde existen datos que se repiten tanto en el conjunto de entrenamiento como en el conjunto de validación la mejor solución es asegurarse que el conjunto de entrenamiento y conjunto de prueba no tengan datos iguales.

1.3.5.2. Validación cruzada (*Cross Validation*)

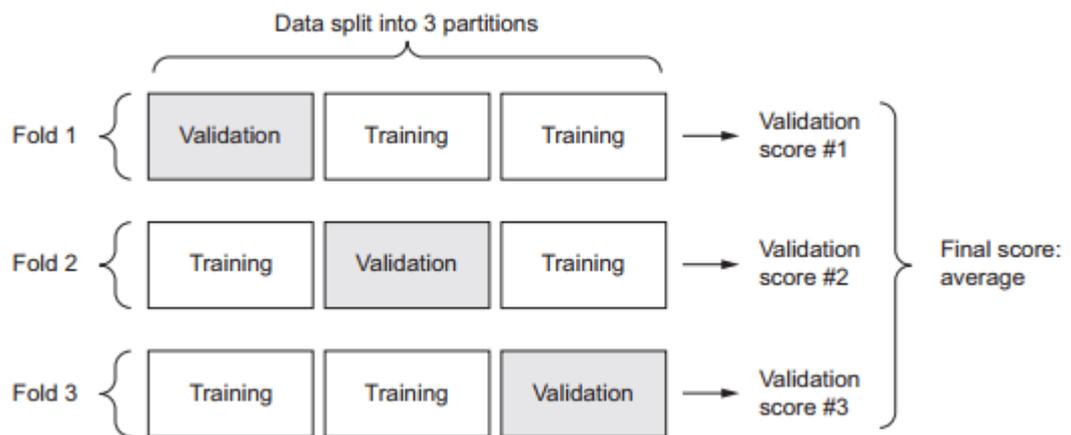
El procedimiento de validación cruzada surge a causa de la escasez de datos en el conjunto de entrenamiento y el conjunto de prueba, esto debido a la poca variabilidad en torno al error estimado, lo que resultaría muy difícil determinar el modelo óptimo. Este procedimiento consiste en dividir el conjunto de entrenamiento en K partes de igual tamaño, dentro de las cuales se van a separar en los conjuntos con los que se van a entrenar ($K - i$) y el conjunto que se va a evaluar (i), Cabe mencionar que, el modelo se divide y entrena repetidas veces. (Chollet, 2018).

A modo de entendimiento en la **Figura 10**, se realizó una validación cruzada, dividiendo así los datos de entrenamiento en 3 partes del tamaño más similar: donde 2 partes son para el entrenamiento y 1 para la validación, a partir de eso se entrena una secuencia de modelos donde se

van intercalando, este proceso se repite hasta completar todas las conjugaciones. Para cada modelo mostrado se calcula la precisión y luego se realiza un promedio de tal.

Figura 10

Validación cruzada: División del conjunto de entrenamiento en 3 particiones



Nota. Para el ejemplo de la imagen, el conjunto de entrenamiento se ha dividido en K igual a 3. Se entrena el modelo con $K - 1$ igual a 2 particiones. Finalmente, se evalúa con la partición restante: i igual a 1. Esto se repite para los siguientes pliegues. De François Chollet, 2017.

1.3.5.3. Preprocesamiento de los datos

Según lo explicado en el capítulo de división de datos, para un correcto procesamiento de datos, las RNA y otros modelos de aprendizaje automático requieren que los datos de entrada tengan características específicas (Casas Roma et al., 2019). En otras palabras, las muestras del conjunto de datos deben ser independiente entre sí, y la distribución del conjunto de entrenamiento y el conjunto de prueba debe estar de forma similar. (Goodfellow et al., 2016).

1.3.6. Redes neuronales artificiales: Optimización de modelos

Los optimizadores son aquellos que se encargan de encontrar los parámetros (pesos) de una función de costo (función de pérdida), esta misma es la que permite cuantificar la calidad de los pesos. La optimización del modelo se establece después de entrenar el modelo, que consiste en realizar ajustes de los parámetros.

1.3.7. Redes neuronales artificiales: Arquitectura

Uno de los aspectos importantes en el diseño de un modelo de RNA es determinar su arquitectura. Cuando hablamos de la arquitectura de una RNA nos referimos a el número de neuronas y la forma en las que están interconectadas. La arquitectura de una RNA depende de la estructura y el problema que se desea resolver, teniendo en consideración los problemas de overfitting y overtraining. (Casas Roma et al., 2019).

- Dimensión de la capa de entrada: el número de neuronas en la capa de entrada debe ser igual al número de características del conjunto de entrenamiento. Para escoger el número de variables se puede ayudar de una revisión de literatura y correlación entre estas, de esta manera se pueden encontrar las variables de mayor significancia. Según Casas Roma *et al.* (2019), dependerá del número de entradas que el investigador considere para el modelo.
- Dimensión de la capa de salida: El número de capa de salida depende del número de objetivos y problemas que se quieren lograr con la RNA. Para modelos de regresión como lo es esta investigación sólo pueden tener un único valor de salida.
- Dimensión de capa oculta: Una capa oculta puede ser suficiente para que una RNA aproxime cualquier función no lineal. Pero para llegar el modelo óptimo de nuestra RNA no existe una fórmula específica para hallar el número de capas ocultas óptimo para nuestra RNA. En cuanto al número de neuronas en cada capa oculta, según Casas Roma *et al.* (2019), es de 1 a 2

veces el número de neuronas de entrada y, según Velasco *et al.* (2019), el número máximo sin que se produzca overfitting está dado por la ecuación (1.10) de Stathakis’.

$$Nh = \frac{Ns}{(\alpha * (Ni + No))} \quad (1.10)$$

Donde, el número de neuronas en la capa oculta Nh se calcula dividiendo el número de datos del conjunto de entrenamiento Ns con el factor de escala arbitraria α multiplicado por la sumatoria del número de neuronas en la capa de entrada Ni y el número de neuronas en la capa de salida No ; el factor α varía de 2 a 10.

1.3.7.1. Algoritmos con tasas de aprendizaje adaptativas

El proceso de establecer la tasa de aprendizaje es importante y problemático, esto es debido a, si se establece una tasa de aprendizaje demasiado grande va a originar que se disperse el mínimo deseado y si se coloca una tasa de aprendizaje pequeña va a originar que el modelo sea muy lento para converger. En la practica la tasa de entrenamiento va disminuyendo con el tiempo (Lathuiliere et al., 2020).

Se han desarrollado una variedad de algoritmos con tasas de aprendizaje adaptativas, estos representan a algunos de los algoritmos de optimización más utilizados en la actualidad:

- **AdaGrad (*Adaptive Gradient*)**: el algoritmo adapta la tasa de aprendizaje de cada peso dividiéndolo por la raíz cuadrada de la suma de sus valores históricos al cuadrado. Muestra una ventaja en disminución de la tasa de aprendizaje para los pesos con gradientes altos, a diferencia de los que tienen pesos con actualizaciones pequeñas tendrá una disminución más lenta en su tasa de aprendizaje.
- **RMSProp (*Root Mean Squared Propagation*)**: Es una modificación del algoritmo AdaGrad con el objetivo de evitar que los índices de aprendizaje se reduzcan de forma agresiva. Este algoritmo mantiene un promedio móvil del gradiente cuadrado para cada peso y divide el gradiente por la raíz

cuadrada de este valor. Este algoritmo ha demostrado un gran rendimiento debido a su aumento de aceleración de la convergencia en el entrenamiento de RNA.

- **AdaDelta:** la característica principal de este algoritmo es que logra brindar a cada uno de los parámetros una tasa de aprendizaje diferente. Realiza cambios pequeños en los parámetros que más se actualizan y para los que no, genera cambios más grandes.
- **Adam (*Adaptive Moment*):** Según Yi *et al.* (2020), actualmente es el algoritmo que mejor rendimiento brinda. Este algoritmo tiene las ventajas que ofrece el algoritmo de optimización del momento y algoritmo AdaGrad, este algoritmo acelera la convergencia en las fases iniciales del entrenamiento. (Yixiang Wang *et al.*, 2019)

1.3.7.2. Selección del número de *epochs*

Se denomina época (*epoch* en inglés) al proceso de pasar los datos de entrenamiento por la RNA. Pasar todos los por un solo *poch* no es buena idea debido a que la RNA no va a lograr aprender lo suficiente, por esta razón la actualización de los pesos debe hacerse mediante varios *epochs*. El número de *epoch* es un factor que determina el aprendizaje de la RNA; mediante se incrementa el número de *epochs* la RNA pasa de ajuste insuficiente a un ajuste satisfactorio y a un ajuste excesivo.

Según Amirabadi *et al.* (2020), el número de *epochs* va a depender del número de conjunto de datos y a la diversidad de los datos. Por lo general es recomendable utilizar un número fijo de *epochs* durante el proceso de optimización de los hiperparámetros para evaluar el efecto de rendimiento (Díaz *et al.*, 2017)(Kim, 2019).

Dinamarca (2018), en su investigación nos recomienda que el número de *epochs* se explore en escala logarítmica, debido a que los efectos son multiplicativos, por ejemplo, en rangos como, {10, 30, 100, 300, 800, 1000, 1500}.

1.3.7.3. Selección del *batch size*

Según Goodfellow *et al.* (2016), el número de lote o *batch size* en inglés consiste en el número de datos de entrenamiento utilizados en la fase de entrenamiento. Este factor permite evaluar la gradiente de la función de pérdida y, en función al número de datos se puede clasificar en:

- a) Descenso de gradiente por dividir el conjunto de entrenamiento en lotes.
- b) Descenso de gradiente por dividir el conjunto de entrenamiento por mini lotes.
- c) Descenso de gradiente por utilizar un solo dato del conjunto de entrenamiento se hace llamar estocástico.

Según Amirabadi *et al.* (2020), nos recomienda dividir el número de *batch size* en diversos lotes debido a las limitaciones del hardware. Esto es debido a la limitada capacidad de memoria del GPU, lo cual limita el número de datos que se pueden procesar a la vez, por lo que se recomiendan potencias de 2 por ejemplo, un *batch size* de 2, 4, 8, 16, 32, 64, 128 y 256 (Balles *et al.*, 2017).

1.3.7.4. Optimización bayesiana (Bayesian optimization)

La Bayesian optimization es una herramienta para poder obtener los hiperparámetros óptimos. El procedimiento consiste en probar de forma aleatoria hiperparámetros con todos los resultados obtenidos, este procedimiento se repite constantemente hasta lograr el objetivo. La decisión del software se basa en la relación entre la cantidad de datos explorados y explotados. El fin de la exploración es probar con tantos hiperparámetros posibles que nos den la mejor arquitectura, y con la explotación se empieza a buscar nuevos valores en base a los mejores resultados.

1.4. Formulación del problema

¿Cómo se evaluarán departamentos habitacionales en la ciudad de Lima aplicando un modelo de Redes Neuronales Artificiales?

1.5. Justificación e importancia del estudio

Varios autores han investigado la creación de modelos de Redes Neuronales Artificiales para valorizar inmuebles. Donde demuestran una ventaja de la aplicación de redes neuronales artificiales sobre los métodos tradicionales, pues este método tiene la capacidad de representar complicadas conexiones dentro las variables de entrada y salida, además, la posibilidad de aprender y generalizar su conocimiento. Estas ventajas que presenta este método han permitido que se acepten como herramientas fiables, pues permiten modelar los complejos comportamientos y la variabilidad de los inmuebles a valorizar.

A pesar de que las redes neuronales artificiales es una metodología antigua y que en otros países se han realizado diversas investigaciones, son pocas las disciplinas de la ingeniería civil en Perú que utilizan esta metodología. Dicho esto, es necesario dar a conocer a la comunidad científica la ventaja de utilizar las redes neuronales artificiales. Debido a que resuelven problemas más complicados y permiten predecir valores de manera precisa, siendo esta metodología muy superior a los métodos tradicionales.

Esta investigación se ha realizado pensando en el alcance social, debido a que sirve como una herramienta para valorizar departamentos en Lima de una forma sencilla. De igual manera esta investigación cuenta con el aspecto económico porque los resultados de la valorización pueden obtenerse sin la necesidad de realizar una extensa campaña de valorización. Esto llega a significar un ahorro a la hora de realizar valorizaciones de departamentos.

1.6. Hipótesis

A través de la aplicación de un modelo de Redes Neuronales Artificiales permitirá evaluar departamentos habitacionales en la ciudad de Lima.

1.7. Objetivos

1.7.1. Objetivo general

Desarrollar un modelo de Redes Neuronales Artificiales para evaluar departamentos habitacionales, que permita mejorar el proceso de valorización de un departamento con los datos brindados por el Banco Central de Reserva del Perú.

1.7.2. Objetivos específicos

1. Determinar los hiperparámetros óptimos para el modelo de Redes Neuronales Artificiales usando la optimización bayesiana con el lenguaje de programación de Python con el software Spyder.
2. Entrenar el modelo de Redes Neuronales Artificiales con los mejores hiperparámetros con el conjunto de datos obtenidos del Banco Central de Reserva del Perú, mediante el lenguaje de programación de Python.
3. Comprobar que la estructura del modelo de Redes Neuronales Artificiales haya aprendido correctamente, comparando los valores de costo de departamentos habitacionales estimados por el modelo con los valores de departamentos habitacionales reales.

CAPITULO II:
MATERIAL Y METODO

II. MATERIAL Y MÉTODO

2.1. Tipo y diseño de investigación

2.1.1. Tipo de investigación

La tesis presentada es de tipo explicativa, puesto que, para desarrollar un modelo de red neuronal artificial, se determinó el efecto de los hiperparámetros y las entradas en la predicción del valor de los departamentos habitacionales. Así es como se determinó la causa de que la arquitectura del modelo red neuronal artificial determine la precisión del modelo. Así mismo se evaluó la correlación entre las variables de entrada y la variable de salida. (Hernández-Sampieri & Mendoza Torres, 2018).

2.1.2. Diseño de investigación

El diseño de investigación fue experimental, debido a que se realizaron diferentes modelos con distintas modificaciones en los hiperparámetros tomando el criterio de prueba y error, logrando obtener el mejor modelo como producto final.

2.2. Población y muestra

2.2.1. Población

Debido a que la finalidad de este trabajo es realizar el avalúo de Departamentos habitacionales en la ciudad de Lima, se recopiló información brindada por la Gerencia Central de Estudios Económicos del BCRP, por lo que tomaremos todos los datos brindados.

- Medición de Valorización de Departamentos agrupados por Trimestres (Primer Trimestre 1998 – Cuarto Trimestre 2020). Haciendo un total de 66 671 datos.

2.2.2. Muestra

No se requiere establecer la muestra, debido a que los datos están almacenados en el sistema de estadística del BCRP; por lo tanto, se tomará como población muestra, dado a que se evita la manipulación virtual.

2.3. Variables y operacionalización

Tabla 1

Operacionalización de la variable independiente

Variable	Dimensiones	Indicadores	Ítem	Técnica e Instrumento de recolección de datos
Modelo de redes neuronales artificiales	Conjunto de datos	Conjunto de entrenamiento	Base de datos	Recolección y análisis de datos secundarios y Ficha de registro
		Conjunto de validación		
		Conjunto de prueba		
	Hiperparámetros asociados al modelo	Número de entradas	X	Análisis del contenido y Ficha bibliográfica
		Número de salidas	y	
		Número de capas ocultas	Num layers	Observación y Ficha de registro
		Número de neuronas ocultas	Units 1; Units 2	
		Función de activación	Activation	
	Hiperparámetros asociados al	Número de épocas	Epochs	Observación y Ficha de registro
		Tasa de aprendizaje	Learning rate	

descenso del gradiente	Optimizador	Optimizer
	Tamaño de lote	Batch size

Tabla 2

Operacionalización de la variable dependiente

Variable	Dimensiones	Indicadores	Ítem	Técnica e Instrumento de recolección de datos
Predicción de los Precios en dólares de los departamentos	Rendimiento del modelo	MSE	Entrenamiento	Observación y Ficha de registro
			Validación	
			Prueba	
		MAE	Entrenamiento	Observación y Ficha de registro
			Validación	
			Prueba	
		Coeficiente de correlación	R	Observación y Ficha de registro

2.4. Técnicas e instrumentos de recolección de datos, validez y confiabilidad

2.4.1. Técnicas de recolección de datos

En esta parte de la investigación se utilizó como base de datos los registros de ventas de departamentos brindadas por el BCRP por medio de su Gerencia Central de Estudios Económicos.

Cuando analizamos el contenido nos damos cuenta que se encuentra organizadas debidamente que posteriormente se utilizarán como las variables de entrada y de salida. Al momento de observar las muestras es visible la cantidad de datos en blanco de estas, ese es un gran problema para nuestro modelo de RNA debido a que esas muestras no se pueden tomar como valor "0" debido a que no está explícita.

2.4.2. Instrumentos de recolección de datos

Después de tener la base de datos a utilizar para la RNA se utilizó una ficha de registro de datos (ver Anexo 1) para organizar de forma conveniente para la investigación. El siguiente paso fue realizar una depuración de muestras que tengan datos faltantes, los cuales no permitirían que nuestro modelo de RNA tenga el mejor rendimiento; esto se desarrolla en la sección 3.1.1. El último paso fue realizar una ficha de registro con los datos de entrada y salida que se utilizarán en la RNA, dejando de lado datos que sean irrelevantes, este proceso se desarrolla en la sección 3.1.2.

2.4.3. Procedimientos para la recolección de datos

Para realizar un modelo de RNA es necesario tener una base de datos confiable, debido a esto se realizó una búsqueda de alguna data confiable, por lo que se ingresó a un registro brindado al público por el BCRP donde nos muestran datos disgregados de su registro de ventas de departamentos en la ciudad de Lima. Esta base de datos brinda desde el

año 1998 hasta el 2020, separados debidamente en los cuatro trimestre de cada año, así como el precio de venta de los inmuebles en dólares, precio soles corrientes del presente año de realizada la transacción y el precio en soles corrientes del 2009, además del distrito, la superficie, el número de habitaciones, el número de baños, el número de garajes, piso de ubicación, vista exterior y la antigüedad de cada inmuebles en el año que se realizó la venta respectivamente. Luego de obtenido la base de datos se realizó a depurar, ordenar, revisar y codificar convenientemente para la investigación. Quedando un producto confiable que serán usadas como nuestra base de datos para las variables de entrada y salida de los modelos de RNA.

Como segundo paso se tomó en cuenta los IVDH que se utilizarán en la investigación, para ello se empleó la ficha de registros utilizada en la RNA (ver Anexo 2) en ella se recolectaron los datos que se utilizarán para los datos del ítem “X” (número de entradas) y el ítem “Y” (número de salidas).

Teniendo la base de datos el siguiente paso es importarlos a el programa Spyder, donde se definirán las entradas y salidas del modelo de RNA. Para el procesamiento del conjunto de datos se utilizó el estimador MinMaxScaler, para cada una de las variables. Para importar el estimador a Spyder se utilizó el siguiente código:

```
from sklearn.preprocessing import MinMaxScaler
```

La función MinMaxScaler se utiliza para escalar cada dato utilizando la siguiente ecuación ((2.1). Done x es el valor que se desea escalar, x_{min} el valor mínimo de todo el conjunto de datos y x_{max} el valor máximo de todo el conjunto de datos.

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2.1)$$

El tercer paso fue realizar una ficha de registro para encontrar la mejor arquitectura (ver Anexo 3). Antes de aplicarla se dividió el conjunto de datos en un conjunto de entrenamiento y un conjunto de validación, después se pasó a realizar el código para entrenar el modelo inicial y se luego a definir la cuadrícula de búsqueda de los hiperparámetros.

Se entrenó 50 modelos para la predicción utilizando la optimización bayesiana y la validación cruzada. Se entrenó 2 veces cada modelo, al finalizar se exportaron los resultados de Spyder y se aplicó la ficha para ordenar y registrar los datos sobre los ítems: num layers, units 1 y units 2, activation, epochs, learning rate, optimizer y batch size. Aquí es donde se registran los hiperparámetros usados en cada modelo y el error cometido en esta etapa de entrenamiento.

El cuarto paso fue seleccionar el mejor modelo para evaluar departamentos habitacionales en la ciudad de Lima, con lo cual se registró los datos para comprobar el rendimiento del mejor modelo (ver Anexo 4). En esta ficha se registraron el error cuadrático medio, el error absoluto medio y el coeficiente de correlación. El registro de estos datos ayudo para comparar las predicciones del modelo con los datos reales.

Todo quedo registrado en una hoja de cálculo de Microsoft Excel, la primera hoja comprende de datos registrados, los cuales fueron corroborados con la base de datos; este procedimiento se realiza para verificar si los datos predichos se asemejan a los datos reales.

2.5. Procedimiento de análisis de datos

2.5.1. Entradas y salidas del modelo

La revisión de la base de datos se utilizó para el uso de las muestras como variables de los modelos de RNA. Una manera en la que se calculó la relación entre las variables de entrada y salida a usar fue utilizando un análisis de dispersión entre las propiedades. En otras palabras, se identificó la existencia de una correlación entre el precio de los inmuebles y las variables a utilizar en los modelos de RNA. Esto con la finalidad de ver la influencia de las propiedades en la predicción.

2.5.2. División del conjunto de datos

Tomando en cuenta la recomendación de CHOLLET (2018), donde menciona que no utilicemos el mismo conjunto de datos en la etapa de entrenamiento y la etapa de prueba debido a la filtración de información. En otras palabras, cuando queramos hacer una predicción y no hemos separado nuestro conjunto de datos, el modelo de RNA ha aprendido los patrones de todas las muestras, lo que va a originar que no se haga una predicción sino un dato ingresado.

SINGH & SINGH (2016), en su investigación recomienda que utilicemos el 20% de los datos para la prueba, 80% de datos restantes para el entrenamiento y el 20% (de los datos restantes) para la validación.

2.5.3. Determinación de la arquitectura del modelo

En las fases de entranamiento, ajuste de hiperparametros y al momento de calificar el comportamiento de los modelos se utilizó como métrica el error cuadrático medio (MSE) y el error absoluto medio. En esta fórmula, y_i es el valor predicho, x_i es el valor real y n el número de datos.

$$MSE = \frac{\sum_{i=1}^n (y_i - x_i)^2}{n} \quad (2.2)$$

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (2.3)$$

2.6. Criterios éticos

Según Belmont (1979), los principios éticos que rigen una investigación son los siguientes :

- **información:** brindar la información del procesamiento y los fines de investigación para trabajos a futuro.
- **comprensión:** realizar una investigación que permita a la comunidad científica comprender la información brindada.
- **voluntariedad:** la participación debe ser de manera voluntaria.

2.7. Criterios de rigor científico

La presente investigación fue realizada teniendo en cuenta los siguientes principios de rigor científico:

- **credibilidad:** los datos recogidos en la investigación tienen concordancia con la realidad, es decir, los resultados obtenidos presentan una relación con lo que se observó.
- **neutralidad:** los resultados obtenidos en el presente estudio no fueron manipulados por el investigador.
- **relevancia:** permite determinar el logro de los objetivos propuestos en la investigación.

CAPITULO III:
RESULTADOS

III. RESULTADOS

En el presente capítulo se van a presentar los resultados para medir la eficacia del método redes neuronales artificiales para avaluar precios de departamentos habitacionales en la ciudad de lima. El desempeño del modelo se evalúa en función al termino de disminución del error de predicción. De igual forma se ha presentado los resultados para ajustar los hiperparámetros del modelo de redes neuronales artificiales, así como el número de neuronas, capas ocultas, entre otros. Se aplicó la mejor predicción con el mejor modelo y datos para el modelo. Por último, se realiza una comparación entre los valores predichos con los reales para evaluar el rendimiento del modelo de redes neuronales artificiales.

3.1. Resultados en tablas y figuras

Los resultados demuestran la hipótesis planteada en la presente investigación. En esta se sostiene que el modelo de redes neuronales artificiales propuesto para avaluar departamentos habitacionales en la ciudad de Lima, teniendo en cuenta la base de datos disgregadas brindada por el Departamento de Gerencia Central de Estudios Económicos del BCRP, el cual predice el precio con mayor facilidad, precisión y menor tiempo.

3.1.1. Exploración del conjunto de datos

La población y muestra de la presente investigación está en la base de datos brindada por el Departamento de Gerencia Central de Estudios Económicos del BCRP, donde se encontraron una cantidad de 66 610 muestras con 13 características cada una de ellas. En la Figura 11 y en la

Figura 12 indica las características de los departamentos habitacionales vendidos en la ciudad de Lima.

Dentro de las cuales se separan en el año en el que se realizó la venta (desde 1998 hasta el 2020), separados en el trimestre del año en que se realizó la venta, el precio en dólares corrientes vendida dicho año, el precio en soles constantes de 2009, el distrito perteneciente del departamento, la superficie del suelo, el número de habitaciones, el número de baños, el número de garajes, el piso de ubicación, si cuenta con vista exterior y el año de antigüedad en el año de realizada la venta.

Figura 11

Distribución del conjunto de datos según el año de venta del departamento.

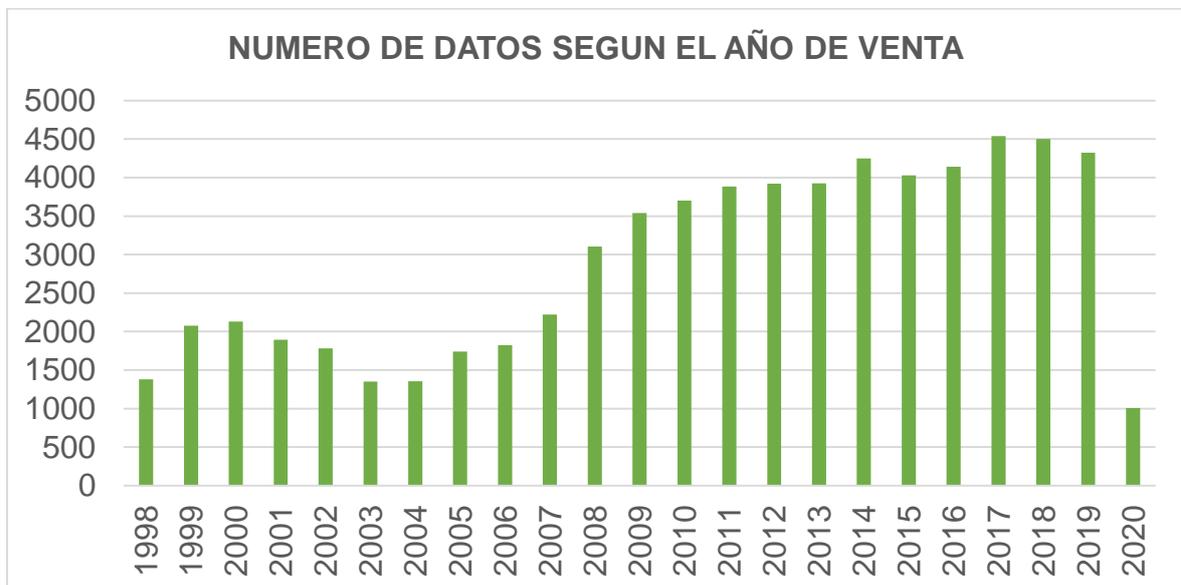
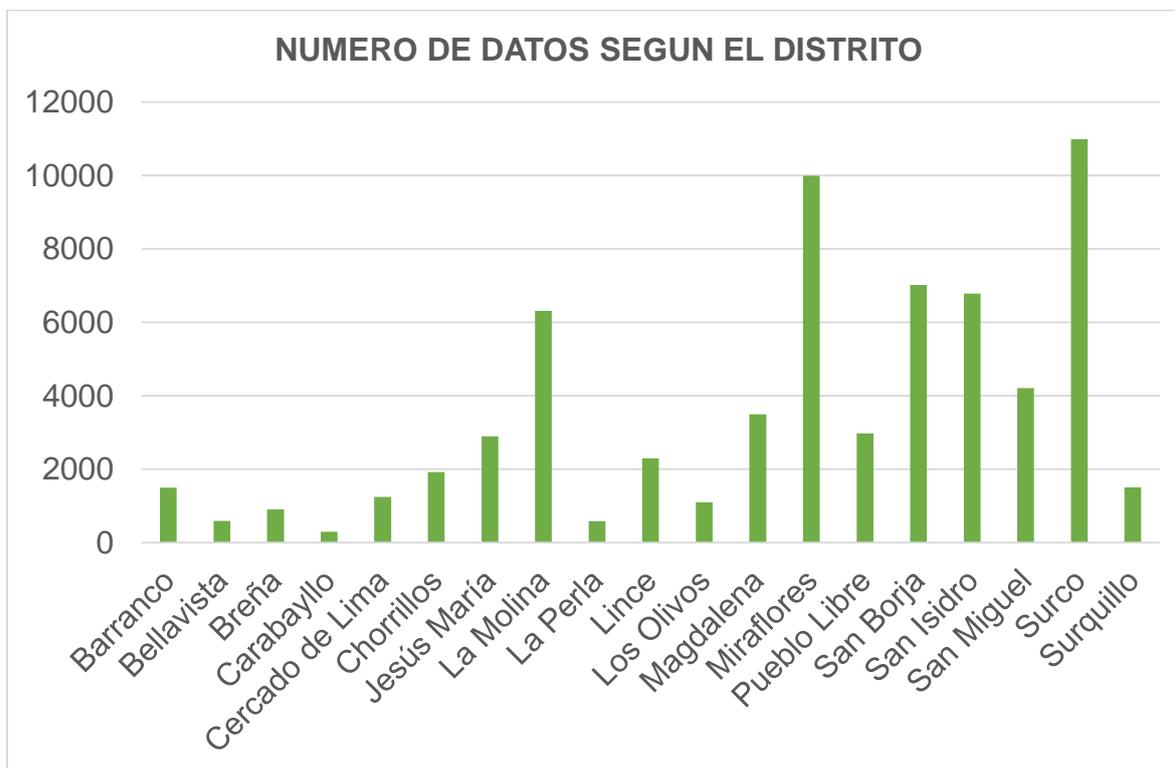


Figura 12

Distribución del conjunto de datos según el distrito del departamento.



De todo este conjunto de datos se llegó a la conclusión que se debería depurar una gran cantidad de muestras, si bien es cierto que la RNA aprende mucho mejor con una base de datos amplia, no quiere decir que el exceso de datos sea lo más conveniente para dicho método, pues va a alterar el algoritmo de aprendizaje. Para ello se depuró la data, tomando en cuenta tan solo datos a partir del primer trimestre del 2015 hasta el cuarto trimestre del 2020. Además de depuró algunos datos inconclusos, casilleros vacíos, pues la RNA no puede aprender de forma correcta debido a que esos datos no se pueden interpretar como “0” si no son verificados. Teniendo así un conjunto final de 19 616 muestras.

3.1.2. Entradas y salidas del modelo

La selección de las variables de entrada y salida para la RNA se realizó a partir de la base de datos brindada por el Departamento de Gerencia Central de Estudios Económicos del BCRP.

En diferentes investigaciones relacionadas con el tema se ha tomado en cuenta modelos de diferentes literaturas, donde se fundamentan las variables más influyentes dentro del modelo de RNA, pero debido a que contamos con variables limitadas por tener una única base de datos, pues se tomarán tan solo aquellas que guarden mayor correlación con la salida que deseamos.

Debido a que la finalidad del estudio es llegar a el costo de un departamento y la mayoría de los procesos de venta son realizados en dólares corrientes, entonces se tomó aquella variable como salida.

Para las variables de entrada se han tomado la elección de los IVDH. Para la elección de los IVDH se ha considerado la información recopilada del BCRP, para ser precisos en su base de datos disgregada de “indicadores de precios de alquiler y ventas de departamentos”, esta recopilación de información ha sido tomada de la ciudad de Lima, en algunos distritos, cómo Barranco, Bellavista, Breña, Carabayllo, Cercado de Lima, Chorrillos, Jesús María, La molina, La perla, Lince, Los olivos, Magdalena, Miraflores, Pueblo Libre, San Borja, San Isidro, San Miguel, Surco y Surquillo. La información utilizada para la presente investigación me brinda datos desde el 2015 hasta el 2020, dónde sus índices son año de la venta, trimestres del año realizado, precio en dólares corrientes, precio en soles corrientes, precio en soles constantes de 2009, el distrito del departamento, la superficie del departamento, número de habitaciones, numero de baños, numero de garajes, piso de ubicación, vista exterior y años de antigüedad del departamento en el año que se realizó la venta.

En las variables de entrada se descartaron las siguientes:

- ~~Año~~: esta variable se descarta debido a que la finalidad de esta investigación es llegar a avaluar departamentos habitacionales a futuro, no a avaluar algo que ya se vendió, pues a pesar de que se llega a comprobar en algún punto de la investigación, esta variable no llega a influir.
- ~~Trimestre~~: esta variable se descarta pues no llega a influir en la salida de la red neuronal artificial
- ~~Precio en dólares corrientes~~: esta variable se descarta debido a que es la variable de salida de la red neuronal artificial.
- ~~Precio en soles corrientes~~: Esta variable se descarta debido a que ya se toma como variable de salida el precio en dólares del departamento habitacional.
- ~~Precio en soles del 2009~~: Esta variable se descarta debido a que ya se toma como variable de salida el precio en dólares del departamento habitacional.

Tabla 3

Entradas y salida del Modelo de Red Neuronal Artificial propuesto.

Variables del BCRP	ENTRADAS DE LA RED NEURONAL ARTIFICIAL	SALIDA DE LA RED NEURONAL ARTIFICIAL
Año	x	x
Trimestre	x	x
Precio en dólares corrientes	x	✓
Precio en soles corrientes	x	x
Precio en soles del 2009	x	x
Distrito	✓	x
Superficie	✓	x

Número de habitaciones	✓	x
Número de baños	✓	x
Número de garajes	✓	x
Piso de ubicación	✓	x
Vista al exterior	✓	x
Años de antigüedad	✓	x

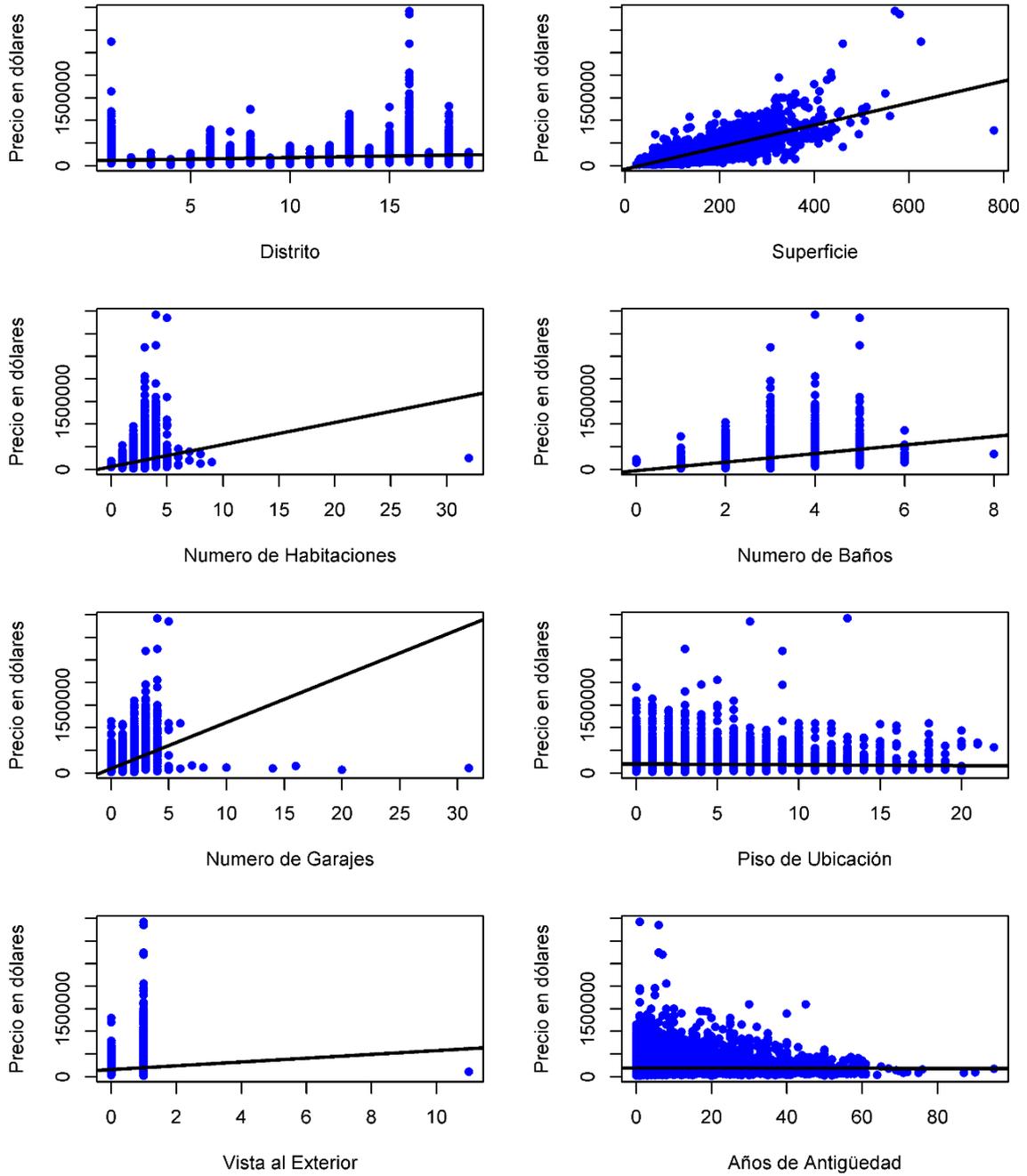
En la **Tabla 3** se ha colocado la lista final de las variables que se utilizarán en la Red Neuronal Artificial que se va a tratar.

La **Figura 13** muestra los diagramas de dispersión entre el precio en dólares del inmueble y sus posibles variables de entrada. Esto se realizó con el fin de observar que no es posible realizar una metodología lineal para avaluar departamentos habitacionales en la ciudad de Lima, por ello se optó por la aplicación de la metodología de RNA.

El precio en dólares del departamento tuvo una relación significativa con Distrito, Superficie, Número de habitaciones, Número de baños, Número de Garajes, Piso de Ubicación, Vista al Exterior y Años de antigüedad. Lo cual nos da a entender que todas las variables propuestas guardan una relación significativa con la variable de salida (ver Anexo 6)

Figura 13

Diagrama de dispersión para el precio en dólares del inmueble y las variables de entrada



3.1.3. División del conjunto de datos

En la presente sección para el desarrollo del modelo de RNA consiste en fragmentar el conjunto de datos en un conjunto de entrenamiento y conjunto de prueba, debido a que estamos utilizando cross validation (validación cruzada).

Los resultados de análisis de medida central y medidas de variabilidad del conjunto de entrenamiento y el conjunto de prueba de las variables se muestran en la **Tabla 4** y **Tabla 5**.

Tabla 4

Estadística del conjunto de datos utilizados como entradas de la Red Neuronal Artificial.

Variabes	Media	σ	Mínimo	Máximo
Superficie (m2)				
Entrenamiento	111.846	51.42	24.00	755.00
Prueba	113.564	52.81	30.00	550.00
Número de habitaciones				
Entrenamiento	2.781	0.67	0.00	32.00
Prueba	2.787	0.63	1.00	7.00
Número de baños				
Entrenamiento	2.361	0.85	0.00	6.00
Prueba	2.384	0.86	0.00	8.00
Número de garajes				
Entrenamiento	0.914	0.85	0.00	20.00
Prueba	0.932	0.95	0.00	31.00

Piso de ubicación				
Entrenamiento	3.010	3.256	0.00	22.00
Prueba	2.989	3.207	0.00	20.00
Vista al exterior				
Entrenamiento	0.942	0.248	0.00	11.00
Prueba	0.937	0.244	0.00	1.00
Años de antigüedad				
Entrenamiento	9.999	10.962	0.00	95.00
Prueba	9.937	10.886	0.00	87.00

Tabla 5

Estadística del conjunto de datos utilizados como salida de la Red Neuronal Artificial.

Variables	Media	σ	Mínimo	Máximo
Precio en dólares corrientes				
Entrenamiento	195268.32	150269.88	16871.00	3403129.00
Prueba	198742.11	155527.98	24567.00	3325433.00

En esta investigación se realizó una comparación entre las medias del conjunto de entrenamiento y el conjunto de entrada con la finalidad que estos sean representativos. En las tablas ya mostradas se observa que las medias de ambos conjuntos son similares. Pero es necesario asegurarse de que la selección de los conjuntos se haya realizado de forma correcta,

es decir que haya la misma diversidad de muestras y representatividad en ambas.

En la **Tabla 6** se muestran los resultados de aplicar la prueba U de Mann-Whitney para la igualdad de medias entre el conjunto de entrenamiento y conjunto de prueba para las variables de entrada, donde se muestra que no existe una diferencia significativa entre ambos conjuntos. Probando que el conjunto de entrenamiento y conjunto de prueba para las variables de entrada son representativas.

Tabla 6

Prueba U de Mann-Whitney para la igualdad de medias entre el conjunto de entrenamiento y conjunto de prueba para las variables de entrada

Variables de entrada	W	Sig. (bilateral)	99 % de intervalo de confianza de la diferencia	
			Inferior	Superior
Distrito	30324178	0.14	-0.00005	0.00006
Superficie	30242976	0.09	-2.99993	0.00003
Número de habitaciones	30594381	0.46	-0.00003	0.00003
Número de baños	30424091	0.21	-0.00002	0.00008
Número de garajes	30451259	0.26	-0.00004	0.00002
Piso de ubicación	30781008	0.98	-0.00003	0.00004
Vista al exterior	30947966	0.22	-0.00003	0.00008
Años de antigüedad	30799979	0.97	-0.00005	0.00005

Como se observa en la **Tabla 7**, se muestran los resultados de aplicar la prueba U de Mann-Whitney para la igualdad de medias entre el conjunto de entrenamiento y conjunto de prueba para la variable de salida,

donde se muestra que no existe una diferencia significativa entre ambos conjuntos. Probando que el conjunto de entrenamiento y conjunto de prueba para la variable de salida son representativas.

Tabla 7

Prueba U de Mann-Whitney para la igualdad de medias entre el conjunto de entrenamiento y conjunto de prueba para las variables de salida

Variables de salida	W	Sig. (bilateral)	99 % de intervalo de confianza de la diferencia	
			Inferior	Superior
Precio en dólares Corrientes	30454586	0.2937	-5000.0	2000.0

3.1.4. Desarrollo del modelo de Red Neuronal Artificial para evaluar departamentos habitacionales en la ciudad de Lima.

Al finalizar se propone 1 modelo para evaluar departamentos habitacionales en la ciudad de Lima. El procedimiento consistió en entrenar diferentes modelos para encontrar los hiperparámetros óptimos, realizar predicciones con el mejor modelo y, finalmente, evaluar el rendimiento del modelo comparando los valores reales con los predichos. Se entrenó 50 modelos de donde se seleccionó el que tuvo mejor rendimiento y se le llamó J.A.R.V.I.S. Los modelos utilizan un número diferente de neuronas en la capa de entrada; sus características se verán más adelante.

3.1.4.1. Ajuste de hiperparámetros.

La **Tabla 8** muestra un resumen de los 10 mejores modelos de RNA de un total de 50 modelos entrenados para evaluar departamentos habitacionales en la ciudad de Lima de acuerdo al precio en dólares corrientes, utilizando como variables de entrada Año, Trimestre, Precio en dólares corrientes,

Precio en soles corrientes, Distrito, Superficie, Número de habitaciones, Número de baños, Número de garajes, Piso de ubicación, Vista al exterior, Años de antigüedad,. En la tabla se puede observar que el modelo N.º 36 compuesto por: 8 neuronas en la capa de entrada, 20 neuronas en la primera capa oculta, 19 neuronas en la segunda capa oculta, 19 neuronas en la tercera capa oculta, 19 neuronas en la cuarta capa oculta y 1 neurona en la capa de salida; fue el mejor modelo con un MSE de 2.12E-01. En esta investigación llamó a este modelo como J.A.R.V.I.S.

Tabla 8

Mejores 10 Redes Neuronales Artificiales para evaluar departamentos habitacionales en la ciudad de Lima

Ranking	Modelo	Capas ocultas	Neuronas		Learning rate	MSE
			(a)	(b)		
1	36	4	20	19	3.08E-04	2.12E-01
2	3	2	17	12	8.13E-04	4.99E-02
3	40	1	20	4	4.14E-04	8.93E-02
4	11	1	17	42	1.03E-04	5.03E-02
5	29	6	20	19	8.74E-05	1.37E-01
6	34	6	20	14	3.57E-06	1.76E-01
7	5	1	20	43	1.29E-03	4.99E-02
8	16	1	19	45	7.34E-06	8.70E-02
9	42	1	20	20	3.75E-05	4.61E-02
10	31	1	20	47	4.47E-03	4.77E-02

Nota. Exploración de hiperparámetros por optimización Bayesiana y *Cross Validation* con K = 10 y número de repeticiones igual a 2. (a) = N.º de

neuronas en la primera capa oculta; (b) = N.^o de neuronas en las capas ocultas restantes.

3.1.4.2. Predicciones

En esta parte de la investigación es donde se plasman los resultados de las predicciones de los valores hechos por la RNA J.A.R.V.I.S. Los conjuntos representativos son los mismos para ambos modelos, donde se trató de hacer una comparación entre los datos predichos y los datos reales. Los datos utilizados son los datos de prueba que constan del 20% de nuestro conjunto de datos (3923 muestras).

En el Anexo 8 se muestran gráficos de líneas que también se asemejan los valores reales de los precios en dólares corrientes de los departamentos a los valores predichos por J.A.R.V.I.S.

Se puede apreciar que la RNA con una arquitectura de 8 neuronas en la capa de entrada, 20 neuronas en la primera capa oculta, 19 neuronas en la segunda capa oculta, 19 en la tercera capa oculta, 19 en la cuarta capa oculta y 1 neurona en la capa de salida tiene un buen rendimiento. Esto se debe a que los valores predichos son similares a la variación de los ejemplos reales.

La **Figura 14** y **Figura 15** la muestran el error absoluto medio (MAE) y el error cuadrático medio (MSE) respectivamente, donde se observa que la red a predicho de manera óptima.

Figura 14

Error Absoluto Medio (MAE) en las pérdidas de entrenamiento y validación.

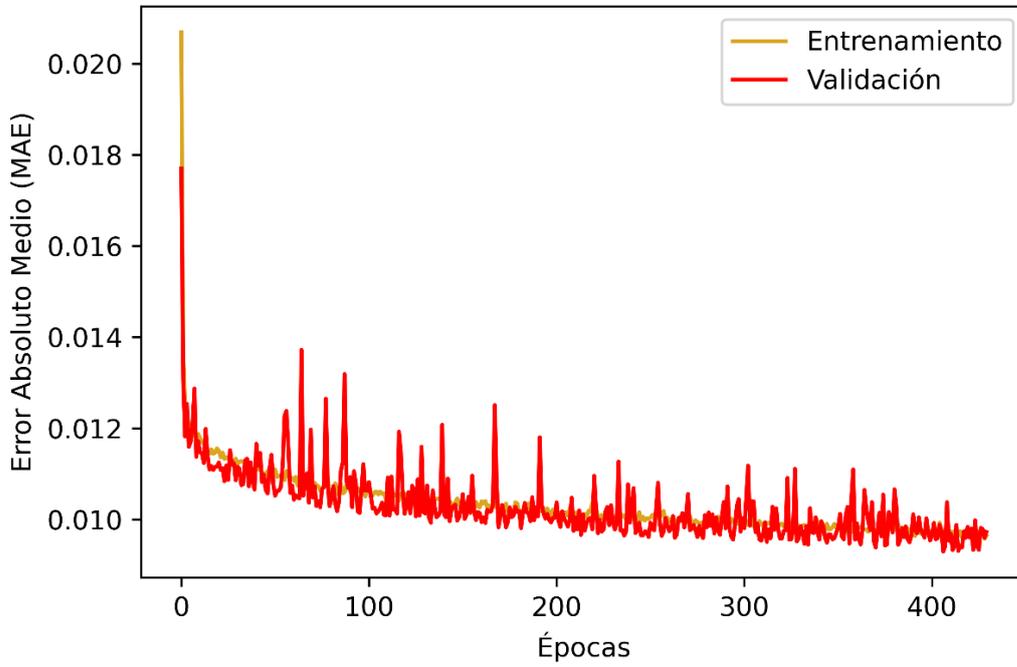
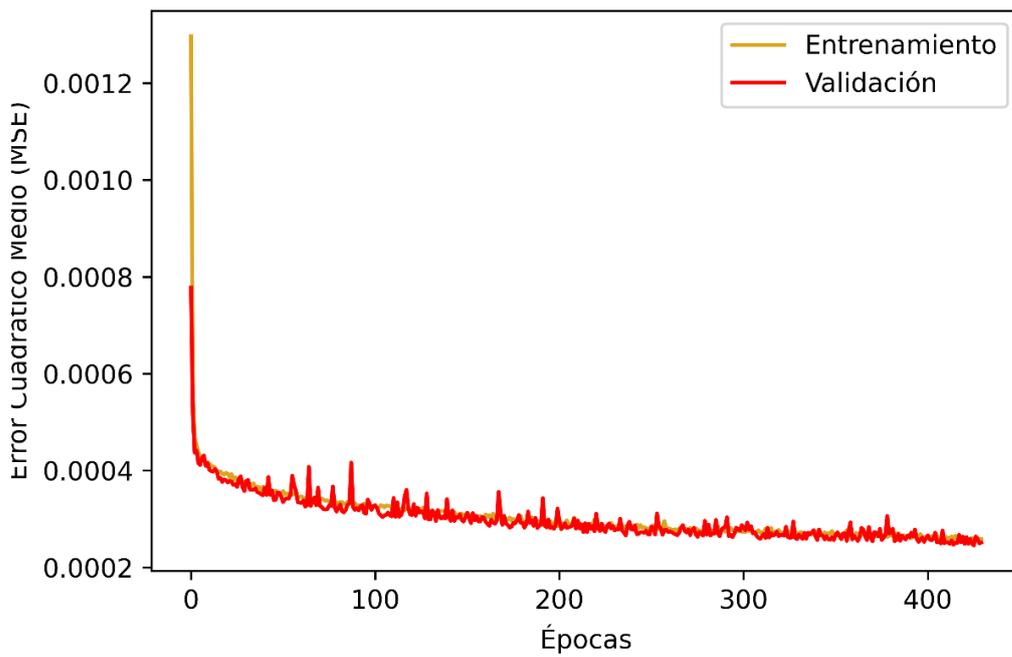


Figura 15

Error Cuadrático Medio (MSE) en las pérdidas de entrenamiento y validación.



3.1.5. Mejorando el modelo J.A.R.V.I.S.

A partir de la arquitectura obtenida de la optimización bayesiana, se optó por mejorar el modelo, con la finalidad de obtener una mayor precisión, proponiendo 4 modelos de RNA. Estos modelos tendrán la arquitectura ya definida.

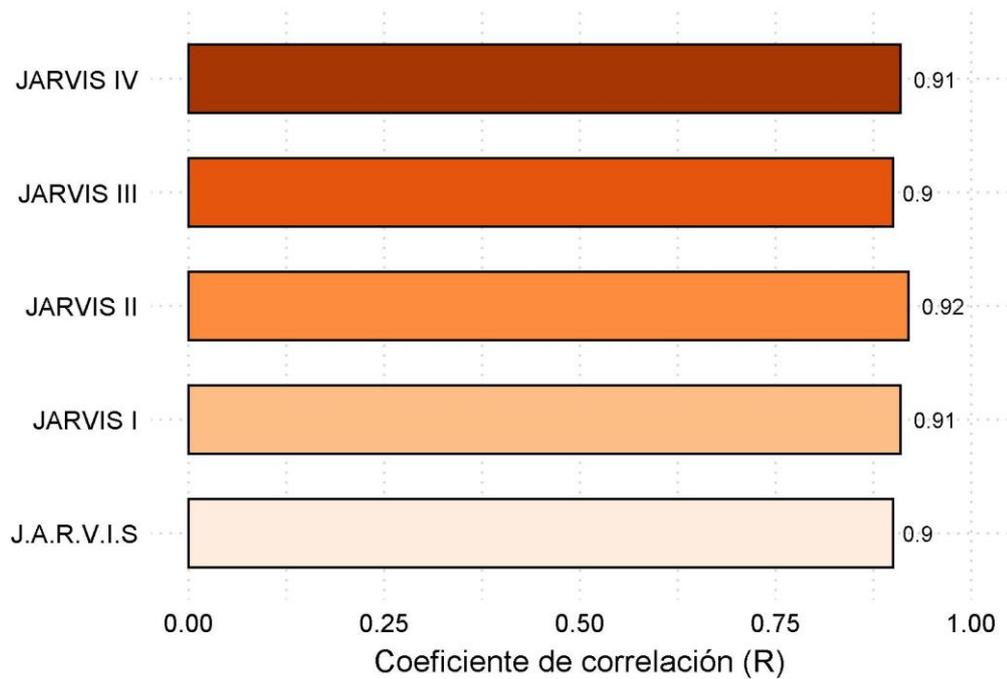
Se ha propuesto 4 modelos de RNA cuya arquitectura se seleccionó con ayuda de la optimización bayesiana. Se puede destacar el JARVIS II para avaluar un departamento habitacional en la ciudad de Lima.

La **Figura 16** deja visualizar una pequeña pero importante mejora a el modelo definido con la optimización bayesiana J.A.R.V.I.S.

Entre los modelos desarrollados los que obtuvieron mejor rendimiento son el JARVIS I y el JARVIS II. En el gráfico de barras se puede ver que el JARVIS II tuvo un mejor rendimiento (R igual a 0.91743) a comparación del JARVIS I (R^2 igual a 0.91159), inclusive, es mejor que el J.A.R.V.I.S (0.90072). Estos resultados muestran que se llegó a mejorar el modelo J.A.R.V.I.S, propuesto inicialmente.

Figura 16

Rendimiento de los modelos de Redes Neuronales Artificiales desarrollados para mejorar el J.A.R.V.I.S.



El criterio para mejorar el modelo J.A.R.V.I.S fue de aumentar el número de neuronas por capa oculta a 1.5, 2 y 2.5 veces las del modelo inicial, pero conservando los mismos hiperparámetros del modelo inicial. El modelo JARVIS I tuvo el mismo número de neuronas que el J.A.R.V.I.S con la diferencia de que fue entrenado durante 3000 *epochs*. El JARVIS II, JARVIS III Y JARVIS IV también fueron entrenados durante 3000 *epochs*. JARVIS I utilizó 19 neuronas en cada capa oculta. JARVIS II utilizó 30 neuronas en cada capa oculta. JARVIS III utilizó 40 neuronas en cada capa oculta. JARVIS IV utilizó 50 neuronas en cada capa oculta.

La **Figura 17** deja visualizar el rendimiento de los modelos conforme se aumenta el número de neuronas ocultas. En la parte izquierda los modelos J.A.R.V.I.S y JARVIS I a pesar de que utilizan el mismo número de neuronas en las capas ocultas, se observa una mejora en cuanto a la precisión con tan sólo aumentar el número de *epochs*. También se logra apreciar que conforme se aumenta el número de neuronas en las capas

ocultas estas mejoran la precisión como es el caso de JARVIS II. A pesar de eso se logra apreciar que JARVIS III tiende a decrecer y JARVIS IV aumenta, pero no tanto como JARVIS II, esto es debido a un fenómeno llamado overfitting que llega a ocurrir cuando se sobre entrena un modelo.

Figura 17

Rendimientos de los modelos desarrollados para mejorar el J.A.R.V.I.S según el número de neuronas ocultas.

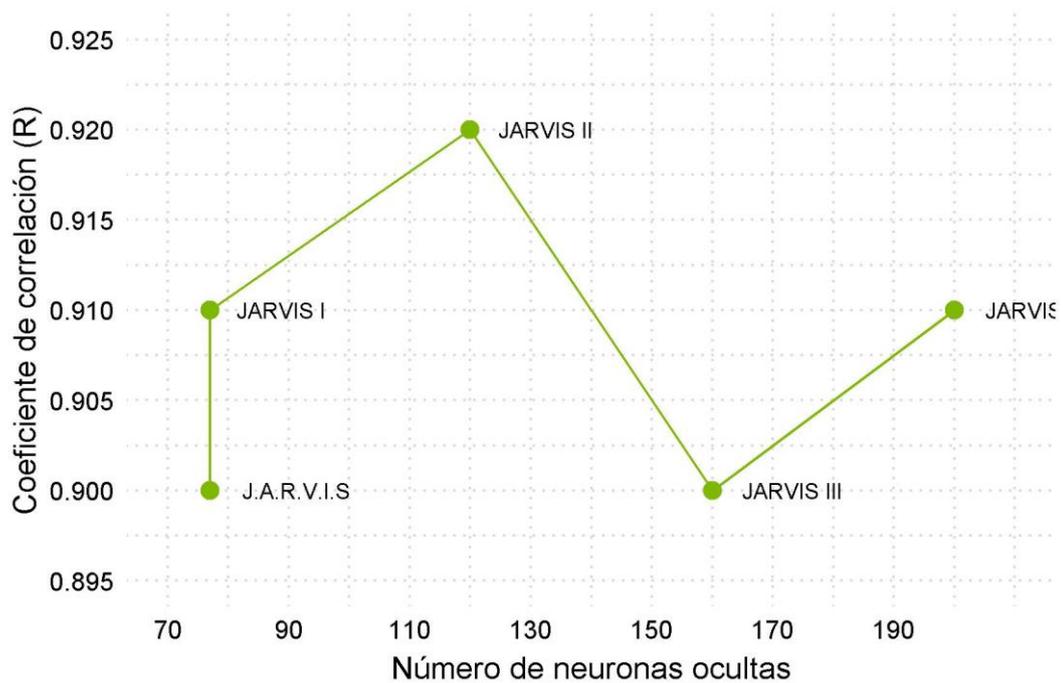
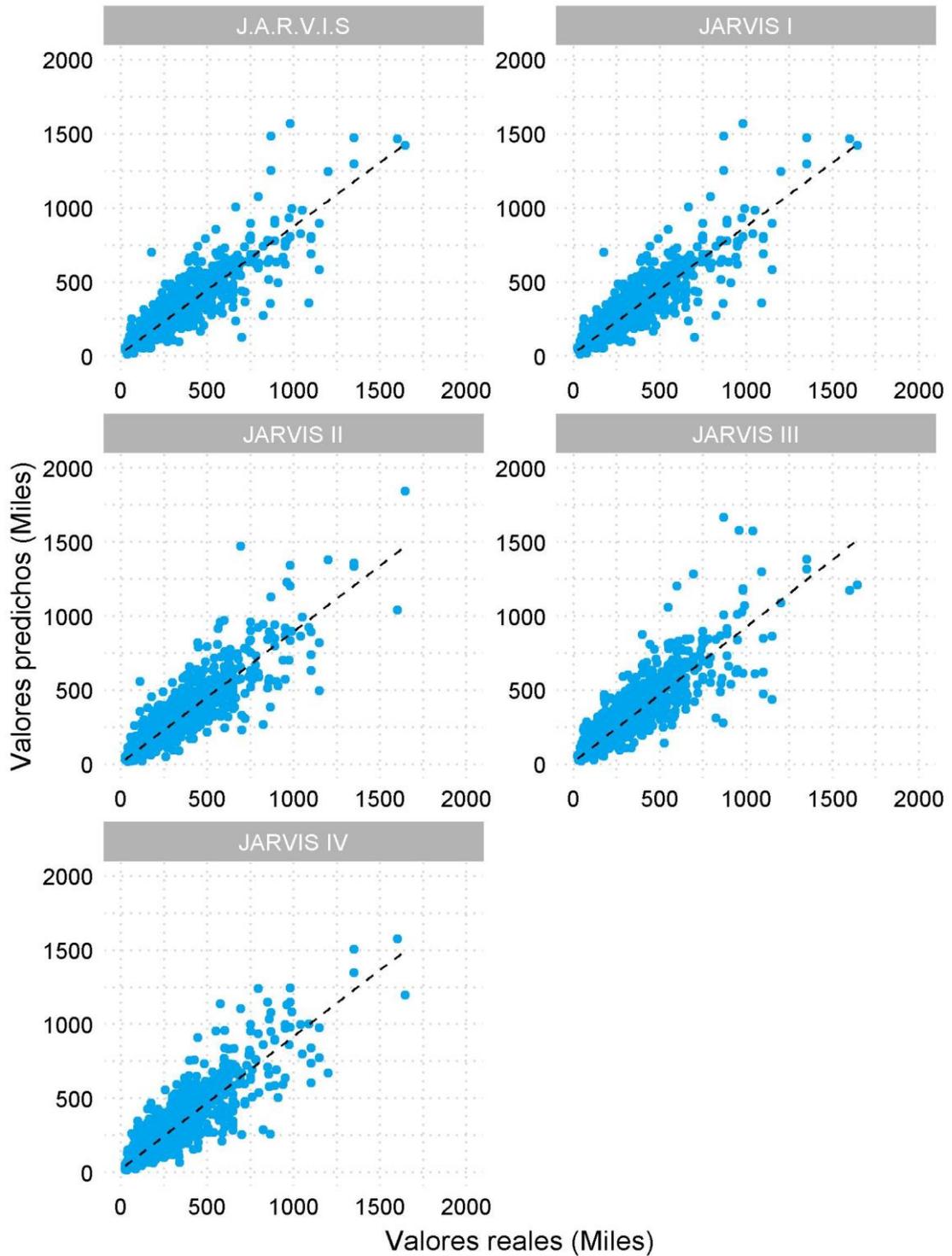


Figura 18

Gráfica de la correlación de las predicciones de los modelos propuestos.



3.1.5.1. Comparando los valores predichos y los valores reales

Tomando como modelo final a JARVIS II como el modelo que logró una mejor precisión se puede observar que logró obtener un R de 0.91743 a comparación del modelo inicial J.A.R.V.I.S que tenía un R de 0.90072. Con estos resultados se puede saber que el modelo JARVIS II fue superior a el modelo inicial a la hora de realizar predicciones con datos que el modelo nunca antes había visto.

3.1.6. Resumen de los modelos de Redes Neuronales Artificiales desarrollados

Se presenta los resultados del comportamiento de cada modelo de RNA para hacer las predicciones. Debido a esto la **Tabla 9** muestra el error en las fases de entrenamiento, validación y prueba cometidos por los modelos RNA. Los resultados para la función de pérdida MSE y la métrica MAE para la fase de prueba se calcularon utilizando 3923 ejemplos que los modelos de RNA no conocían.

Esta sección presenta los resultados del comportamiento de cada modelo de RNA para hacer predicciones. Por ello, la **Tabla 9** muestra el error cometido por los modelos de RNA durante la etapa de entrenamiento, etapa de validación y etapa de prueba. Cabe recordar que los resultados para la función de pérdida MSE y la métrica MAE para la etapa de prueba se calcularon con 3923 muestras que los modelos de RNA no habían visto antes.

La **Tabla 9** muestran 5 modelos de RNA desarrollados para evaluar departamentos habitacionales en la ciudad de Lima; estos son el J.A.R.V.I.S, JARVIS I, JARVIS II, JARVIS III Y JARVIS IV. Datos que se pueden visualizar en el Anexo 7.

La **Tabla 10** muestra la lista completa de hiperparámetros utilizados en los modelos de RNA con los que se hicieron las predicciones, con los cuales se obtuvieron los resultados ya mostrados. Cabe recordar que los

hiperparámetros mostrados fueron encontrados con la optimización bayesiana con el cross validation.

Tabla 9

Estructura y rendimiento de los modelos de redes neuronales artificiales desarrollados.

Redes Neuronales Artificiales	Redes Neuronales Artificiales	Neuronas ocultas	MSE			MAE			R
			E	V	P	E	V	P	
	J.A.R.V.I.S.	77	0.000305	0.000300	0.000401	0.010309	0.010281	0.010747	0.90072
Avaluar departamentos habitacionales	JARVIS I	77	0.000208	0.000204	0.000360	0.008816	0.008785	0.009678	0.91159
	JARVIS II	120	0.000162	0.000159	0.000345	0.008104	0.008063	0.010172	0.91743
	JARVIS III	160	0.000128	0.000125	0.000414	0.007321	0.007268	0.010387	0.89899
	JARVIS IV	200	0.000109	0.000106	0.000367	0.000367	0.006680	0.010460	0.90895

Nota. E = entrenamiento; V = validación; P = prueba; R = coeficiente de correlación.

Tabla 10

Arquitectura e Hiperparámetros de los mejores modelos aplicando optimización bayesiana.

Hiperparámetros	J.A.R.V.I.S.	JARVIS I	JARVIS II	JARVIS III	JARVIS IV
Neuronas de entrada	8	8	8	8	8
Neuronas de salida	1	1	1	1	1
Capas ocultas	3	3	3	3	3
Neuronas ocultas	77	77	120	160	200
Arquitectura	8-19-19-19-1	8-19-19-19-1	8-30-30-30-1	8-40-40-40-1	8-50-50-50-1
Función de activación	ReLU	ReLU	ReLU	ReLU	ReLU
Tasa de aprendizaje	0.0003	0.0003	0.0003	0.0003	0.0003
Optimizador	Adam	Adam	Adam	Adam	Adam
Número de épocas	430	3000	3000	3000	3000
Tamaño de lote	32	32	32	32	32

Dropout	0.0	0.0	0.0	0.0	0.0
MSE	0.000401	0.000360	0.000345	0.000414	0.000367

3.2. Discusión de resultados

El propósito de este estudio era evaluar departamentos habitacionales en la ciudad de Lima aplicando un modelo de RNA. Para ellos se realizó un modelo inicial llamado J.A.R.V.I.S, que a partir del cual se mejoró con otros 4 modelos adicionales llamados JARVIS I, JARVIS II, JARVIS III y JARVIS IV. En esta parte de la investigación se discuten los hallazgos en relación a la literatura sobre el desarrollo de modelos de RNA. Además, se discuten la correlación entre las variables de entrada y la variable de salida, así como la determinación de los hiperparámetros óptimos que se utilizan en el modelo final.

3.2.1. Aplicación de modelos de redes neuronales artificiales para evaluar departamentos habitacionales en la ciudad de Lima

La hipótesis planteada en la investigación pudo llegar a cumplirse según los resultados ya expuestos, pues se ha sobrepasado el límite de confiabilidad del 80%. Se puede concluir que en esta investigación la aplicación de modelos de RNA permitieron evaluar departamentos habitacionales en la ciudad de Lima con los datos desgregados de venta de departamentos en la ciudad de Lima brindados por el BCRP. ((Çelik, 2021; Chaphalkar & Sandbhor, 2019; Epley, 2017; Horvath et al., 2021; Kamire et al., 2021; Lee & Jae, 2021; Núñez Tabales et al., 2017; Preciado Carrillo, 2019; Sridhar & Sathyanathan, 2022; Štubňová et al., 2020; B. Zhou & Lu, 2021)).

El parámetro para demostrar la posibilidad de predecir el precio de un departamento en la ciudad de Lima aplicando la metodología de REDES NEURONALES ARTIFICIALES se basó en el coeficiente de determinación (R). El resultado del modelo final (JARVIS II) para evaluar departamentos habitacionales en la ciudad de Lima tiene un R de 0.91743. fueron similares a comparación al R de 0.9231 en zonas residenciales y el R de 0.9156 de Zonas industriales de otra investigación. (Çelik, 2021)

3.2.2. Avaluar departamentos habitacionales en la ciudad de Lima y los IVDH utilizadas en la red neuronal artificial

La valuación de departamentos habitacionales en la ciudad de Lima y los IVDH utilizados en el modelo de RNA obtuvieron valores satisfactorios debido a que los indicadores que se usaron tuvieron una relación significativa con el precio del inmueble, estos indicadores fueron: Distrito, Superficie, Número de habitaciones, Número de baños, Número de Garajes, Piso de Ubicación, Vista al Exterior y Años de antigüedad. Datos que se ven en otras investigaciones a nivel de estudio de arte. (Epley, 2017; Horvath et al., 2021; Lee & Jae, 2021; Sridhar & Sathyanathan, 2022)

A pesar de que se obtuvieron valores satisfactorios también es necesario notar que la base de datos disgregada brindados por el BCRP carecen de diversidad de indicadores, los cuales hubieran influido en gran proporción en la precisión del modelo, esto se respalda en la investigación de Del Giudice et al. (2017), donde nos comenta que el mayor problema para realizar investigación de valoración de bienes raíces es debido a la poca accesibilidad de la base de datos.

3.2.3. El número de neuronas y la precisión

En la etapa del mejoramiento del modelo de RNA se cambió el número de neuronas en cada capa oculta. Se llegó a mejorar la precisión de la predicción del modelo aumentando el número de neuronas en cada capa oculta (como se puede apreciar en los modelos JARVIS I y JARVIS II), sin embargo, también se demostró que el exceso de neuronas en la capa oculta origina que el modelo deje de mejorar a comparación de sus predecesores (JARVIS III y JARVIS IV).

Velasco *et al.* (2019) en su investigación “la fórmula de Stathakis” que sirve para determinar el número aproximado de neuronas en cada capa oculta. Nos recomienda utilizar un modelo inicial que se va a obtener con la fórmula, pero los resultados de su investigación concluyen que se obtendrán mejores modelos añadiendo neuronas en las capas ocultas. Esta

investigación apoya nuestros resultados obtenidos, debido a que se lograron obtener los hiperparámetros con ayuda de la optimización bayesiana y se encontró el mejor modelo inicial, a partir de allí se optó por mejorar el modelo añadiendo neuronas en las capas ocultas hasta obtener un modelo final óptimo.

3.2.4. El tiempo de entrenamiento afecta en la precisión

En esta investigación se demostró que aumentar el número de tiempo de entrenamiento en cada modelo de RNA a desarrollar logró mejorar la precisión. Sin embargo, teniendo en cuenta que esta metodología se aplica para computación blanda, el hecho de aumentar tiempo de entrenamiento puede generar un exceso el cual provoca un mal comportamiento al momento de predecir (Underfitting y overfitting). (Diaz et al., 2017; Kim, 2019)

Teniendo en cuenta lo ya mencionado se puede decir que no existe una fórmula mágica que permita obtener el número de neuronas en cada capa oculta o el número de epochs óptimo(Casas Roma et al., 2019) .De acuerdo a esto el desarrollo de modelos de RNA consiste en un procedimiento de prueba y error, es en esta parte donde se observa el comportamiento de los modelos a la hora de predecir.

CAPITULO IV:
CONCLUSIONES

IV. CONCLUSIONES

4.1. Conclusiones

La finalidad de este trabajo es evaluar de forma precisa y en menor tiempo departamentos habitacionales en la ciudad de Lima por medio de desarrollar un modelo de red neuronal artificial, a partir de los datos desgregados de venta del BCRP. La conclusión de la investigación de acuerdo a los resultados obtenidos en cada fase de la investigación y tomando en cuenta los objetivos, se describirán de la siguiente manera.

- Para determinar los hiperparámetros óptimos de cada modelo de red neuronal artificial se realizó mediante lenguaje de programación a partir de la programación Python en el software Spyder. Lo más importante en esta parte de la investigación fue encontrar el modelo de red neuronal artificial con mayor capacidad para generalizar. El MSE fue la herramienta que se utilizó para identificar los hiperparámetros óptimos, pues esta muestra la reducción del error y el comportamiento entre las muestras reales y las predichas por el modelo de red neuronal artificial. El mayor obstáculo para identificar los hiperparámetros fue el tipo de técnica a utilizar en el código que se iba a procesar por el software Spyder.
- Se realizaron las pruebas empleando el conjunto de prueba utilizando modelos de red neuronal artificial con los hiperparámetros óptimos. Lo más resaltante dentro de este paso fue hallar el valor de MAE al evaluar departamentos habitacionales en la ciudad de Lima, con un valor de 0.010172, dando a demostrar que se pueden aplicar redes neuronales artificiales para evaluar departamentos habitacionales en la ciudad de Lima utilizando la base de datos desgregada de venta de departamentos en la ciudad de Lima brindada por el BCRP. Lo que más ayudó al momento de la verificación de las predicciones fue haber realizado la partición del conjunto de entrenamiento y el conjunto de prueba, pues se pudo determinar el buen rendimiento del modelo con ejemplos que nunca había visto.

- Se demostró que los modelos de redes neuronales artificiales propuestos aprendieron de forma correcta haciendo una comparación entre los valores reales con los valores predichos por el modelo de red neuronal artificial. Esta comparación sirvió para obtener el valor de R, el cual fue 0.91743., este valor permite demostrar la gran capacidad de los modelos de redes neuronales artificiales para evaluar departamentos habitacionales en la ciudad de Lima en base a los datos brindados por el BCRP. Estos modelos pueden utilizarse como una herramienta para evaluar departamentos habitacionales en la ciudad de Lima tomando en cuenta tan solo las variables de entrada del modelo presentado de una manera rápida.

4.2. Recomendaciones

Los modelos de redes neuronales artificiales propuestos en esta investigación para evaluar departamentos habitacionales en la ciudad de Lima a pesar de que han tenido un número considerable de datos para procesar y por consecuencia se logró obtener resultados satisfactorios, es necesario que esta base de datos brinde variables más específicas para poder mejorar la precisión. A modo de recomendación se hace hincapié en tomar en cuenta las normativas internacionales y nacionales de tasación vigentes.

Se recomienda que este mismo formato se exija a las diferentes inmobiliarias, las cuales a su vez no comparten información al público con fines de investigación. Estos datos serían de gran ayuda a la comunidad científica para desarrollar mejores modelos apuntados a sectores tanto residenciales como industriales.

En esta investigación se ha demostrado que la aplicación de redes neuronales artificiales presenta un gran rendimiento en avaluar departamentos. Es importante ampliar este conocimiento y aplicarlo en distintas ramas de la ingeniería civil, donde no se han encontrado investigaciones con respecto a este objetivo de investigación a nivel nacional.

REFERENCIAS

- Abadi Abbo, I., Martín Domínguez, F., Capasso Mantini, A., & Rojas Rangel, G. (2009). *Instrumento de Evaluación de Viviendas de Interés Social*. Universidad Central de Venezuela.
- Amirabadi, M. A., Kahaei, M. H., & Nezamalhosseini, S. A. (2020). Novel suboptimal approaches for hyperparameter tuning of deep neural network [under the shelf of optical communication]. *Physical Communication*, 41, Article 101057. <https://doi.org/10.1016/j.phycom.2020.101057>
- Antoniucci, V., & Marella, G. (2017). Immigrants and the city: The relevance of immigration on housing price gradient. *Buildings*, 7(4), 1-14. <https://doi.org/10.3390/buildings7040091>
- Balles, L., Romero, J., & Hennig, P. (2017, diciembre). Coupling adaptive batch sizes with learning rates. *Uncertainty in Artificial Intelligence - Proceedings of the 33rd Conference, UAI 2017*.
- Belmont, I. (1979). *Principios éticos y directrices para la protección de sujetos humanos de investigación*.
- Benítez, R. (2014). *Inteligencia artificial avanzada*. Editorial UOC.
- Casas Roma, J., Bosch Rue, A., & Lozano Bagen, T. (2019). *Deep learning: principios y fundamentos*. Editorial UOC.
- Çelik, K. (2021). Gayrimenkul değerlemesinde geleneksel yöntemler. *Gümüşhane Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 12, 246-259. <https://doi.org/10.17714/gumusfenbil.925678>
- Ceular, N., Caridad, J. M., & Núñez, J. M. (2008). Metodología de Precios Hedónicos vs. Redes Neuronales Artificiales como alteRedes neuronales artificialestivas a la valoración de Inmuebles. Un caso real. *Catastro*, Abril, 27-42.

http://files.eltasadorpampeano.webnode.com/200000344-d8d60dac96/METODO_HEDONICO_VS.pdf

- Chaphalkar, N. B., & Sandbhor, S. (2019). Effect of Training Sample and Network Characteristics in Neural Network-Based Real Property Value Prediction. *Springer*. https://doi.org/https://doi.org/10.1007/978-981-13-1610-4_31
- Chollet, F. (2018). Deep Learning with Python. En *MANNING - Shelter Island*. <https://doi.org/10.23919/ICIF.2018.8455530>
- Colegio De Ingenieros Del Perú. (2018). *REGLAMENTO DE LOS CENTROS DE PERITAJE DEL COLEGIO DE INGENIEROS DEL PERÚ*.
- Cubeddu, L. M., Tovar Mora, C. E., & Tsounta, E. (2012). Latin America: Vulnerabilities Under Construction? *IMF Working Papers*, 12(193), i. <https://doi.org/10.5089/9781475505429.001>
- Del Giudice, V., De Paola, P., Forte, F., & Manganeli, B. (2017). Real estate appraisals with Bayesian approach and Markov Chain Hybrid Monte Carlo Method: An application to a central urban area of Naples. *Sustainability (Switzerland)*, 9(11). <https://doi.org/10.3390/su9112138>
- Diaz, G. I., Fokoue-Nkoutche, A., Nannicini, G., & Samulowitz, H. (2017). An effective algorithm for hyperparameter optimization of neural networks. *IBM JouRedes neuronales artificialesl of Research and Development*, 61(4/5). <https://doi.org/10.1147/JRD.2017.2709578>
- Dinamarca, A. (2018). *Aprendizaje y Análisis de Redes Neuronales Artificiales Profundas*. Tesina de grado, Universidad Nacional de Cuyo.
- Epley, D. (2017). THE NEED TO REFERENCE AUTOMATIC VALUATION MODELS TO THE VALUATION. *JOUREDES NEURONALES ARTIFICIALESL OF REAL ESTATE LITERATURE*, 25(1), 237-252.
- Escamilla-García, A., Soto-Zarazúa, G. M., Toledano-Ayala, M., Rivas-Araiza, E., & Gastélum-Barrios, A. (2020). Applications of Artificial Neural Networks in Greenhouse Technology and Overview for Smart Agriculture Development. *Applied Sciences*, 10(11), Article 3835.

<https://doi.org/10.3390/app10113835>

Fondo Mivivienda. (2018). ESTUDIO DE DEMANDA DE VIVIENDA A NIVEL NACIONAL. *Ministerio de Vivienda, Construcción y Saneamiento*.

Gestión Noticias. (2018). ¿Cómo se evalúa la compra de un inmueble? | TU-DINERO | GESTIÓN. *GESTIÓN NOTICIAS*. <https://gestion.pe/tu-dinero/evalua-compra-inmueble-232662-noticia/>

Gobierno del Perú. (2016). Publicación del proyecto de Resolución Ministerial que aprueba el Reglamento Nacional de Tasaciones. *EL PERUANO*.

Gobierno del Perú. (2017). Ley De Tributación Municipal. *Editora Peru*, 53(9), 21-25. <http://www.elsevier.com/locate/scp>

Gobierno del Perú. (2020). *Reglamento Nacional de Edificaciones*.

Godfrey, L. B., & Gashler, M. S. (2015). A continuum among logarithmic, linear, and exponential functions, and its potential to improve generalization in neural networks. *IC3K 2015 - Proceedings of the 7th InteRedes neuronales artificiales tional Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 1, 481-486. <https://doi.org/10.5220/0005635804810486>

Gómez, A. (2019, junio 7). *Evolución del sector inmobiliario en Lima Metropolitana*.

<http://seccionnoticias.net.pe/index.php/2019/05/29/evolucion-del-sector-inmobiliario-en-lima-metropolitana/>

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*.

Guyton y Hall. (2016). Tratado de Fisiología Médica. En *ELSEVIER SAUNDERS (DECIMOTERC)*. https://doi.org/10.1007/978-88-470-1064-2_13

Hernández-Sampieri, R., & Mendoza Torres, C. P. (2018). *Metodología de la Investigación: las rutas cuantitativas, cualitativa y mixta* (1.ª ed.). McGraw Hill.

Horvath, S., Soot, M., Zaddach, S., Neuner, H., & Weitkamp, A. (2021).

- Deriving adequate sample sizes for ANN-based modelling of real estate valuation tasks by complexity analysis. *Land Use Policy*, 107(October 2020), 105475. <https://doi.org/10.1016/j.landusepol.2021.105475>
- Instituto Nacional de Estadística e Informática. (2020). *Informe Técnico N° 03 -Indicador de la Actividad Productiva Departamental: Segundo Trimestre 2020*.
- InteRedes neuronales artificialescional Valuation Standards Council. (2020). *Normas InteRedes neuronales artificialescionales de Valuacion (IVS) (Número 21)*.
- Jassam Mohammed, A., Hameed Arif, M., & Adil Ali, A. (2020). A multilayer perceptron artificial neural network approach for improving the accuracy of intrusion detection systems. *IAES InteRedes neuronales artificialescional JouRedes neuronales artificialesl of Artificial Intelligence (IJ-AI)*, 9(4), 609-615. <https://doi.org/10.11591/ijai.v9.i4.pp609-615>
- Kamire, A., Chaphalkar, N., & Sandbhor, S. (2021). *Real Property Value Prediction Capability Using Fuzzy Logic and ANFIS. Icsmdi*.
- Kim, M. (2019). Supervised learning-based DDoS attacks detection: Tuning hyperparameters. *ETRI JouRedes neuronales artificialesl*, 41(5), 560-573. <https://doi.org/10.4218/etrij.2019-0156>
- Lathuiliere, S., Mesejo, P., Alameda-Pineda, X., & Horaud, R. (2020). A Comprehensive Analysis of Deep Regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(9), 2065-2081. <https://doi.org/10.1109/TPAMI.2019.2910523>
- Lee, J., & Jae, P. R. (2021). Prediction of Seoul House Price Index Using Artificial Neural Network. *Revista de la Sociedad de Cooperación Industrial-Académica de Corea*, 22(4), 228-234.
- Malgorzata, R.-B., Sabina, Ż., Marek, W., Richard, B., Richard, G., & Maurizio, D. (2022). La aceptación inteRedes neuronales artificialescional del uso de herramientas modeRedes neuronales artificialesl automatizadas es imprescindible para el desarrollo

- sostenible del mercado inmobiliario. *Land Use Policy*, 113. <https://doi.org/https://doi.org/10.1016/j.landusepol.2021.105876>.
- Mora-Esperanza, J. (2004). La inteligencia artificial aplicada a la valoración de inmuebles: un ejemplo para valorar Madrid. *CT: Catastro*, 51-68. <http://dialnet.unirioja.es/servlet/articulo?codigo=892741&orden=66297&info=link>
- Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python: a guide for data scientists*. O'Reilly Media, Inc.
- Müller, A. C., & Guido, S. (2020). Introduction to Machine Learning with Python. En D. Schanafelt (Ed.), *O'REILLY*.
- Núñez Tabales, J. M., Rey Carmona, F. J., & Caridad y Ocerin, J. M. (2017). Redes neuronales (RN) aplicadas a la valoración de locales comerciales. *Informes de la Construcción*, 69(545). <https://doi.org/10.3989/ic.15.053>
- Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2018). *Activation Functions: Comparison of trends in Practice and Research for Deep Learning*.
- Olaechea-cunza, L. (2019). ANÁLISIS COMPARATIVO DE LOS ASPECTOS INFLUYENTES EN LA TASACIÓN DE INMUEBLES. *Repositorio Institucional PIRHUA*.
- Orrego, F. (2014). Precios de viviendas en Lima. *Documentos de Trabajo del BCRP*, 59(Mayo 2014), 1-18. <http://www.bcrp.gob.pe/docs/Publicaciones/Documentos-de-Trabajo/2014/documento-de-trabajo-08-2014.pdf>
- Preciado Carrillo, J. C. (2015). Valoración De Inmuebles Urbanos : Comparativa Modelo Regresión Multivariable Versus Redes Neuronales Artificiales Para La Ciudad de Morelia, Michoacán, Mexico. En *UNIVERSIDAD POLITECNICA DE VALENCIA*. UNIVERSIDAD POLITECNICA DE VALENCIA.
- Preciado Carrillo, J. C. (2019). Redes neuronales artificiales , una aplicación al mercado inmobiliario habitacional de segmento medio de la ciudad de Morelia Michoacán , México. *Revista Nicolaita de*

Estudios Económicos, XIV(1), 49-68.

- Properati. (2019). *El evaluador de Properati - Noticias y Análisis inmobiliarios en Perú*. PROPERATI BLOG. <https://blog.properati.com.pe/el-avaluador-de-properati/>
- Purohit, Y., Sarkar, D., & Pandya, J. V. (2021). *Study of Efficacy of Forecasting Models for Residential Real Estate Sale Prices*. March. <https://www.researchgate.net/publication/350974893>
- Rey Carmona, F., & Núñez Tabales, J. (2017). AltesRedes neuronales artificiales para la valoración de inmuebles urbanos. *Revista Publicando*, 4(11(1)), 3-19.
- Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J. (2019). Deep learning-based electroencephalography analysis: a systematic review. *JouRedes neuronales artificiales / of Neural Engineering*, 16(5). <https://doi.org/10.1088/1741-2552/ab260c>
- Saikia, P., Baruah, R. D., Singh, S. K., & Chaudhuri, P. K. (2020). Artificial Neural Networks in the domain of reservoir characterization: A review from shallow to deep models. *Computers and Geosciences*, 135. <https://doi.org/10.1016/j.cageo.2019.104357>
- Serrano, A., Soria, E., & Martín, J. (2010). *REDES NEURONALES ARTIFICIALES*. Universidad de Valencia.
- Sharma, S., Sharma, S., & Athaiya, A. (2020). ACTIVATION FUNCTIONS IN NEURAL NETWORKS. *InteRedes neuronales artificiales / of Engineering Applied Sciences and Technology*, 04(12), 310-316. <https://doi.org/10.33564/IJEAST.2020.v04i12.054>
- SINGH, G., & SINGH, A. K. (2016). A study on precursors leading to geomagnetic storms using artificial neural network. *JouRedes neuronales artificiales / of Earth System Science*, 125(5), 899-908. <https://doi.org/10.1007/s12040-016-0702-1>
- Snell, R. (2014). NEURONATOMÍA CLÍNICA. En *Wolters Kluwer* (OCTAVA EDI). Wolters Kluwer.
- Sridhar, M., & Sathyanathan, R. (2022). Modelado de precios de casas

- residenciales usando análisis de regresión. *Avances en la Gestión de la Construcción*, 191, 321-331.
https://doi.org/https://doi.org/10.1007/978-981-16-5839-6_28
- Štubňová, M., Urbaníková, M., Hudáková, J., & Papcunová, V. (2020). Estimation of residential property market price: comparison of artificial neural networks and hedonic pricing model. *Emerging Science JouRedes neuronales artificiales*, 4(6), 530-538.
<https://doi.org/10.28991/esj-2020-01250>
- Szandała, T. (2021). Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks. En *Studies in Computational Intelligence* (Vol. 903, pp. 203-224).
https://doi.org/10.1007/978-981-15-5495-7_11
- The Appraisal Foundation. (2019). *Uniform Standards of Professional Appraisal Practice (USPAP)*.
<https://doi.org/10.1002/9781119200130.ch10>
- Tinsa. (2016). *Valor de Mercado o venal de un inmueble (VM) - Tinsa*.
<https://www.tinsa.es/glosario/valor-de-mercado-valor-venal-inmueble/#:~:text=El valor de mercado o el valor venal de un,mercado%2C que las condiciones del>
- Velasco, L. C. P., P., R., Shahin, M., F., B., & C., J. (2019). Performance Analysis of Multilayer Perceptron Neural Network Models in Week-Ahead Rainfall Forecasting. *InteRedes neuronales artificialesl of Advanced Computer Science and Applications*, 10(3), 578-588.
<https://doi.org/10.14569/IJACSA.2019.0100374>
- Vijayashanthar, V., Qiao, J., Zhu, Z., Entwistle, P., & Yu, G. (2018). Modeling Fecal Indicator Bacteria in Urban Waterways Using Artificial Neural Networks. *JouRedes neuronales artificialesl of Environmental Engineering*, 144(6). [https://doi.org/10.1061/\(ASCE\)EE.1943-7870.0001377](https://doi.org/10.1061/(ASCE)EE.1943-7870.0001377)
- Wang, Yingying, Li, Y., Song, Y., & Rong, X. (2020). The Influence of the Activation Function in a Convolution Neural Network Model of Facial

- Expression Recognition. *Applied Sciences*, 10(5).
<https://doi.org/10.3390/app10051897>
- Wang, Yixiang, Liu, J., Mistic, J., Mistic, V. B., Lv, S., & Chang, X. (2019). Assessing Optimizer Impact on DNN Model Sensitivity to Adversarial Examples. *IEEE Access*, 7, 152766-152776.
<https://doi.org/10.1109/ACCESS.2019.2948658>
- Yi, D., Ahn, J., & Ji, S. (2020). An Effective Optimization Method for Machine Learning Based on ADAM. *Applied Sciences*, 10(3).
<https://doi.org/10.3390/app10031073>
- Zhang, X., Chen, X., Wang, J., Zhan, Z., & Li, J. (2018). Verifiable privacy-preserving single-layer perceptron training scheme in cloud computing. *Soft Computing*, 22(23), 7719-7732.
<https://doi.org/10.1007/s00500-018-3233-7>
- Zhou, B., & Lu, N. (2021). Correlation between Atterberg Limits and Soil Adsorptive Water. *JouRedes neuronales artificiales/ of Geotechnical and Geoenvironmental Engineering*, 147(2).
[https://doi.org/10.1061/\(ASCE\)GT.1943-5606.0002463](https://doi.org/10.1061/(ASCE)GT.1943-5606.0002463)
- Zhou, G., Ji, Y., Chen, X., & Zhang, F. (2018). Artificial neural networks and the mass appraisal of real estate. *InteRedes neuronales artificiales/estional JouRedes neuronales artificiales/ of Online Engineering*, 14(3), 180-187. <https://doi.org/10.3991/ijoe.v14i03.8420>

ANEXOS

Anexo 1. Formato de la ficha de registro de datos brindados por el banco central de reserva del Perú

La base de datos brindada por el BCRP se registró en la siguiente ficha:

Año	Trimestre	Precio en dólares corrientes	Precio en soles corrientes	Precio en soles constantes de 2009	Distrito	Superficie	Número de habitaciones	Número de baños	Número de garajes	Piso de ubicación	Vista al exterior	Años de antigüedad
1												
2												
.												
.												
.												

Anexo 2. Formato de la ficha de registro de datos utilizadas en las redes neuronales artificiales

Se utilizó una tabla similar para registrar las variables de entrada y salida utilizadas en esta investigación para cada modelo:

Precio en dólares corrientes	Superficie	Número de habitaciones	Número de baños	Número de garajes	Piso de ubicación	Vista al exterior	Años de antigüedad
1							
2							
⋮							

Anexo 3. Formato de la ficha de registro de datos para los hiperparámetros asociados al modelo e hiperparámetros asociados al descenso del gradiente

En la tabla se registran cada uno de los hiperparámetros probados para el modelo. Además, en ella se registran los resultados obtenidos y la posición del modelo según su rendimiento.

Modelo	Num layers	Units 1	Units 2	Learning rate	Drop rate	Batch size	Epochs	K1	K2	MSE	Ranking
1											
2											
3											
4											
5											
6											
8											
9											
⋮											

Anexo 4. Tabla de registro de datos para el rendimiento de los modelos de redes neuronales artificiales

REDES NEURONALES ARTIFICIALES	MSE	MAE	R
Precio en dólares corrientes			

Anexo 5. Tabla de los 10 mejores modelos procesados

Ranking	Modelo	Capas ocultas	N. ° Neuronas		Learning rate	MSE
			(a)	(b)		
1	36	3	20	19	3.08E-04	2.12E-01
2	3	1	17	12	8.13E-04	4.99E-02
3	40	0	20	1	4.14E-04	8.93E-02
4	11	0	17	12	1.03E-04	5.03E-02
5	29	5	20	19	8.74E-05	1.37E-01
6	34	5	20	14	3.57E-06	1.76E-01
7	5	0	20	13	1.29E-03	4.99E-02
8	16	0	19	15	7.34E-06	8.70E-02
9	42	0	20	20	3.75E-05	4.61E-02
10	31	0	20	17	4.47E-03	4.77E-02

Anexo 6. Características del conjunto de datos

		1	2	3	4	5	6	7	8	9
Distrito	Correlación de Spearman	1.00								
	Sig. (bilateral)									
Superficie	Correlación de Spearman	0.21	1.00							
	Sig. (bilateral)	0.00								
Número de habitaciones	Correlación de Spearman	0.08	0.43	1.00						
	Sig. (bilateral)	0.00	0.00							
Número de baños	Correlación de Spearman	0.21	0.62	0.42	1.00					
	Sig. (bilateral)	0.00	0.00	0.00						
Número de garajes	Correlación de Spearman	0.26	0.57	0.19	0.50	1.00				
	Sig. (bilateral)	0.00	0.00	0.00	0.00		0.00			

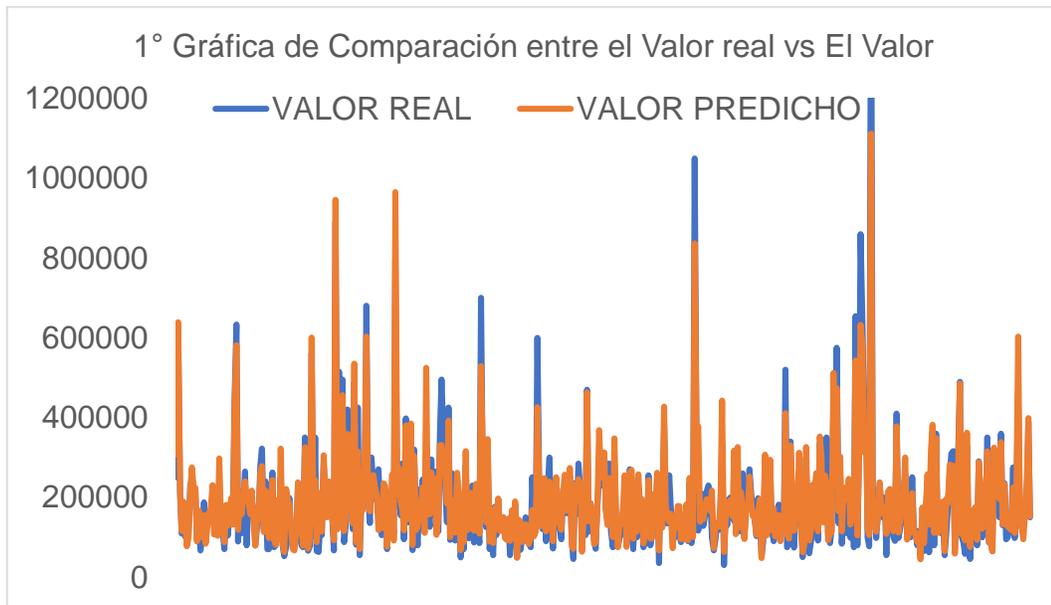
Piso de ubicación	Correlación de Spearman	-0.06	-0.14	0.01	-0.05	-0.16	1.00			
	Sig. (bilateral)	0.00	0.00	0.25	0.00	0.00				
Vista al exterior	Correlación de Spearman	0.04	0.07	0.04	0.07	0.09	-0.03	1.00		
	Sig. (bilateral)	0.00	0.00	0.00	0.00	0.00	0.00		0.01	
Años de antigüedad	Correlación de Spearman	0.01	0.12	0.02	-0.11	-0.04	-0.11	-0.02	1.00	
	Sig. (bilateral)	0.05	0.00	0.02	0.00	0.00	0.00	0.01		
Precio en dólares corrientes	Correlación de Spearman	0.30	0.80	0.25	0.62	0.70	-0.16	0.10	0.03	1.00
	Sig. (bilateral)	0.00								

Anexo 7. Rendimiento de los modelos de Redes Neuronales Artificiales mejorados

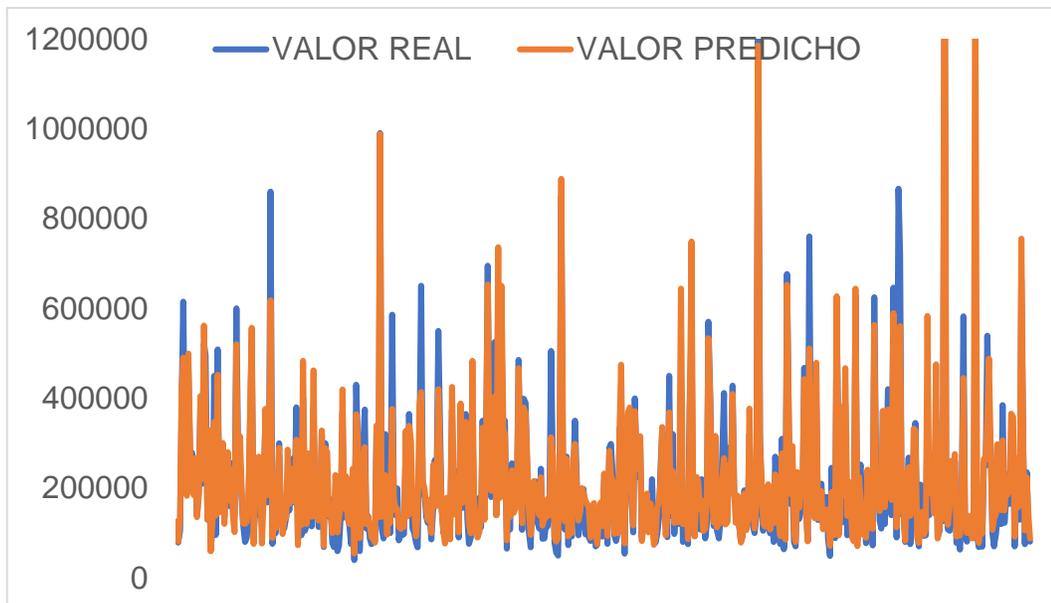
	Jarvis I	Jarvis II	Jarvis III	Jarvis IV
Tiempo de entrenamiento	2254.83032	2305.61746	2233.49114	2310.22425
Training MSE	0.00020809	0.00016182	0.0001278	0.00010892
Validation MSE	0.0002044	0.00015935	0.00012546	0.00010621
Test MSE	0.00036003	0.00034484	0.00041395	0.00036728
Training MAE	0.00881609	0.00810388	0.00732136	0.00675254
Validation MAE	0.00878539	0.00806346	0.00726813	0.00668009
Test MAE	0.00967822	0.01017159	0.01038675	0.01046011
Test MSE unscaled	4169587059	3993684510	4794110765	4253609746
Test MAE unscaled	32936.2466	34615.2311	35347.4385	35597.1096

Anexo 8. Gráfica de comparación entre los valores predicho y los valores reales

Gráfica de Comparación entre el Valor real vs El Valor Predicho (muestra 1 hasta muestra 500)



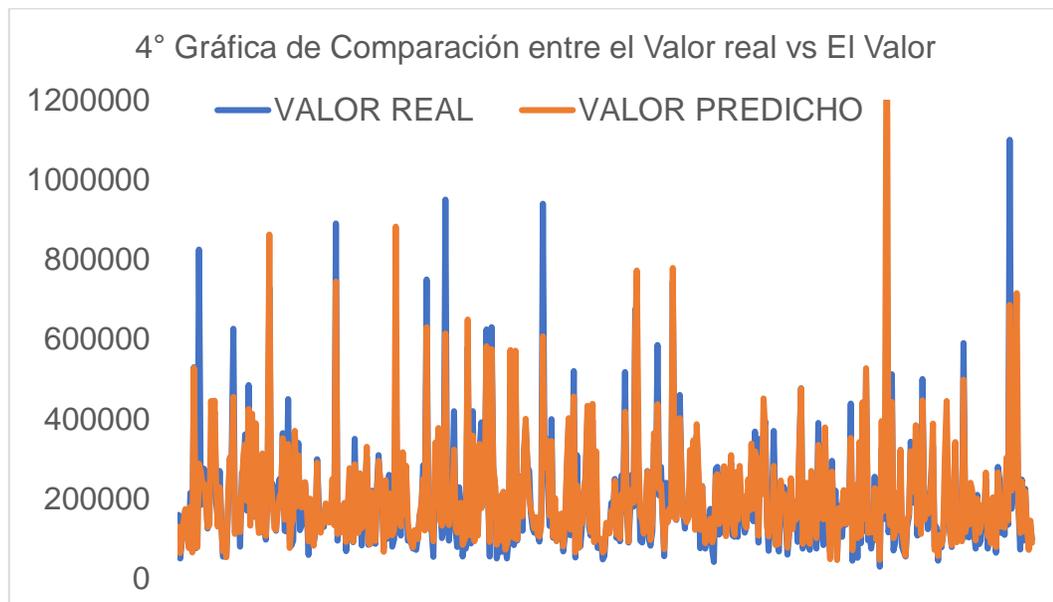
Gráfica de Comparación entre el Valor real vs El Valor Predicho (muestra 501 hasta muestra 1000)



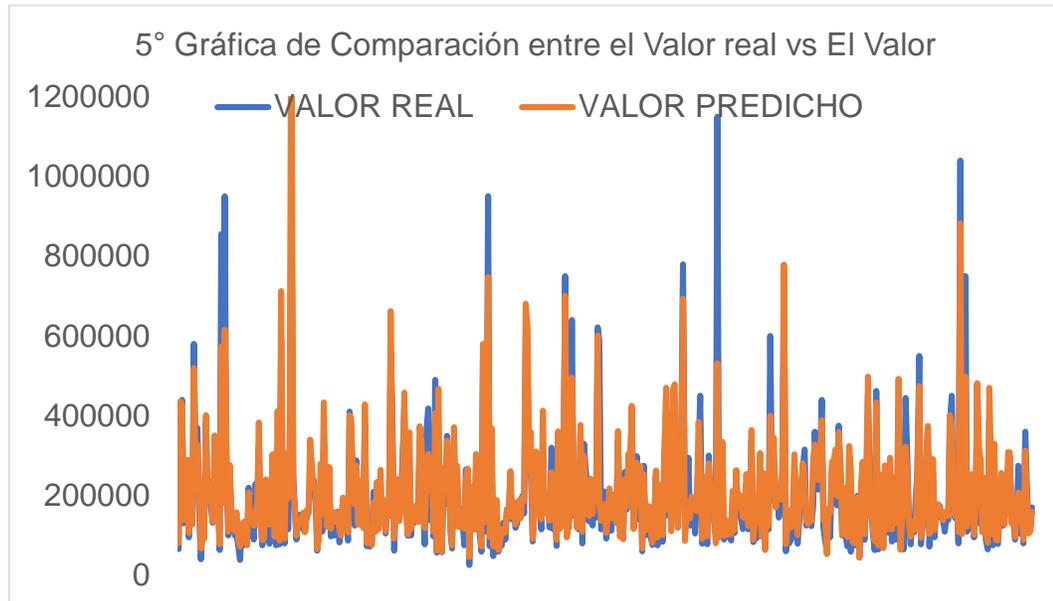
Gráfica de Comparación entre el Valor real vs El Valor Predicho (muestra 1001 hasta muestra 1500)



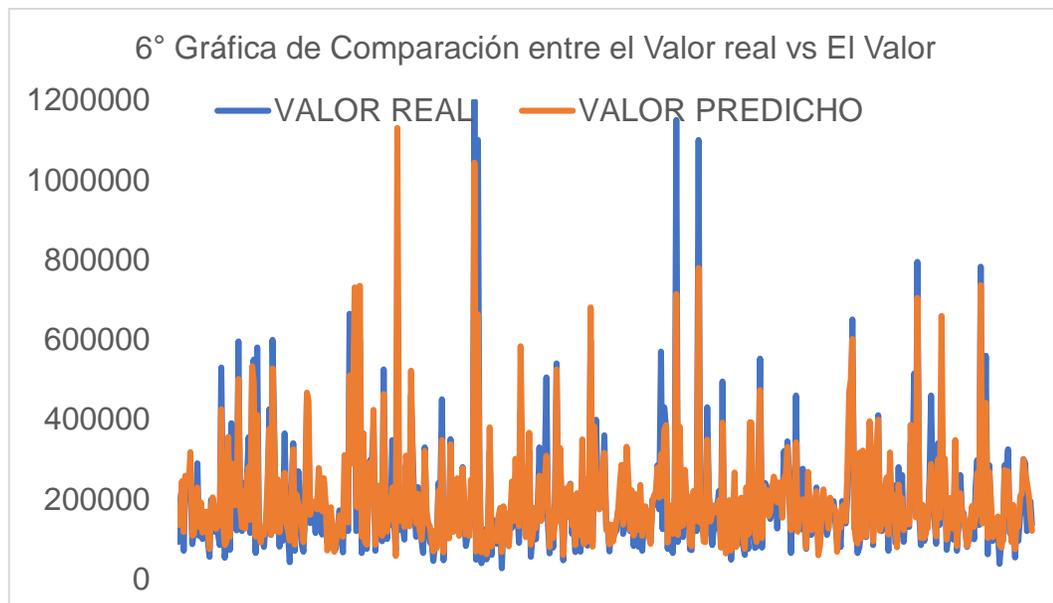
Gráfica de Comparación entre el Valor real vs El Valor Predicho (muestra 1501 hasta muestra 2000)



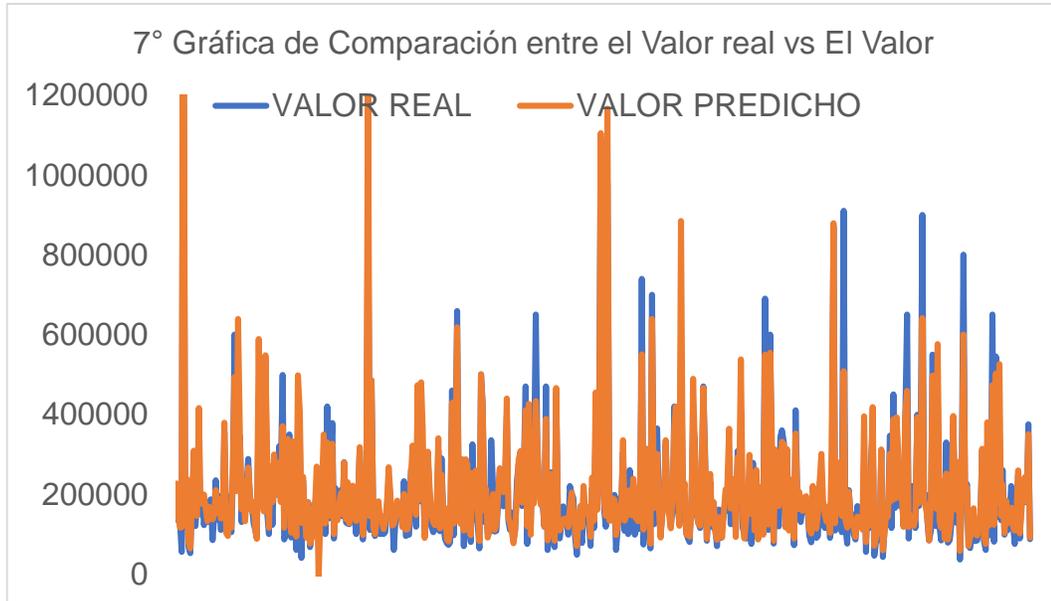
Gráfica de Comparación entre el Valor real vs El Valor Predicho (muestra 2001 hasta muestra 2500)



Gráfica de Comparación entre el Valor real vs El Valor Predicho (muestra 2501 hasta muestra 3923)



Gráfica de Comparación entre el Valor real vs El Valor Predicho (muestra 3001 hasta muestra 3500)



Gráfica de Comparación entre el Valor real vs El Valor Predicho (muestra 3501 hasta muestra 3923)

