

Modelo de valoración de propiedad raíz mediante el uso de redes neuronales

JUAN CAMILO MONTOYA ZULUAGA

**Trabajo de grado para optar al título de
Ingeniero Financiero**

Pablo Andrés Deossa Molina

MsC. en Ingeniería Administrativa y PhD. en Industria y Organizaciones.



**UNIVERSIDAD EIA
INGENIERÍA FINANCIERA
ENVIGADO
2019**

CONTENIDO

	pág.
INTRODUCCIÓN.....	7
1. PRELIMINARES.....	8
1.1 Planteamiento del problema.....	8
1.2 Objetivos del proyecto.....	9
1.2.1 Objetivo General.....	9
1.2.2 Objetivos Específicos.....	9
1.3 Marco de referencia.....	9
1.3.1 Antecedentes.....	9
1.3.2 Marco teórico.....	11
2. METODOLOGÍA.....	15
2.1 Base de datos.....	15
2.2 Formulación de modelos.....	17
2.2.1 Modelo neuronal.....	17
2.2.2 Modelo de flujo de caja descontado.....	21
2.3 Validación de los modelos.....	22
2.3.1 Validación del modelo neuronal.....	22
2.3.2 Validación del modelo de flujo de caja descontado.....	22
3. PRESENTACIÓN Y DISCUSIÓN DE RESULTADOS.....	23
3.1 Base de datos.....	23
3.2 Modelo neuronal.....	25
3.2.1 Validación con regresión múltiple.....	26
3.3 Modelo de flujo de caja descontado.....	28

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

4.	CONCLUSIONES Y CONSIDERACIONES FINALES	30
5.	BIBLIOGRAFÍA	32
6.	ANEXOS	35
6.1	Anexo 1.....	35
6.2	Anexo 2.....	35
6.3	Anexo 3.....	36
6.4	anexo 4	36

LISTADO DE TABLAS

	pág.
1. Tabla 1. Asignación de categorías numéricas.....	15
2. Tabla 2. Descripción de variables con valores binarios.....	16
3. Tabla 3. Descripción de las demás variables del modelo.....	17
4. Tabla 4. Error mínimo en cada repetición.....	26
5. Tabla 5. Comparación de errores de los modelos.....	28

LISTADO DE FIGURAS

	pág.
1. Figura 1. Estructura de una red neuronal.....	13
2. Figura 2. Descripción de la metodología.....	18
3. Figura 3. Estructura de una red neuronal con una capa oculta.....	20
4. Figura 4. Estructura de una red neuronal con dos capas ocultas.....	20
5. Figura 5. Estructura de una red neuronal con tres capas ocultas.....	21
6. Figura 6. Frecuencia de las variables: estrato, tipo de inmueble y municipio.....	23
7. Figura 7. Valores, distribución y media de la variable área.....	24
8. Figura 8. Valores, distribución y media de la variable precio.....	25
9. Figura 9. Ajuste de los modelos.....	27

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

LISTADO DE ECUACIONES

	pág.
1. Ecuación 1. Retorno esperado modelo CAPM.....	12
2. Ecuación 2. Cálculo del Beta.....	12
3. Ecuación 3. Error cuadrático medio.....	14
4. Ecuación 4. Normalización min-max.....	16
5. Ecuación 5. Flujo de caja neto esperado.....	22
6. Ecuación 6. Tasa de crecimiento esperado para vivienda.....	22
7. Ecuación 7. Tasa de crecimiento esperado para destinaciones distintas a vivienda....	22
8. Ecuación 8. Valor terminal.....	22

RESUMEN

El sector inmobiliario es de gran importancia para el país por el producto que representa, la población que involucra y el empleo que genera. La valoración de la propiedad raíz en el país se hace usualmente mediante aproximaciones de mercado que podrían tener errores de estimación y sub o sobrevalorar la propiedad. En el caso de una sobrevaloración de los precios, las consecuencias para el sector y la población que involucra serían considerables. El objetivo de este trabajo de investigación es desarrollar un modelo de valoración de propiedad raíz mediante el uso de redes neuronales. Lo anterior, pasando por una fase inicial de selección y obtención de información, una fase exploratoria de métodos de estimación para el desarrollo de un modelo de red neuronal, una fase de validación de los resultados obtenidos y finalmente una de comparación entre los precios obtenidos y otras valoraciones de mercado. Se espera contar al final de este proceso con un modelo que utilice la red neuronal desarrollada para hacer valoraciones de propiedad raíz mediante el uso de las variables previamente seleccionadas y finalmente tener un análisis de las diferencias, en caso de haberlas, entre las valoraciones del modelo obtenido y otras valoraciones del mercado local. Lo anterior, concluyendo que las redes neuronales cuentan con ventajas frente a las regresiones para la valoración de inmuebles, por su capacidad de generalización y de ajustar relaciones polinómicas entre las variables, y que los resultados obtenidos no permitieron concluir sobre la presencia de una burbuja de precios en el mercado inmobiliario colombiano.

ABSTRACT

The real estate sector is highly important for the country for the production it promotes, the population involved and the jobs that it generates. The valuation of real estate in the country is usually made with market approaches that could bring estimation errors and promote property's misvaluation. The consequences for the population involved and the sector would be important if there were an overvaluation of prices. The objective of this investigative work is to develop a real estate valuation using neural networks. That, going through an initial phase of data selection and procurement, an exploratory phase of methods for the development of a neural network model, a result validation phase and finally a comparison between the prices that were obtained and other market valuations. At the end of this process, a model with the developed neural network is expected to make real estate valuations using the variables that were previously selected and, at last, an analysis of the differences between the valuations obtained by the model and other valuations from the local market, if there were these differences. Concluding that neural networks have various advantages in their ability to generalize data and establish polynomial relations between variables if they are compared to regressions and that the results obtained were did not allow to conclude about the presence of a price bubble in Colombia's real estate market

INTRODUCCIÓN

Esta investigación aborda el problema de las imprecisiones en valoraciones de propiedad raíz, generadas por la propensión de los profesionales del mercado inmobiliario, a tomar decisiones basándose en estimaciones realizadas con modelos lineales y en precios de mercado que no reflejan el valor real de los inmuebles. Las imprecisiones mencionadas, en un mercado como este, pueden tener efectos negativos por su alcance y por la dependencia de la población a él.

Para este fin, se propone un modelo que haga una valoración que considere relaciones polinómicas entre las variables de entrada y el precio, y que realice una valoración objetiva de los inmuebles. Inicialmente, se formuló un modelo neuronal, para el que se formó una base de datos con las características de inmuebles y su precio, se eligió la estructura definitiva de la red neuronal mediante varios métodos y se almacenó información para comparar los resultados del modelo frente a los de una regresión y para identificar si el modelo neuronal tenía beneficios sobre ella.

Además, se formuló un modelo de valoración de flujo de caja descontado, que se basa en los ingresos esperados y hace una valoración objetiva de los inmuebles, para evaluar la presencia de una burbuja de precios en el mercado.

Para el desarrollo de estos objetivos, se desarrolla el documento de la siguiente manera: en la primera sección, se plantea el problema que va a abordarse a lo largo de la investigación, se establecen objetivos para guiarla, se hace una revisión de antecedentes sobre el tratamiento del problema y sobre el método propuesto en fuentes académicas, y se tratan temas teóricos que el lector puede encontrar útiles para la comprensión del documento.

En las secciones siguientes, se expone a detalle la metodología que se usó para el cumplimiento de los objetivos de la investigación, se analizan los resultados obtenidos por los modelos y se concluye a partir de ellos.

Una vez finalizada la investigación, se espera encontrar que, contando con características físicas de un inmueble, los modelos con redes neuronales proporcionan beneficios importantes sobre los modelos de regresión, para la estimación de los precios de propiedad raíz. También, se espera obtener más información sobre la presencia de una burbuja de precios en el mercado inmobiliario colombiano, a pesar de que podría haber algunas limitaciones para ese análisis por características de la base de datos dispuesta para la investigación.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

1. PRELIMINARES

1.1 PLANTEAMIENTO DEL PROBLEMA

Esta investigación busca inicialmente, proponer una alternativa al problema que implica la valoración imprecisa de inmuebles. La tendencia de usar los precios de mercado de los inmuebles como alternativa de valoración puede generar errores, pues estos no corresponden necesariamente al valor real y por lo tanto podrían ser superiores o inferiores al mismo. Algunos factores que hacen riesgosos los errores en las valoraciones de inmuebles se exponen a continuación.

En cuanto al factor del empleo, el mercado contaba con 1.8 millones de personas en febrero de 2017 según el Ministerio de Vivienda Ciudad y Territorio (2017), en el trimestre entre diciembre de 2016 y febrero de 2017 las actividades inmobiliarias, empresariales y de alquiler fueron las mayores generadoras de empleo con un total de 103.000 puestos de trabajo adicionales a los registrados en el año anterior, que si se agregan a los empleos de construcción, completan 3.2 millones de personas según el Ministerio de Vivienda Ciudad y Territorio (2017).

Además del factor relacionado con el empleo, una parte del mercado inmobiliario busca cubrir la necesidad de vivienda que es esencial para los hogares, según Torres (2012), las personas que tienen ingresos laborales de un salario mínimo o menos, utilizan aproximadamente el 63% de su salario en arrendamiento de vivienda y aquellos con los mismos ingresos entre 1.1 y 2 salarios mínimos, el 20%. Según El Espectador (2016), el 37.3% de los hogares colombianos viven bajo la modalidad de arriendo y el 45.9% en vivienda propia.

El alcance y dependencia de la población al mercado inmobiliario colombiano, ocasionada por la alta generación de empleo, la necesidad de la vivienda, la participación en contratos de arrendamiento y la propiedad de vivienda u otros, hace que investigaciones que busquen mejorar prácticas como la valoración de inmuebles en el sector, sean relevantes.

Este trabajo también busca obtener información sobre la presencia o ausencia de una burbuja de precios en el mercado inmobiliario colombiano, teniendo en cuenta que, según Blanchard y Watson (1982), las desviaciones alrededor del precio objetivo de un activo, que refleja sus fundamentales, son señales de irracionalidad y pueden tener fuertes efectos en la economía. Adicionalmente, en el caso de una sobrevaloración de la propiedad raíz, los precios elevados implican mayores retornos para la industria, y, por lo tanto, una mayor oferta futura que; sin haber cambiado la demanda, implica menores rentas en el futuro. Entonces, si hay una burbuja, la disminución repentina de los precios elevados irracionalmente a valores normales que reflejen sus fundamentales, tiene efectos negativos sobre la población, lo que hace relevante investigar sobre las burbujas de precios y su presencia en el mercado local.

Finalmente, considerando los problemas y efectos descritos anteriormente, se hace relevante la exploración de herramientas de valoración, de tipo matemático y de inteligencia

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

computacional, que no estén sesgadas y que sean particulares para el mercado colombiano, considerando este mercado como el marco en el que se aplicarán los temas expuestos.

1.2 OBJETIVOS DEL PROYECTO

Buscando dar solución a los problemas planteados, se establecen los siguientes objetivos general y específicos para mantener claro el alcance que busca tener la investigación y guiar el proceso.

1.2.1 Objetivo General

Desarrollar un modelo de valoración de propiedad raíz mediante el uso de redes neuronales.

1.2.2 Objetivos Específicos

- Identificar las variables pertinentes para la estimación del modelo.
- Diseñar una red neuronal que componga el modelo de valoración.
- Validar los resultados de la red neuronal propuesta y realizar ajustes, en caso de ser necesario, incluyendo otras herramientas de estimación.
- Analizar las diferencias; en caso de haberlas, entre las valoraciones arrojadas por el modelo y valoraciones comerciales del mercado local.

1.3 MARCO DE REFERENCIA

1.3.1 Antecedentes

En el pasado, otros autores han hecho referencia a las prácticas de valoración de inmuebles en Colombia, según Dinero (2017), lo tradicional en el sector es que los inmuebles se valoren a precios de mercado, aspirando a que los inmuebles sean vendidos a precios similares a los de otros en la misma zona, por lo tanto, predomina una valoración asociada simplemente a una diferencia entre el precio de compra y de venta de los inmuebles.

La propensión a utilizar valoraciones de mercado en operaciones reales de compra y venta, podrían impulsar diferencias entre el valor real de los inmuebles y el valor al que se negocian, ocasionando incluso burbujas de precios.

En la literatura, se han hecho múltiples menciones a una posible burbuja inmobiliaria en el país. Según Dinero (2014), la Asociación Nacional de Instituciones Financieras (ANIF) alertó sobre la presencia de una burbuja e indicó que los precios de las viviendas habían

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

sobrepasado el crecimiento del ingreso real de los hogares. Según Clavijo y Ortiz (2016), durante el periodo entre 2005 y 2014, los precios de la vivienda usada acumularon crecimientos que evidenciaban la presencia de una burbuja inmobiliaria. En el caso de Bogotá, un análisis estadístico realizado por Cediel y Velásquez (2015) proporciona evidencia de crecimientos exuberantes de precios de vivienda en algunas zonas de la ciudad desde 2009, con aumentos no homogéneos, ni coyunturales.

En cuanto a métodos de valoración de propiedad raíz, varios autores han desarrollado investigaciones con objetivos similares a los de esta exploración, realizando estimaciones con modelos neuronales y comparando los resultados con los de otros métodos para la valoración de activos. Un trabajo realizado por Ceular, Caridad, & Núñez (2008) en España, busco realizar una comparación entre la metodología de valoración de precios hedónicos, que considera el valor de los inmuebles como un conjunto de atributos distintos con el uso de regresiones múltiples, y las redes neuronales, Concluyendo que la red neuronal que usó resultó ser más apropiada, pues permite evitar la rigidez que representan los planteamientos de la metodología hedónica y que la red neuronal fue más efectiva al momento de reconocer y calcular valores de inmuebles atípicos.

Chaphalkar y Sandbhor (2013) concluyeron en otra investigación que las redes neuronales presentan un mejor rendimiento en la valoración si se las compara con modelos que utilizan regresiones múltiples y que son comúnmente usados, además señala varios beneficios del uso de redes neuronales y específicamente la facilidad que tienen en el reconocimiento de patrones de los datos suministrados y que no requieren un alto conocimiento de los datos que las alimentan para realizar predicciones acertadas. También, Mimis, Rovolis, y Stamou (2013) realizaron un estudio comparando los resultados de las redes neuronales y un modelo de regresión basado en auto correlaciones espaciales y obteniendo resultados similares en cuanto a la superioridad de los modelos neuronales. Peterson y Flanagan (2009) afirman que los modelos de regresión lineal (usados comúnmente en la valoración de propiedad raíz) generan errores significativos que son evitables y que podrían resultar costosos, lo anterior ocasionado por la no linealidad en la relación entre las características y el valor de los inmuebles y que dichos modelos de regresión asumen como lineal.

En el caso colombiano, se han realizado varios trabajos de investigación que usan redes neuronales para la estimación de modelos, algunos en otros sectores, como el modelo de pronóstico de exportaciones de café planteado por Gil Serna (2015), las evaluaciones de resistencia a compresión del concreto planteadas por González Salcedo, Guerrero Zúñiga, Arjona, & Will (2012) y González, Guerrero, Delvastro, & Will (2012) y otros orientados al sector financiero, como la predicción de valores de la inflación del país de dos trabajos propuestos por Santana (2006) y Aristizábal (2006).

Casas (2014), realizó una investigación para hacer valoraciones de propiedad raíz con modelos neuronales, específicamente, buscando realizar mejoras al modelo de valoración que utiliza el Instituto Geográfico Agustín Codazzi para el cálculo del avalúo catastral en Colombia. Además de incorporar dentro del modelo, variables de Sistemas de Información Georreferenciada, enfatiza en el impacto que podría tener la implementación de un modelo de este tipo para el avalúo catastral, y por consecuencia, para el cálculo del pago del impuesto predial e incluso para valoraciones implicadas en préstamos bancarios.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

La literatura del tema presente en bases de datos auditadas y fuentes confiables de información e investigación evidencia que las redes neuronales han sido usadas, con resultados predominantemente favorables, para la valoración de propiedad raíz en otros lugares del mundo y que en Colombia las mismas han sido usadas en múltiples ocasiones para la estimación de modelos en el sector financiero y otros.

1.3.2 Marco teórico

Teniendo en cuenta que los objetivos de esta investigación están encaminados a formar un modelo de valoración de propiedad raíz, se hace necesario discutir cuáles con los métodos usados para ese fin comúnmente, también abordar otros métodos que esta investigación pretende abordar y aclarar conceptos necesarios para entender el funcionamiento de las redes neuronales, que son la base del modelo de valoración que se propone e introducir aspectos importantes sobre las burbujas de precios.

○ Métodos de valoración

Como aclaran IFRS Foundation (2013) e IVSC (2011), organizaciones encargadas de generar estándares de reporte de información financiera y de valoración de activos en sus estándares IFRS 13 (International Financial Reporting Standards) e IVS 2017 (International Valuation Standards), las técnicas de valoración generalmente usadas en la valoración de inmuebles son las de mercado, ingreso y costo. Ernst & Young (2013) recopiló la siguiente información sobre esas técnicas:

- La aproximación de mercado busca determinar un valor basado en otras transacciones comparables resultantes de negociaciones con activos idénticos o comparables y normalmente estableciendo bases de comparación entre propiedades como el valor del metro cuadrado para hacerlo. Esta técnica es comúnmente usada por los agentes del mercado inmobiliario.
- La aproximación del ingreso se basa en una estimación de los ingresos futuros y las ganancias o flujos de dinero que se espera entregará un activo para llegar a un estimado del valor. Aunque tiene variantes como el método de capitalización del ingreso y el del flujo de caja descontado, se conserva el principio que establece que el ingreso neto para un periodo futuro se ajusta al presente usando una tasa de descuento.
- Por su parte, la aproximación del costo refleja la cantidad de dinero que sería necesaria para reemplazar el activo en cuestión.

Un modelo de “flujo de caja descontado” es una variación de la aproximación del ingreso que concentra la idea de que el valor de cualquier activo que produce un flujo de caja es el valor presente de los flujos de caja esperados que producirá en el futuro (Damodaran, 2012), propone proyectar al futuro los flujos de ingreso de ese activo y valorar el activo como la suma del valor presente de cada uno de esos flujos, con el uso de una tasa de descuento.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

La tasa mencionada anteriormente, conocida como costo de capital, se calcula como una ponderación de la tasa de patrimonio (requerida por los inversionistas de un activo) y la tasa de la deuda (requerida por oferentes de liquidez), esta ponderación corresponde a la relación entre el patrimonio y la deuda de la fuente de financiación de un proyecto, empresa, o en este caso, un inmueble.

El costo de patrimonio antes mencionado es el retorno que un inversionista espera de un activo, este retorno puede ser calculado con modelos como el "CAPM". De acuerdo con Womack & Zhang (2003) El modelo CAPM ("Capital Asset Pricing Model") busca cuantificar la relación entre el Beta (β) de un activo con el mercado y su retorno esperado. El modelo generaliza la relación entre los retornos esperados de un activo y su exposición al riesgo de mercado, así:

$$\text{Retorno esperado} = R_f + \beta (R_m - R_f) \quad (1)$$

Donde R_f es la tasa libre de riesgo, R_m el rendimiento del mercado y β el Beta entre el activo en cuestión y el mercado, definido como:

$$\beta_{\text{Act.i,mercado}} = \frac{\text{Cov}(R_{\text{Activo } i}, R_m)}{\delta_{\text{Rend.mercado}}^2} \quad (2)$$

La técnica de valoración de mercado es ampliamente usada y aunque es de fácil aplicación por los pocos requerimientos técnicos que tiene, puede llevar a juzgamientos equivocados del precio precisamente por el mismo motivo. La técnica del ingreso y su variante de flujo de caja descontado, solo se basa en los ingresos esperados del inmueble, por eso está alejada de factores del mercado que podrían llevar a una valoración errónea.

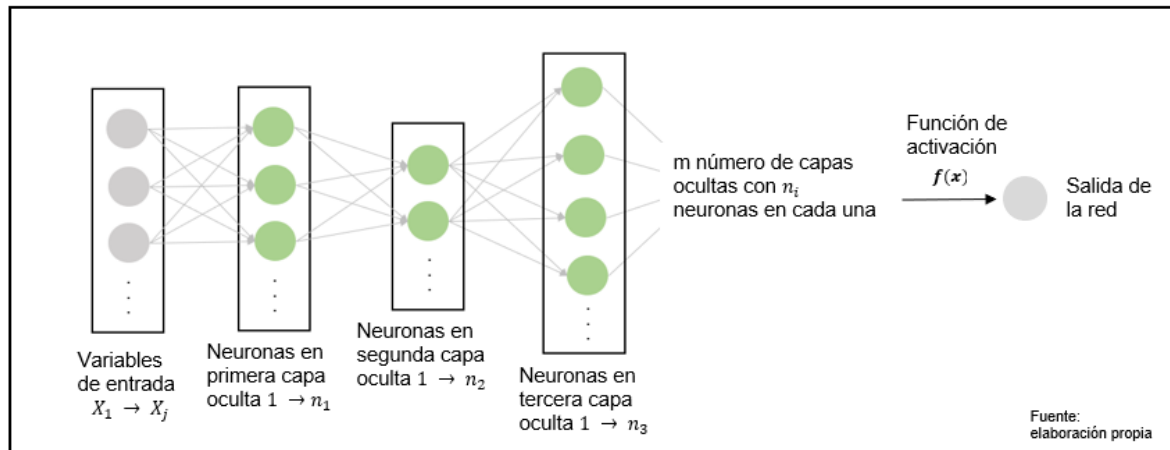
o **Redes neuronales**

Las redes neuronales artificiales son modelos matemáticos diseñados a imitación del cerebro y que buscan específicamente imitar actividades relacionadas con la estructuración y desarrollo del conocimiento (García, Gamez, & Alfaro, 2006). Entre los beneficios de las redes neuronales frente a otros métodos, según Chaphalkar & Sandbhor (2013), tienen un mejor rendimiento y menores errores de estimación en comparación a otras herramientas como las regresiones múltiples, además cuentan con una mayor facilidad para el reconocimiento de patrones de un conjunto de datos y no requieren de mucho conocimiento de los datos que van a utilizarse para tener resultados válidos. Adicionalmente, Peterson & Flanagan (2009) señala el beneficio que representa el hecho de que las redes neuronales no asuman relaciones lineales entre las variables de entrada y salida, linealidad que si asumen los modelos de regresión normalmente usados.

Algunos elementos importantes para resaltar en la estructura base de una red neuronal son las variables de entrada, las neuronas, las capas ocultas, la función de activación y la salida de la red. Esa estructura se detalla en la figura 1 a continuación:

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Figura 1. Estructura de una red neuronal.



Mirjadili (2015) aclara que la primera capa (izquierda) se conoce como la capa de entrada, la última como la capa de salida, las demás capas entre la de entrada y la de salida como capas ocultas.

En el proceso conocido como “entrenamiento”, en el que la red neuronal define los pesos con el que las entradas y las neuronas del modelo se relacionan, la información de las variables de entrada en la primera capa alimenta a las neuronas de la segunda capa ponderándose con un peso $W_{i,j}$ y las neuronas alimentan a su vez a las demás capas ocultas, el resultado de la ponderación de todos los pasos a través de las capas ocultas pasa finalmente por una “función de activación” para obtener el resultado final de la red. Finalmente, este resultado final se evalúa frente a un valor real de entrenamiento obteniendo un error, a partir del cual se generan nuevos $W_{i,j}$ buscando lograr un menor error del modelo.

En el caso particular del modelo propuesto, la estructura de red neuronal utilizada se conoce como “red neuronal de propagación hacia adelante” (FNN por sus siglas en inglés) que de acuerdo con Mirjadili (2015), es una red en la que hay una conexión unidireccional entre las neuronas, las variables de entrada alimentan con información a las neuronas de la primera capa oculta y esas neuronas hacen lo mismo con las de las siguientes capas, sin que nunca una neurona devuelva información a una neurona de una capa anterior. En caso de tener una sola capa oculta, u una FNN se conoce como “perceptrón multi capa” (MLP por sus siglas en inglés).

Es importante aclarar que la elección del número de capas ocultas y el número de neuronas en cada una de esas capas son factores importantes para el mejor entrenamiento y

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

desempeño del modelo, lo que hace importante buscar configuraciones que logren disminuir los errores de estimación.

Una manera de medir los errores de entrenamiento y prueba del modelo, para así medir su capacidad de ajuste de los datos y su generalización es el error cuadrático medio (ECM), calculado como (Casas, 2017):

$$ECM = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 \quad (3)$$

- **Burbujas de precios**

Kindleberger (2016) define las burbujas de precios como un gran aumento en el precio de un activo o un rango de activos en un proceso continuo, en el que el aumento inicial genera expectativas de aumentos posteriores y atrae nuevos compradores, generalmente especuladores interesados en obtener ingresos por negociar el activo y no por su uso o su capacidad de generar ingresos.

Desde una vista más orientada al valor de los fundamentales, Rosser (2000) considera que las burbujas especulativas existen cuando el precio de algo no equivale a lo que reflejan sus fundamentales de mercado; planteados como equilibrios en el largo plazo consistentes con un equilibrio general, por algún periodo de tiempo y por razones diferentes a choques aleatorios.

Es importante mencionar que en la literatura no hay referencia al límite de sobrevaloración a partir del cual se considera que los precios en un mercado hacen parte de una burbuja. Para efectos prácticos, puede definirse un cierto límite para calificar el precio como uno de burbuja y asumir que la diferencia no corresponde a imperfecciones normales del mercado ni está justificada.

2. METODOLOGÍA

Esta sección presenta el método usado en esta investigación para el logro de los objetivos propuestos, partiendo desde la consecución de los datos, los ajustes necesarios para ser usados, continuando con la formulación de los modelos propuestos y las mediciones propuestas para validarlos.

2.1 BASE DE DATOS

La base de datos utilizada está compuesta por la información de 502 inmuebles, recaudados por una compañía intermediaria del sector inmobiliario en la ciudad de Medellín, Colombia. La base de datos contiene inmuebles con destinación tanto comercial como de vivienda ubicados en la ciudad de Medellín y municipios aledaños, con información de las características físicas del inmueble, su precio de arrendamiento y en algunos casos, de arrendamiento y venta.

Teniendo en cuenta que la información primaria de un estudio no está normalmente formada de una forma adecuada para usarse en modelos de regresión y debe ser inspeccionada y en algunos casos modificada antes de ser usada. (Steyerberg, 2019), se propone hacer inicialmente, una transformación de las variables categóricas alfanuméricas en variables numéricas; asignando a cada categoría un código distinto (numérico), las variables transformadas y las categorías asignadas fueron las siguientes:

Tabla 1. Asignación de categorías numéricas.

Categoría	Tipo de inmueble	Municipio	Tipo de cocina	Tipo de gas	Parqueadero
1	Apartamento	Envigado	Semi - integral	Gas con pipeta	Descubierto
2	Aparta estudio	Itagüí	Integral	Red de gas	Cubierto
3	Local comercial	Bello	Integral renovada	Sin gas	Sótano
4	Bodega	Medellín	Integral abierta	NA	NA
5	Casa	Sabaneta	Integral con barra	NA	NA
6	Oficina	Rio negro	NA	NA	NA
7	Consultorio	San Jerónimo	NA	NA	NA

Adicionalmente, se propone realizar la normalización de las variables que, según Jain, Nandakumar y Ross (2005), hace homogéneas las escalas numéricas de los datos para facilitar su uso como entradas de un modelo. Lo anterior, usando el método de

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

normalización “min-max”, desplazando los valores de las variables al rango [0,1] con la siguiente transformación:

$$X_{i\ norm} = \frac{X_i - \min_X}{\max_X - \min_X} \quad (4)$$

Un conjunto representativo de las variables de la base de datos describe las características físicas de los inmuebles y son, en su mayoría, variables con un valor binario (0 cuando el inmueble no cuenta con la característica *i* y 1 cuando cuenta con ella). En las siguientes tablas, se resumen estadísticamente las variables usadas para la investigación, inicialmente con las variables que toman valores binarios y después con las que toman valores categóricos y continuos.

Tabla 2. Descripción de variables con valores binarios.

Característica	Moda	Característica	Moda
Garaje cubierto	0	Alcoba de servicio	0
Agua caliente	0	Baño de servicio	0
Unidad cerrada	1	Patios	0
Sala integral	1	Turco	0
Administración	0	Sala independiente	0
Parqueadero de visitantes	1	Zona deportiva	0
Ascensor	1	Sauna	0
Balcón	0	Baño social	0
Cuarto útil	0	Hall	0
Juegos	0	Comedor separado	0
Salón social	0	Solárium	0
Calentador	0	Biblioteca	0
Piscina	0	Gimnasio	0
Vestier	0	Terraza	0
Zona de ropas	0	Horno	0
Citófono	1	Jacuzzi	0
Portería	1	Estufa	0
Zona verde	0	Puerta para camión	0

Tabla 3. Descripción de las demás variables del modelo.

Característica	Mínimo	Media	Moda	Máximo
Tipo de inmueble	1	2.817	1	7
Estrato	1	5.141	6	7
Municipio	1	3.663	4	7
Tipo de cocina	1	2.717	2	5
Tipo de gas	1	1.835	2	3
Número de habitaciones	0	1.727	3	6
Número de closets	0	1.502	0	6
Número de baños	0	1.875	2	6
Número de sanitarios	0	1.299	0	6
Número de lavamanos	0	1.289	0	6
Número de parqueaderos	0	0.9801	1	3
Tipo de parqueadero	0	1.747	3	3
Área	15	86.66	26	450
Precio	260,000	1,400,000	1,200,000	4,900,000

Finalmente, para la estimación de los modelos neuronales planteados, deben seleccionarse datos para formar una subbase de datos de “entrenamiento” y otra de “prueba”, compuesta en este caso por el 60% y el 40% de los datos respectivamente. Estos conjuntos de datos son seleccionados aleatoriamente.

2.2 FORMULACIÓN DE MODELOS

El proyecto plantea el desarrollo de un modelo neuronal para la valoración de propiedad raíz y la formulación de un modelo basado en el método de valoración del ingreso para ser comparado con precios de mercado.

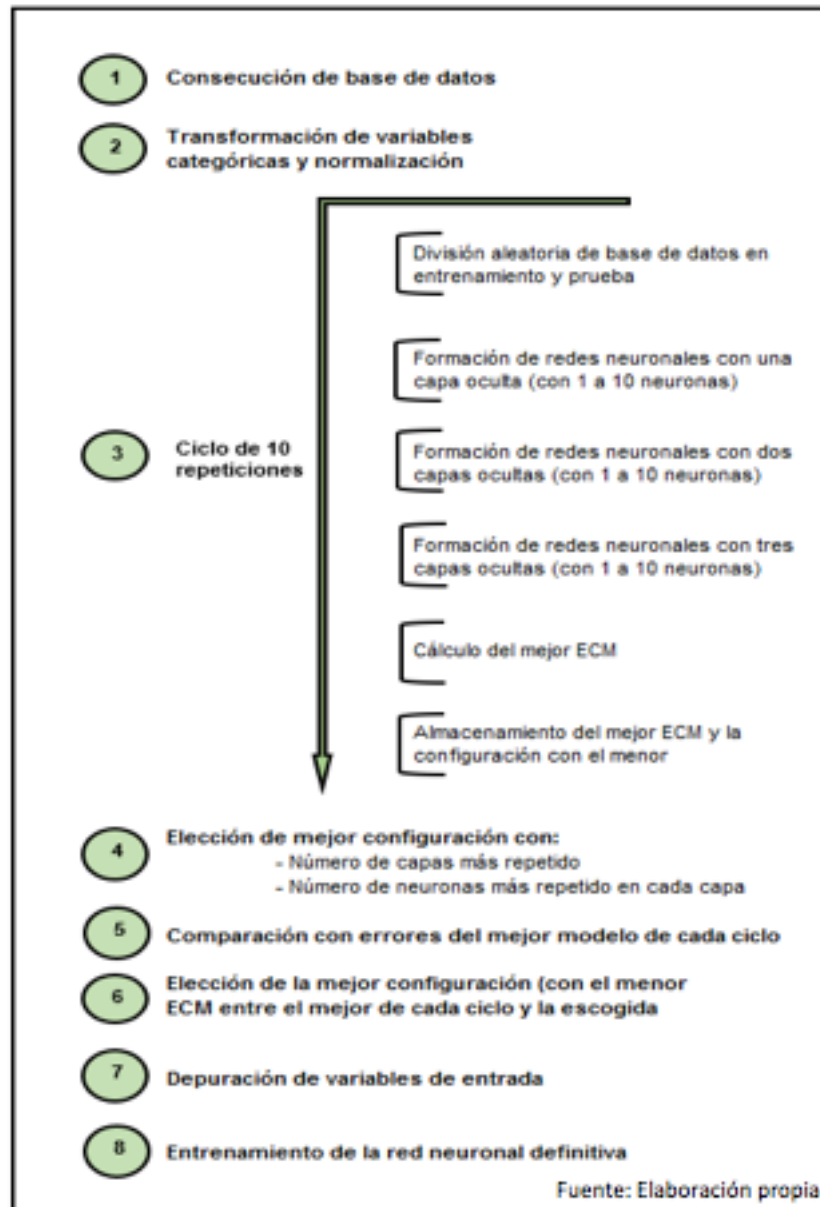
2.2.1 Modelo neuronal

Se propone ajustar un modelo neuronal con la configuración que permita obtener el menor error; medido como el error cuadrático medio (ECM) del modelo, para estimar los precios de inmuebles en una base de datos de prueba que no fue usada para el entrenamiento y comparar los resultados de esas estimaciones con otro método.

Para el ajuste del modelo se propone usar el paquete “Neuralnet v. 1.44.2” (Fritsch, Guenther, Wright, Suling, Mueller, 2019) del lenguaje de programación R 3.3.2 (R core team, 2017). El proceso propuesto puede detallarse en la figura 2:

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Figura 2. Descripción de la metodología.



El proceso consiste en la repetición en 10 ocasiones del entrenamiento de redes neuronales con distintas configuraciones, después de realizadas, se habrán entrenado 11100 redes neuronales distintas (1110 en cada repetición) para concluir sobre cuál de las configuraciones fue la mejor. A continuación, se describen los modelos que se formarán en cada una de esas repeticiones, en el Anexo 1, puede observarse el código utilizado para la ejecución de las 10 repeticiones y en los anexos 2,3 y 4, el código usado para la estimación de los modelos neuronales:

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

- 10 redes neuronales con una capa oculta y con neuronas en esa capa en un rango entre 1 y 10. La estructura se detalla en la figura 3.
- 10^2 redes neuronales con dos capas ocultas y de una a 10 neuronas en cada capa, formando todas las combinaciones posibles de redes en esas dos capas. La estructura se detalla en la figura 4.
- 10^3 redes neuronales con tres capas ocultas y de una a 10 neuronas en cada capa, formando todas las combinaciones posibles de redes en esas tres capas. La estructura se detalla en la figura 5.

Los valores de ECM calculados para cada red neuronal entrenada en el proceso son almacenados, de manera que al final de cada una de las repeticiones pueda definirse cuál fue la configuración con la que se logró tener una menor medición de error y que, por lo tanto, logró una mejor generalización (una mejor estimación para valores de una base de datos de prueba que no tuvo que ver con el entrenamiento de la red). Adicionalmente, en cada repetición se almacenan: la configuración (número de capas ocultas), el número de neuronas en cada capa oculta y el error de la red neuronal elegida.

Una vez ejecutadas 10 pruebas con las configuraciones descritas en las que se repite el proceso de entrenamiento de la red, medición del error (ECM) y selección de la configuración con el menor ECM (la “mejor” configuración), se propone elegir inicialmente para el modelo la configuración que fue “mejor” en el mayor número de repeticiones de 10 y para cada capa oculta, el número de neuronas que más se repitió en las ocasiones en que esa configuración fue la mejor, para después realizar el entrenamiento de una red neuronal con la mejor configuración elegida, para después calcular el ECM arrojado por esa red.

Después de realizado ese proceso, se elegirá la configuración de red neuronal con la mínima medición del error entre el de la escogida en el paso previo y de todas las configuraciones elegidas como mejores. Finalmente, se hará una depuración de las entradas de la red elegida, removiendo las variables de entrada con pesos inferiores, de manera que la red no se alimente con variables que no estén aportando significativamente a su entrenamiento.

Figura 3. Estructura de una red neuronal con una capa oculta.

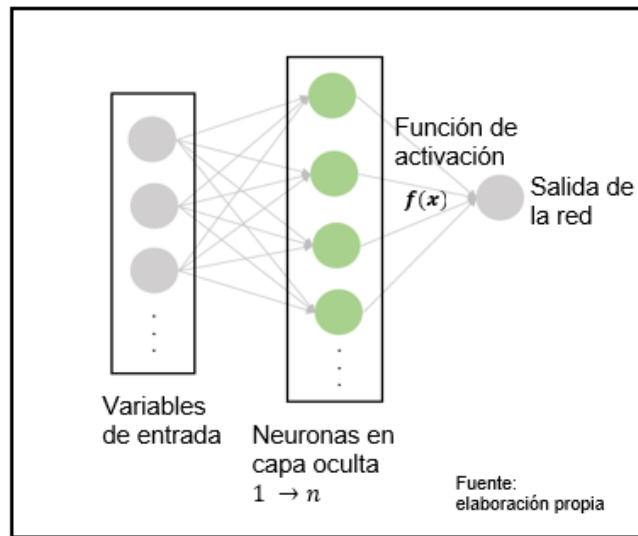


Figura 4. Estructura de una red neuronal con dos capas ocultas.

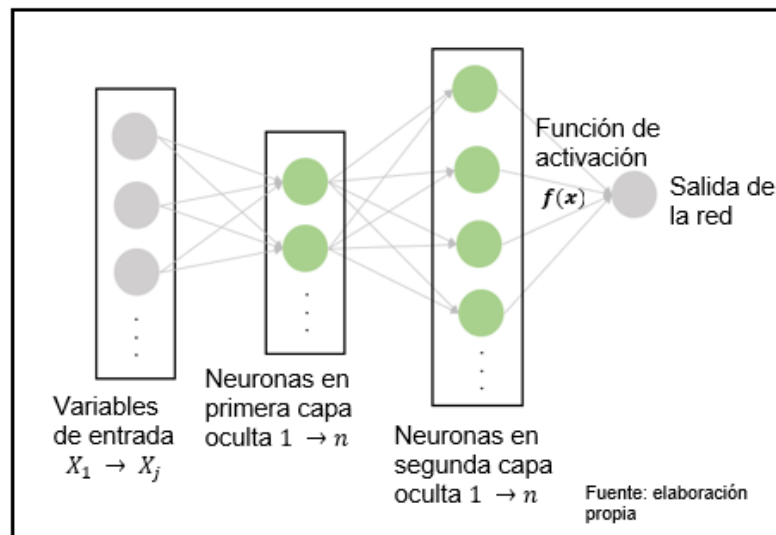
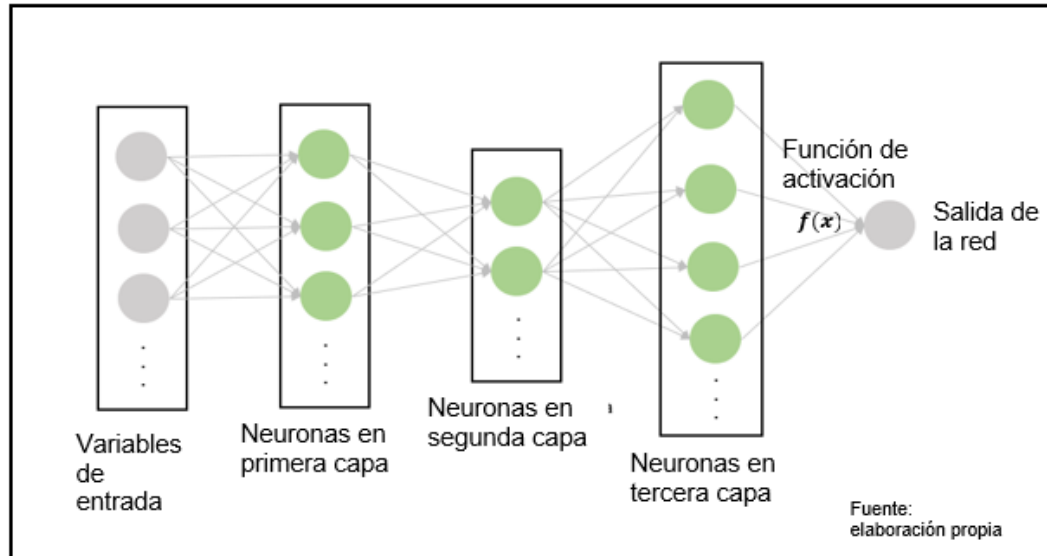


Figura 5. Estructura de una red neuronal con tres capas ocultas.



2.2.2 Modelo de flujo de caja descontado

Con el objetivo de comparar la base de datos mencionada al inicio de esta sección con una medición objetiva del valor de cada inmueble, se propone hacer un modelo de valoración breve. El valor de cualquier activo que produce un flujo de caja es el valor presente de los flujos de caja esperados que producirá en el futuro (Damodaran, 2012), a partir de eso se propone un modelo de descuento de los flujos de caja estimados de cada inmueble a valorar, teniendo en cuenta:

- Estimación de una tasa de descuento que refleje el riesgo de los activos inmobiliarios. Lo anterior, calculando un costo de capital que corresponde a una ponderación entre el costo del patrimonio y el costo de la deuda, usando en este caso, una ponderación del 30 % de patrimonio y el 70% de deuda, debido a que el 30% es un consenso comercial sobre el monto de patrimonio exigido por instituciones financieras para operaciones de crédito hipotecario en Colombia. El costo del patrimonio mencionado, se estima mediante el modelo CAPM descrito en la sección anterior y utilizando como Beta (β) del mercado inmobiliario, el calculado entre el principal índice del mercado accionario colombiano (COLCAP) y el índice de precios de la vivienda usada (IPVU); reportado por el Banco de la República de Colombia.

Adicionalmente, se toma como costo de la deuda, la tasa promedio de colocación de créditos hipotecarios para adquisición de inmuebles de las principales instituciones financieras del país, que es reportado por la Superintendencia Financiera de Colombia en el "Proyecto Tasa de Interés".

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

- Estimación del flujo de caja neto estimado mensual de cada inmueble a partir de información contenida en la base de datos inicial, que se plantea así:

$$\text{Flujo de caja}_i = \text{Ingresos por arrendamiento}_i - \text{Gasto de administración}_i - \text{Impuesto predial}_i \quad (5)$$

- Elección de una tasa de crecimiento esperado como la tasa de inflación prevista en el Marco Fiscal de Mediano Plazo 2019 (MFMP) para los próximos diez años para inmuebles con destinación de vivienda y de la tasa de inflación prevista más dos puntos porcentuales para los inmuebles con destinación comercial. (Ministerio de Hacienda y Crédito Público, 2019)

$$\text{tasa de crecimiento esperado vivienda} = \text{tasa de inflación prevista MFMP} \quad (6)$$

$$\text{tasa de crecimiento esperado otros} = \text{tasa de inflación prevista MFMP} + 2\% \quad (7)$$

- Cálculo de un valor terminal con el que se estima el valor del inmueble al final del horizonte de valoración (10 años) y que considera que la tasa de crecimiento de los flujos de caja del inmueble será constante durante los años posteriores a ese período, dónde, según Damodaran (2012):

$$\text{Valor terminal}_t = \frac{\text{flujo de caja estimado}_t}{\text{tasa de descuento} - \text{tasa de crecimiento esperado}} \quad (8)$$

Dónde t es el año del horizonte de valoración (año 10).

2.3 VALIDACIÓN DE LOS MODELOS

2.3.1 Validación del modelo neuronal

El modelo propuesto previamente se validará comparando el error (ECM) de la configuración elegida como mejor con el error arrojado por un modelo de regresión múltiple con la misma información de entrada usada para el entrenamiento de la red neuronal.

2.3.2 Validación del modelo de flujo de caja descontado

Una vez realizada la estimación de los precios de los inmuebles, se comparan dichos precios con los precios de mercado presentes en la base de datos mencionada para determinar cuál es la diferencia entre la valoración realizada por el modelo y las valoraciones y su dirección, si es superior o inferior al valor de mercado.

Para el análisis del modelo propuesto, según la opinión de un experto en el sector, un inmueble cuyo valor comercial supere en 20% al valor calculado con el método de flujo de caja descontado, se considerará que indica la presencia de una burbuja, suponiendo que ese límite supera las diferencias que podrían ocasionarse por imperfecciones del mercado.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

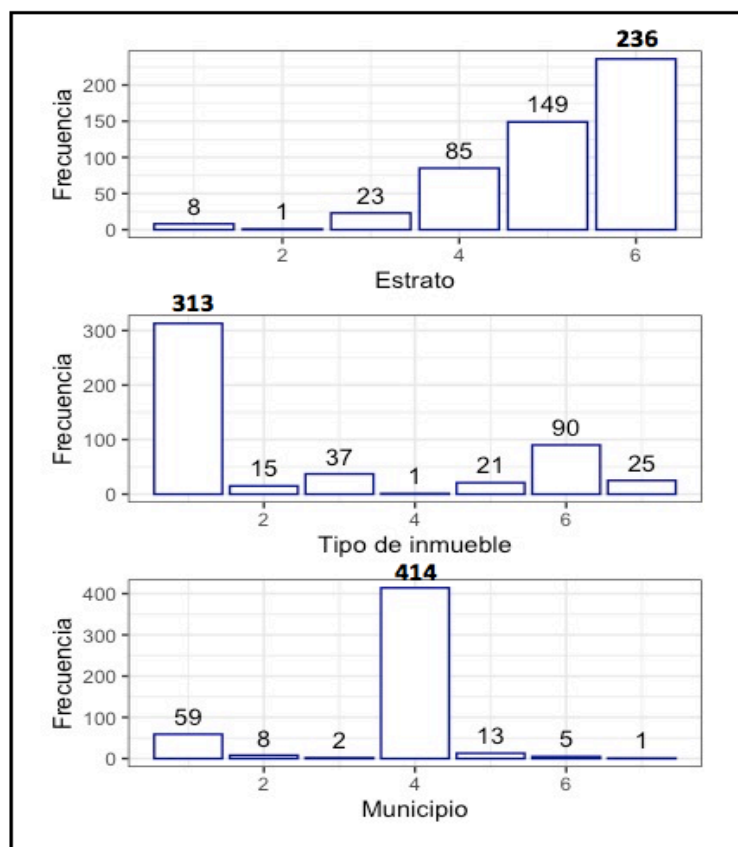
3. PRESENTACIÓN Y DISCUSIÓN DE RESULTADOS

3.1 BASE DE DATOS

A partir de la información disponible en la base de datos, es necesario hacer varias observaciones sobre las variables, debido a que estas pueden tener repercusiones en el alcance del modelo por la concentración de algunas de ellas.

Inicialmente, se destacan las observaciones realizadas sobre las variables categóricas de la base de datos. La figura 6 permite ver la frecuencia de las variables estrato, tipo de uso y municipio. Como puede observarse, los inmuebles contenidos en la base de datos, pertenecen en su mayoría a estratos 4, 5 y 6, adicionalmente, una proporción considerable de ellos son apartamentos y, por lo tanto, están destinados al uso de vivienda y una proporción incluso mayor está localizada en la ciudad de Medellín. Lo anterior, indica una particularidad de los inmuebles contenidos en la base de datos, esto debe tenerse en cuenta al momento de concluir sobre los resultados de los modelos planteados en la investigación

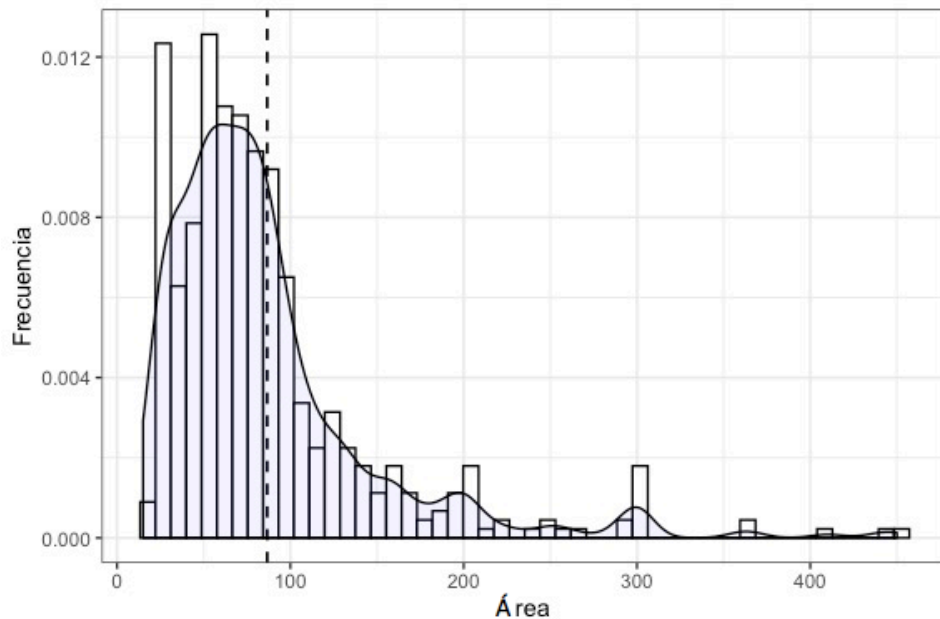
Figura 6. Frecuencia de las variables: estrato, tipo de inmueble y municipio.



La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

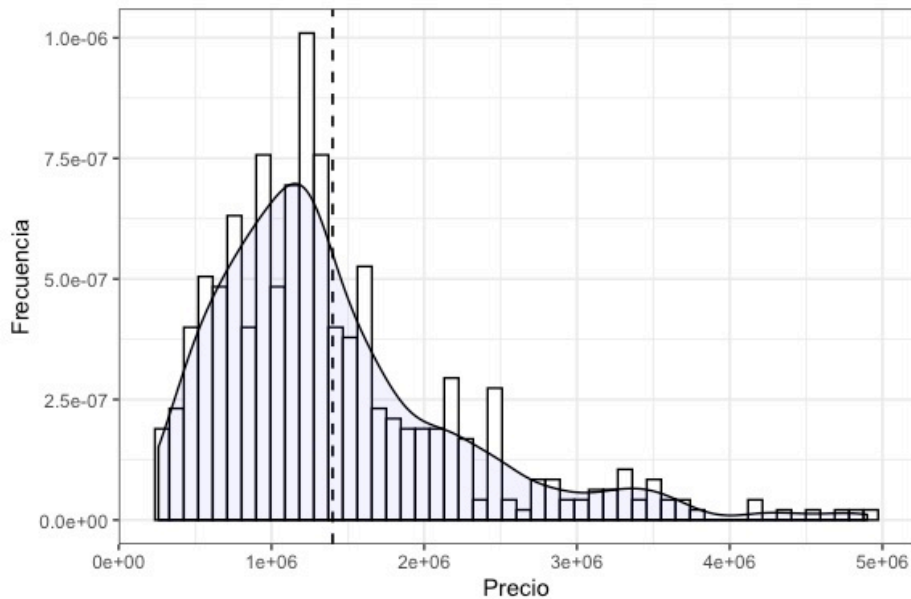
Adicionalmente, es pertinente analizar los valores de las variables área y precio para los inmuebles. En la figura 7, se detalla la variable área, con los valores que toma, la función de densidad de los mismos y su media en línea punteada. El coeficiente de asimetría de esta variable es de 2.38 y su curtosis, de 10.54, estas medidas y lo observado en la gráfica indican que la distribución tiene asimetría positiva y es leptocúrtica, por lo que, el área de los inmuebles está concentrada en valores inferiores a la media, pero con colas menos gruesas y, por lo tanto, una mayor proporción de valores cercanos a la media y una menor proporción de datos en sus colas.

Figura 7. Valores, distribución y media de la variable área.



Finalmente, en la figura 8, se detalla la variable precio, con los valores que toma, la función de densidad de los mismos y su media en línea punteada. El coeficiente de asimetría de esta variable es de 1.46 y su curtosis, de 5.54. Lo observado en la gráfica y las medidas anteriores, permiten conocer que la variable tiene una asimetría positiva y que es leptocúrtica, lo anterior implica que precios de los inmuebles disponibles para esta investigación están concentrados en valores inferiores a su media y que su distribución cuenta con menos datos acumulados en sus colas a los de una distribución normal.

Figura 8. Valores, distribución y media de la variable precio.



Las observaciones realizadas con las variables disponibles, constituyen un conjunto de consideraciones sobre los datos, que deben ser tenidas en cuenta como factores importantes en la definición del alcance de esta investigación y al momento de concluir sobre los resultados de los modelos planteados.

3.2 MODELO NEURONAL

Una vez realizadas las 10 repeticiones en las que se formaron las diferentes configuraciones mencionadas, en ninguna de las repeticiones el mejor modelo tenía una sola capa oculta, en una de ellas dos capas ocultas y en 9 de ellas tres, por lo que pudo concluirse que el mejor modelo neuronal para el set de datos es uno con 3 capas ocultas. Los errores obtenidos en cada repetición y la configuración elegida para cada una pueden observarse en la Tabla 4.

Con la metodología que plantea tomar en cuenta los números de neuronas que más se repitieron en cada una de las capas ocultas de las redes neuronales seleccionadas, se definió como mejor número de neuronas: una en la primera capa, tres en la segunda y siete en la tercera. Sin embargo, el error mínimo obtenido entre los resultados de todas las repeticiones y el de la red seleccionada fue el de la quinta repetición y no el seleccionado (de 1,3 y 7 neuronas en cada capa oculta), por eso se eligió como mejor configuración la obtenida en esa repetición, que contaba con tres capas ocultas y 3, 3 y 7 neuronas respectivamente en cada una.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Tabla 4. Error mínimo en cada repetición.

Repetición	Error mínimo	Número de capas	Número de neuronas en cada capa
1	0.0138	3	1-3-6
2	0.0166	2	1-9
3	0.0140	3	3-4-8
4	0.0140	3	5-3-10
5	0.0124	3	3-3-7
6	0.0149	3	1-10-6
7	0.0164	3	4-10-9
8	0.0139	3	2-1-10
9	0.0142	3	1-9-10
10	0.0158	3	1-5-7

Después, se realizó el proceso de depuración de las variables de entrada, dejando solo las variables cuyos pesos fueron superiores a la mitad de la media, y se logró una disminución del error en prueba de la red (ECM) de 0.046870 a 0.000195. Finalmente, la red neuronal utilizada para las estimaciones de los precios, con 3, 3 y 7 neuronas respectivamente en cada capa oculta, tuvo solo como variables de entrada las que se conservaron en el proceso de depuración.

Los resultados de error de prueba de la red no tuvieron un aumento considerable frente a los resultados de error de entrenamiento, de 0.000176 a 0.000195, lo que hace ver que el modelo pudo hacer predicciones precisas con entradas que no habían sido usadas para su entrenamiento y por lo que puede inferirse que el modelo logró una buena generalización de los datos.

Por lo anterior, podría asumirse que el modelo tiene la habilidad de estimar correctamente el precio de inmuebles por fuera de las bases de datos iniciales, teniendo como variables de entrada sus características físicas y otros datos definidos.

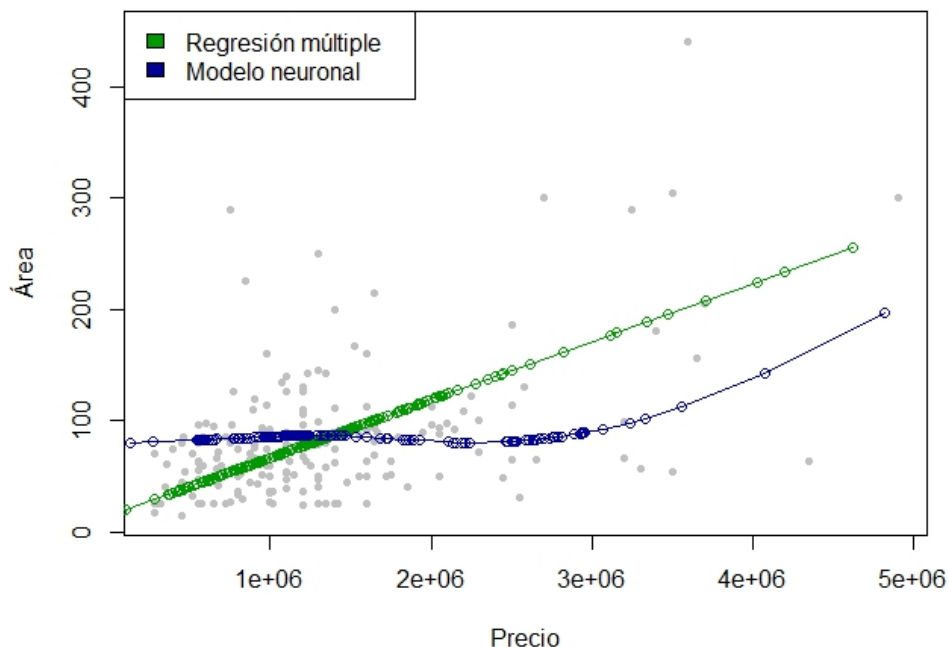
Adicionalmente, corresponde validar los resultados del modelo frente a los resultados obtenidos por una regresión múltiple, teniendo en cuenta que la última es una práctica comúnmente usada en el sector.

3.2.1 Validación con regresión múltiple

El modelo neuronal elegido, que fue entrenado con la misma base de datos que fue usada para la estimación de la regresión, logró obtener un menor error de prueba al obtenido por la regresión. Sin embargo, es importante analizar las relaciones que establecieron ambos modelos entre las variables de entrada y la de salida para ver las bondades del modelo neuronal.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Figura 9. Ajuste de los modelos.



Como puede verse en la figura 9, que contiene la relación entre el área y el precio de arrendamiento de los inmuebles, la relación entre estas variables no es lineal. Sin embargo, el ajuste realizado por la regresión múltiple subestima el precio de inmuebles con menores áreas y sobrestima los de inmuebles con áreas mayores. En comparación, el modelo neuronal hizo un ajuste polinómico de la relación de las variables y debido a eso puede reflejar esa relación de una manera más cercana a la realidad, pudiendo estimar con mayor exactitud los precios para inmuebles con cualquier área, con menores posibilidades de sub o sobrestimarlos.

En cuanto a los resultados de error obtenidos por los modelos, la comparación de las estimaciones de ambos con una misma base de datos de prueba que no fue utilizada para su formación es útil para validar la calidad de generalización. Como es visible en la tabla 3, que hace una comparación entre los errores de prueba del modelo neuronal y la regresión, la media de los errores del modelo neuronal es considerablemente inferior a la del modelo de regresión. El rango en el que se encuentran los errores de la regresión (0.3021999) es mucho mayor al del modelo neuronal (0.0026850), lo que implica que, con ese modelo, se obtuvieron algunas estimaciones más lejanas a la realidad. Además, el cuartil 3 de los errores del modelo neuronal es incluso menor que el cuartil 1 de los errores de la regresión, lo que indica que la mayoría de los errores del modelo neuronal son menores a la mayoría de los errores de la regresión.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Tabla 5. Comparación de errores de los modelos.

	Modelo neuronal	Regresión
Mínimo	-	0.0000001
Máximo	0.0026850	0.3022000
Media	0.0000195	0.0033750
Cuartil 1	0.0000004	0.0008617
Cuartil 3	0.0000024	0.0122800

Podría entonces decirse que la capacidad de generalización del modelo neuronal es superior a la de la regresión para los datos utilizados en esta investigación y que tiene una habilidad mayor a la de la regresión múltiple para estimar precios de inmuebles que fueron ajenos a su entrenamiento.

3.3 MODELO DE FLUJO DE CAJA DESCONTADO

Inicialmente, se exponen los resultados del cálculo de variables de entrada de importancia para el modelo. El Beta estimado entre el mercado inmobiliario y accionario colombiano fue calculado con el “Índice de Precios de la Vivienda Usada” (IPVU) y el índice accionario COLCAP, el resultado de fue de 0.043, por lo que el costo del patrimonio usado, es el rendimiento de los bonos del tesoro colombiano con vencimiento en 10 años (asumida tasa libre de riesgo) más 0.043 veces el exceso de la rentabilidad del índice COLCAP sobre la tasa libre de riesgo, lo que es 5.93% EA.

La tasa promedio de colocación de créditos de adquisición de vivienda, calculada a partir de la información de cada entidad financiera que reporta la Superintendencia Financiera de Colombia, fue del 11.22%. Considerando esta tasa como costo de la deuda para el modelo y en combinación con el costo del patrimonio hallado, ponderando ambas como 70% deuda y 30% patrimonio, el costo de capital encontrado y, por lo tanto, la tasa con la que fueron descontados los flujos netos calculados de cada inmueble, es de 9.63%.

En la comparación del modelo de valoración con flujos de caja descontados frente a las valoraciones de mercado, el porcentaje de sobrevaloración de cada inmueble es calculado como el porcentaje que representa la diferencia entre la valoración de mercado presente en la base de datos y la valoración arrojada por el modelo de flujo de caja descontado. Además, en los casos en los que hay sobrevaloración, se valida si el límite establecido para considerar que la sobrevaloración es injustificada y no obedece simplemente a imperfecciones del mercado fue superado.

El número de casos en las que la diferencia entre ambos valores supera el límite establecido es de 23 de los 53 inmuebles valorados, la proporción que lo supera, representa el 43.40% de las valoraciones totales. Este resultado, que no supera la mitad de los inmuebles valorados, no es claro en cuanto a la presencia de una burbuja de precios en el mercado.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

A partir de los resultados obtenidos y el análisis realizado sobre las burbujas de precios puede resaltarse además que son un fenómeno temporal, por lo que es adecuado analizarlo con datos de diferentes períodos de tiempo y no con información longitudinal como la que se dispuso en esta investigación. Además de las limitaciones por la única temporalidad de los datos, el origen privado e individual de la información, sugiere los datos podrían ser una muestra sesgada de la población.

No se consideró posible concluir sobre la presencia o no de una burbuja de precios en el mercado porque lo impiden las limitaciones de los datos disponibles para la investigación el análisis insuficiente que pudo hacerse por el mismo motivo.

4. CONCLUSIONES Y CONSIDERACIONES FINALES

Basándose en los resultados obtenidos con el uso de la metodología propuesta es posible sacar algunas conclusiones y consideraciones de la investigación.

Inicialmente, pudo observarse que existen algunas características de los inmuebles de la base de datos que pueden llevar a particularidades del modelo y limitar su alcance a la valoración de inmuebles con esas características particulares, entre las que están la concentración de inmuebles en estratos altos, destinados al uso de vivienda, en la ciudad de Medellín.

Pudo verificarse que el método de elección del número de neuronas en cada capa oculta del modelo, como el número de neuronas de mayor frecuencia en cada una de las diez repeticiones realizadas, no fue efectivo. La configuración con el menor error, que se eligió para hacer las estimaciones y validaciones del modelo neuronal, se obtuvo en una de las redes neuronales entrenada en las repeticiones y no correspondió a la configuración inicialmente elegida.

Adicionalmente, el método de depuración de las variables de entrada, removiendo las variables con pesos inferiores, fue efectivo, ya que logró disminuciones considerables en el error de entrenamiento de la red neuronal y permitió que tuviera una mejor generalización.

Los métodos de entrenamiento utilizados para la elección de la configuración del modelo neuronal lograron disminuir los errores de estimación de precios, la depuración de las variables de entrada con menores pesos logró la mayor disminución de esos errores en validaciones con información no usada en el entrenamiento del modelo. Se recomienda explorar la utilización de otros métodos de elección de configuraciones para el entrenamiento de redes neuronales que disminuyan los errores de estimación del modelo.

Con los resultados obtenidos, se concluye que el uso de modelos neuronales tiene beneficios frente a otros métodos de valoración como las regresiones, por el mejor ajuste de las relaciones entre variables que no tienen una relación lineal con la variable de salida y la mejor generalización del modelo, evidenciada gráficamente en los ajustes de los modelos y en los errores con la estimación del precio de inmuebles de una base de datos de prueba que no fue usada para el entrenamiento.

Finalmente, no fue posible concluir sobre la presencia de una burbuja de precios en el mercado debido a que los resultados del modelo no lo indicaron y adicionalmente debido a limitaciones impuestas por la disponibilidad de los datos para esta investigación. Se deben tener presentes algunas consideraciones para futuras investigaciones sobre el tema en cuanto a la complejidad del modelo de valoración y la disponibilidad de información en varios períodos de tiempo.

En estudios futuros se recomienda trabajar en la elección de una base de datos que contenga inmuebles que sean representativos para todas las zonas de la ciudad de Medellín y municipios aledaños, pues por la elección de la muestra expuesta anteriormente, se presenta un sesgo de selección. Adicionalmente, se recomienda analizar las burbujas de

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

precios con información de varios períodos de tiempo y con modelos de panel. Por su naturaleza temporal, no es apropiado estudiarlas con bases de datos longitudinales como la dispuesta para esta investigación.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

5. BIBLIOGRAFÍA

- Aristizábal Restrepo, M. C. (2006). Evaluación asimétrica de una Red Neuronal Artificial: aplicación al caso de la inflación en Colombia, 65(65), 73–116. Retrieved from <http://www.scielo.org.co/pdf/le/n65/n65a3>
- Blanchard, O., & Watson, M. (1982). Bubbles, Rational Expectations and Financial Markets, (July 1983). <http://doi.org/10.3386/w0945>
- Casas Fajardo, A. E. (2014). Propuesta metodológica para calcular el avalúo de un predio empleando redes neuronales artificiales, 156. Retrieved from <http://www.bdigital.unal.edu.co/49434/>
- Casas Sánchez José M. (2017). Inferencia estadística para economía y administración de empresas. Editorial Centro de Estudios Ramón Areces, S.A.
- Cediel Sánchez, V., & Velásquez Vega, C. (2015). ¿Hay una burbuja inmobiliaria en bogotá? Un estudio por segmentos de Mercado. Revista de Economía Institucional, 17(32), 233–257. <http://doi.org/10.18601/01245996.v17n32.08>
- Ceular, N., Caridad, J. M., & Núñez, J. M. (2008). Metodología de Precios Hedónicos vs. Redes Neuronales Aritificiales como alternativas a la valoración de Inmuebles. Un caso real. Catastro, (Abril), 27–42.
- Chaphalkar, N. B., & Sandbhor, S. (2013). Use of artificial intelligence in real property valuation. International Journal of Engineering and Technology, 5(3), 2334–2337.
- Chiarazzo, V., Caggiani, L., Marinelli, M., & Ottomanelli, M. (2014). A neural network based model for real estate price estimation considering environmental quality of property location. Transportation Research Procedia, 3(July), 810–817. <http://doi.org/10.1016/j.trpro.2014.10.067>
- Clavijo, S., & Ortiz, M. C. (2016). CICLO HIPOTECARIO EN COLOMBIA Y SU AJUSTE 2015-2017.
- Damodaran, A. (2012). En Investment Valuation: Tools and Techniques for Determining the Value of Any Asset, Third Edition. John Wiley & Sons.
- Dinero, R. (2014). ¿Burbuja hipotecaria en Colombia? Retrieved from <http://www.dinero.com/economia/articulo/burbuja-hipotecaria-colombia/194507>
- Dinero, R. (2017). Hasta dónde llegará el boom de las titularizaciones inmobiliarias.
- El Espectador, P. (2016). Fenalco dice que el 45,9% de los hogares colombianos tienen casa propia. Redacción Negocios Y Economía. Retrieved from <https://www.elespectador.com/noticias/economia/fenalco-dice-el-459-de-los-hogares-colombianos-tienen-c-articulo-637734>

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

- Ernst & Young. (2013). Applying IFRS in real estate, IFRS 13 Fair value measurement: Fair value implications for the real estate sector and example disclosures for real estate entities, (January), 37. Retrieved from <http://bit.ly/2fhB4ot>
- Ferrari, C., & Javeriana, U. (2013). Burbuja inmobiliaria - Columnista invitado. Retrieved from <https://www.youtube.com/watch?v=e9VA28674BQ>
- Fritsch Stefan, Guenther Frauke, Wright Marvin, Suling Marc, Mueller Sebastian (2019). neuralnet: Training of neural networks. R package version 1.44.2. <https://cran.r-project.org/web/packages/neuralnet/index.html>
- Garcia, N., Gamez, M., & Alfaro, E. (2006). Redes neuronales artificiales: un largo e irregular camino hasta la actualidad. En H. d. estadística.
- Gil Serna, J. G. (2015). Estimación de un pronóstico de exportaciones de café suave colombiano. Redes neuronales artificiales y ARDL : ene-dic 2012. Tesis de Maestría. Universidad EAFIT.
- González, L., Guerrero, A., Delvastro, S., & Will, A. (2012). Red neuronal artificial para estimar la resistencia a compresión, en concretos fibro-reforzados con polipropileno. Ventana Informativa, 26, 11–28. Retrieved from <http://revistasum.umanizales.edu.co/ojs/index.php/ventanainformatica/article/viewFile/217/267>
- González Salcedo, L. O., Guerrero Zúñiga, A. P., Arjona, S. D., & Will, A. L. E. (2012). Exploración con redes neuronales artificiales para estimar la resistencia a la compresión, en concretos fibroreforzados con acero. Universidad Militar Nueva Granada.
- IFRS Foundation. (2013). International Financial Reporting Standard 13: Fair Value Measurement, 1(c), 1–28.
- IVSC. (2011). International Valuation Standards (2011). Ivs.
- Kindleberger, C. P. (2016). Bubbles in History. Palgrave Macmillan.
- Jain, A., Nandakumar, K., & Ross, A. (2005). Score normalization in multimodal biometric systems. Pattern Recognition, 2270-2285.
- Mimis, A., Rovolis, A., & Stamou, M. (2013). Property valuation with artificial neural network: The case of Athens. Journal of Property Research, 30(2), 128–143. <http://doi.org/10.1080/09599916.2012.755558>
- Ministerio de Hacienda y Crédito Público. (2019). Marco Fiscal de Mediano Plazo 2019.
- Ministerio de Vivienda Ciudad y Territorio. (2017). Actividades inmobiliarias puntean en generación de empleo en Colombia. Retrieved from <http://www.minvivienda.gov.co/sala-de-prensa/noticias/2017/marzo/actividades->

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

inmobiliarias-puntean-en-generacion-de-empleo-en-colombia

- Mirjadili, S. (2015). How effective is the Grey Wolf optimizer in training multi-layer perceptrons. *Applied Intelligence*.
- Peterson, S., & Flanagan, A. B. (2009). Neural Network Hedonic Pricing Models in Mass Real Estate Appraisal. *Journal of Real Estate Research*, 31(2), 147–165. Retrieved from <http://ares.metapress.com/index/M3H27210W6411373.pdf>
- R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rosser, J. B. (2000). *From Catastrophe to Chaos: A General Theory of Economic Discontinuities (Volume I)*.
- Santana, J. C. (2006). Predicción de series temporales con redes neuronales: Una aplicación a la inflación Colombiana. *Revista Colombiana de Estadística*, 29(1), 77–92.
- Steyerberg E.W. (2019) Coding of Categorical and Continuous Predictors. In: *Clinical Prediction Models. Statistics for Biology and Health*. Springer, Cham
- Torres Ramírez, J. E. (2012). Estudio sobre el mercado de arrendamiento de vivienda en Colombia.
- Womack, K., & Zhang, Y. (2003). Understanding Risk and Return, the CAPM, and the Fama-French Three-Factor Model.

6. ANEXOS

6.1 ANEXO 1

```
for (s in 1:repeticiones){  
  
  # Separa la base de datos en una de entrenamiento y otra de test.  
  filas_train<-sample(1:nrow(bd_scaled),round(0.6*nrow(bd_scaled)))  
  bd_train<-bd_scaled[filas_train,1:ncol(bd_scaled)]  
  bd_test<-bd_scaled[-filas_train,1:ncol(bd_scaled)]  
  
  #[Ciclos con una neurona]  
  #[Ciclos con dos neuronas]  
  #[Ciclos con tres neurona]  
  
}
```

6.2 ANEXO 2

```
# Ciclo con una capa oculta con neuronas entre 1 y 10.  
ek2_train_una<-matrix(0,n,1)  
ek2_test_una<-matrix(0,n,1)  
for (j in 1:n){  
  mse_train<-NULL  
  mse_test<-NULL  
  print(j)  
  vc<-as.vector(c(as.integer(j)))  
  nn<-neuralnet(formula,as.data.frame(bd_train),vc,stepmax=10000,err.fct="sse",act.fct="logistic",linear.output = TRUE)  
  # Error de entrenamiento.  
  train_nn<-compute(nn,bd_train[,2:ncol(bd_train)])  
  train_nn<-train_nn$net.result  
  mse_train<-sum((as.data.frame(bd_train[,1])-train_nn)^2)/nrow(train_nn)  
  # Error de test.  
  test_nn<-compute(nn,bd_test[,2:ncol(bd_test)])  
  test_nn<-test_nn$net.result  
  mse_test<-sum((bd_test[,1]-test_nn)^2)/nrow(test_nn)  
  # Guardar ek2 en matriz.  
  ek2_train_una[j,1]<-mse_train  
  ek2_test_una[j,1]<-mse_test  
}
```

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

6.3 ANEXO 3

```
# Ciclo con dos capas ocultas con neuronas entre 1 y 15 en cada una.
ek2_train_dos<-matrix(0,n,n)
ek2_test_dos<-matrix(0,n,n)
for (j in 1:n){
  for (m in 1:n){
    mse_train<-NULL
    mse_test<-NULL
    print(paste(j," - ",m))
    vc<-as.vector(c(as.integer(j),as.integer(m)))
    nn<-neuralnet(formula,as.data.frame(bd_train),vc,stepmax=10000,err.fct="sse",act.fct="logistic",linear.output = TRUE)
    # Error de entrenamiento.
    train_nn<-compute(nn,bd_train[,2:ncol(bd_train)])
    train_nn<-train_nn$net.result
    mse_train<-sum((as.data.frame(bd_train[,1])-train_nn)^2)/nrow(train_nn)
    # Error de test.
    test_nn<-compute(nn,bd_test[,2:ncol(bd_test)])
    test_nn<-test_nn$net.result
    mse_test<-sum((bd_test[,1]-test_nn)^2)/nrow(test_nn)
    # Guardar ek2 en matriz.
    ek2_train_dos[j,m]<-mse_train
    ek2_test_dos[j,m]<-mse_test
  }
}
```

6.4 ANEXO 4

```
ek2_train_tres<-array(rep(0,n^3),c(n,n,n))
ek2_test_tres<-array(rep(0,n^3),c(n,n,n))
for (j in 1:n){
  for (m in 1:n){
    for (k in 1:n){
      mse_train<-NULL
      mse_test<-NULL
      print(paste(j," - ",m," - ",k))
      vc<-as.vector(c(as.integer(j),as.integer(m),as.integer(k)))
      nn<-neuralnet(formula,as.data.frame(bd_train),vc,stepmax=10000,err.fct="sse",act.fct="logistic",linear.output = TRUE)
      # Error de entrenamiento.
      train_nn<-compute(nn,bd_train[,2:ncol(bd_train)])
      train_nn<-train_nn$net.result
      mse_train<-sum((as.data.frame(bd_train[,1])-train_nn)^2)/nrow(train_nn)
      # Error de test.
      test_nn<-compute(nn,bd_test[,2:ncol(bd_test)])
      test_nn<-test_nn$net.result
      mse_test<-sum((bd_test[,1]-test_nn)^2)/nrow(test_nn)
      # Guardar ek2 en matriz.
      ek2_train_tres[j,m,k]<-mse_train
      ek2_test_tres[j,m,k]<-mse_test
    }
  }
}
```

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.