



UNIVERSIDAD NACIONAL DE COLOMBIA

Propuesta Metodológica para Calcular el Avalúo Catastral de un Predio Utilizando Redes Neuronales Artificiales

ADRIANA EMILCEN CASAS FAJARDO
Ingeniera Agrónoma

Universidad Nacional de Colombia
Facultad de Agronomía, Maestría en Geomática
Bogotá D.C., Colombia
2014

Propuesta Metodológica para Calcular el Avalúo de un Predio Empleando Redes Neuronales Artificiales

ADRIANA EMILCEN CASAS FAJARDO
Ingeniera Agrónoma

Tesis o trabajo de investigación presentado como requisito parcial para optar al título de:
Magister en Geomática

Director:

MSc. Alberto Boada Rodríguez

Línea de Investigación:
Geoinvestigación para el uso sostenible de los recursos naturales

Universidad Nacional de Colombia
Facultad de Agronomía, Maestría en Geomática
Bogotá D.C., Colombia
2014

Hay momentos en la vida en que se siente que tu camino se termina, y todo parece oscuro, pero de repente te das cuenta que no estás solo que estas rodeado de seres maravillosos que te extienden su mano para no dejarte sucumbir

Dedico este trabajo de grado a:

A Dios, porque me mostró el camino y me dio fortaleza en los momentos más duros.

A mi hijo Christian, por su amor, apoyo permanente, porque más que un hijo, es un amigo y un ser excepcional.

A mi Paulita, por ser la mejor hija que la vida pudo darme, que siempre me transmitió valor, por sus cuidados y ternura permanente.

A mi madre por ser mi fortaleza

A Matías, por ser el peor gato del mundo, pero el mejor guardián de mis largos desvelos

Agradecimientos

De manera muy especial doy mis agradecimientos a:

Al Instituto Geográfico Agustín Codazzi, por darme la oportunidad de hacer esta Maestría y por permitirme el uso de la información para efectos del desarrollo de la presente tesis.

Al Doctor Alberto Boada Rodríguez, Director de la presente tesis, por su apoyo, confianza y valiosos aportes en el desarrollo de este trabajo.

A mi jefe, el doctor Edgar Santiago Benítez Acevedo, por su apoyo, comprensión y por esa calidad humana que difícilmente volveré a encontrar en otro ser humano.

A mi querida amiga María Cristina Giraldo Uribe por su apoyo, consejos y palabras de aliento constantes para lograr culminar este trabajo de grado.

Resumen

El presente estudio tiene como objetivo la determinación del avalúo de un predio urbano empleando la metodología de Redes Neuronales Artificiales (RNA) que es el método econométrico menos tradicional del campo de la inteligencia artificial. Para el desarrollo del modelo se usó las variables que el IGAC emplea tradicionalmente para el cálculo de los avalúos catastrales y se incorporaron variables de ubicación espacial de los predios apoyados en un Sistema de Información Geográfica. Los datos que se emplearon para la estimación del modelo fueron obtenidos en la investigación de mercado directa e indirecta realizada en el proceso de Actualización de la Formación Catastral de la zona urbana del Municipio de Fusagasugá que entró en vigencia el primero de enero de 2013. Para obtener el mejor modelo de RNA, en el que se seleccionen las variables que más contribuyen a la determinación del valor de un inmueble, se realizaron, numerosas pruebas, utilizando el algoritmo Perceptrón Multicapa de tipo supervisado MLP y el software estadístico SPSS V. 21. Las variables explicativas incluidas en el modelo final para calcular el avalúo catastral de un predio han sido: la Área Terreno, Área C Privada, USOPOT, X Centroide, Distancia, Comuna, Puntaje Estructura, Puntaje Total, No Pisos, No Habitaciones, Edad, Y Centroide, Puntaje Cocina, Puntaje Acabados, Puntaje Baños y No Baños. La prueba paramétrica t de *Student* para datos pareados o muestras relacionadas, corroboró el mayor poder de la Red Neuronal Artificial para estimar los avalúos respecto de resultados obtenidos con la metodología del IGAC. Tras esto, se considera que las variables utilizadas para caracterizar el valor de un predio en el municipio de Fusagasugá, fácilmente se pueden adaptar a otras ciudades de Colombia.

Palabras Clave: Redes Neuronales Artificiales, variables de localización, Sistema de Información Geográfica, avalúos.

Abstract

The present study aims at determining the valuation of an urban property using the methodology of Artificial Neural Networks (ANN) which is the least traditional econometric method in the field of artificial intelligence. To develop the model variables IGAC traditionally used to calculate the assessed valuations of variables and spatial location of the properties supported by a Geographic Information System was used were included. The data were used to estimate the model were obtained in the investigation of direct and indirect market made in the process of training the Cadastral Updating the urban area of the Municipality of Fusagasugá which came into force on January 2013. For the best model RNA, wherein the variables that contribute to determining the value of a property is selected, were conducted extensive testing using the algorithm type supervised Multilayer Perceptron MLP and statistical software SPSS V. 21. The explanatory variables included in the final model to calculate the assessed value of a property have been: Land Area, Private C, USOPOT, Centroid X, Range, City, Structure Points, Total Points, N° Flats, N° Rooms, Age, and Centroid Points Kitchen, Finished Score, Score Bathrooms and N° Bathrooms. The parametric Student's t test for paired data or related samples, confirmed the greater power of Artificial Neural Network to estimate the relative valuations of results obtained with the methodology of IGAC. After this, it is considered that the variables used to characterize the value of a property in the town of Fusagasugá, can easily be adapted to other cities of Colombia.

Keywords: Artificial Neural Networks, location variables, Geographic Information System, appraisals.

Tabla de contenido

	Pág.
Resumen.....	..VII
Lista de figuras.....	XII
Lista de tablas.....	.XIV
Lista de Ecuaciones.....	XV
Introducción.....	17
Objetivos.....	21
Objetivo General.....	21
Objetivos Específico.....	21
1. El Catastro Colombiano.....	.22
1.1 Definición	22
1.2 Aspectos del catastro.....	23
1.2.1 Aspecto Físico.	23
1.2.2 Aspecto Jurídico.....	23
1.2.3 Aspecto Fiscal.	23
1.2.4 Aspecto Económico.....	24
1.3 Actividades Catastrales.....	25
1.3.1 Formación.....	25
1.3.2 Actualización de la Formación.....	26
1.3.3 Conservación Catastral.....	26
1.4 Importancia del catastro.....	26
2. Métodos de Valoración de inmuebles.....	28
2.1 Métodos de valoración de acuerdo con la Colombiana.	28
2.1.1 Método de Comparación de Mercado.....	28

2.1.2	Método de Costo de Reposición	29
2.1.3	Método – Técnica Residual	30
2.1.4	Método de Capitalización de Rentas o Ingresos	31
2.2	El Avalúo Catastral en Colombia	32
2.2.1	Normatividad que rige los avalúos Catastrales	32
2.2.2	Metodología de Zonas Homogéneas empleada por el IGAC para la determinación del Avalúo Catastral	34
2.2.3	Determinación de los valores unitarios para las construcciones	39
3.	Revisión de Literatura	44
3.1	Redes Neuronales Artificiales	44
3.1.1.	Características de las Redes Neuronales Artificiales	45
3.1.2	Estructura de una Red Neuronal Artificial	48
3.1.3	La activación de las neuronas y las funciones de activación	50
3.1.4	Proceso de aprendizaje y entrenamiento de una RNA	52
3.1.6	Clasificación de los modelos de RNA	54
3.1.7	Las redes Perceptrón multi-capas	56
3.2	Análisis previo de los datos antes de ser procesados por la RNA	56
3.3	Uso de modelos de redes neuronales artificiales para la estimación del avalúo de inmuebles.	62
3.4	Modelos de ubicación de los Inmuebles	67
3.4.1	Modelo Monocéntrico	69
3.4.2	El Stock de Capital No Maleable	71
3.4.3.	El modelo de Tiebout	72
3.5	Aplicación de la Geomática para el cálculo de las variables de localización.	74
4.	Metodología	77
4.1.	Municipio de Fusagasugá – Caso de Estudio	77
4.1.1	Entorno físico y geográfico	77
4.1.2	División político administrativa	80
4.1.3	Información Predial	80

4.2 Materiales.....	81
4.3 Métodos.....	82
4.3.1 Constitución de la muestra.....	82
4.3.2 Análisis previo de los datos antes de ser procesados por la RNA	84
5. Análisis de Resultados.....	98
5.1 Análisis Exploratorio de las Variables	98
5.1.1 Variables Físicas del Predio.....	98
5.1.2 Variables de Localización	107
5.2 Análisis de Componentes Principales.....	110
5.3 Estimación de la Red Neuronal.....	114
5.4 Comparación de los Avalúos obtenidos con Metodología del IGAC y la Red Neuronal. .	134
Conclusiones	139
Recomendaciones.....	143

Lista de figuras

	Pág.
Figura 1. Flujograma del proceso para determinar el avalúo catastral de un predio en Colombia.	43
Figura 2. Diagrama de una red neuronal artificial.....	49
Figura 3. La activación de una neurona.....	50
Figura 4. Fases de pre-procesamiento y post-procesamiento en una red neuronal.....	58
Figura 5. Teoría de las Zonas Concéntrica.....	70
Figura 6. Ubicación general municipio de Fusagasugá.....	79
Figura 7. Distribución de predios por rango de avalúos zona urbana municipio de Fusagasugá.	81
Figura 8. Plano de conjunto urbano Municipio de Fusagasugá cálculo de distancia al centro de la ciudad.....	86
Figura 9. Plano de conjunto urbano Municipio de Fusagasugá con espacialización de Usos permitidos POT.....	88
Figura 10. Plano de conjunto urbano Municipio de Fusagasugá con División político Administrativa de comunas Municipio Fusagasugá.....	89
Figura 11. Plano urbano con distribución espacial de predios de la muestra, fuente IGAC.....	91
Figura 12. Flujograma de procesamiento de redes neuronales de la red Perceptrón Multicapa....	93
Figura 13. Flujograma de la metodología para calcular el avalúo catastral de un predio utilizando RNA.....	97
Figura 14. Esquema de la red neuronal (<i>salida de SPSS</i>) - Fuente: <i>Elaboración propia en el SPSS</i>	123
Figura 15. Red Neuronal Artificial. Parte A, donde se muestra variables cualitativas, COMUNA y USOPOT.....	124
Figura 16. Red Neuronal Artificial. Parte A, donde se muestra variables cuantitativas.....	125

Figura 17. Importancia relativa de las variables utilizadas en la estimación de la red neuronal ..	126
Figura 18. Precio observado vs el precio estimado por la red neuronal	127

Lista de tablas

	Pág.
Tabla 1. RNA según tipo de aprendizaje	53
Tabla 2. Límites Municipio Fusagasugá.....	77
Tabla 3. Distribución de Comunas Municipio de Fusagasugá.....	80
Tabla 4. Relación de los materiales empleados en la investigación	81
Tabla 5. Relación de Usos del suelo encontrados con codificación.....	87
Tabla 6. Relación de Comunas del Municipio de Fusagasugá y la conversión a variable cuantitativa	87
Tabla 7. Estadísticos Descriptivos Análisis de Componentes Principales	110
Tabla 8. Matriz de Correlaciones	111
Tabla 9. Comunalidades de Análisis de Componentes Principales	112
Tabla 10. Varianza total explicada.....	113
Tabla 11. Tabla Matriz de Componentes ACP	113
Tabla 12. Opciones de Entrenamiento de la Red.....	117
Tabla 13. Información sobre la red	120
Tabla 14. MAPE y RMSE.....	121
Tabla 15. Resultado de la red neuronal en términos de precisión (<i>salida</i> de SPSS).....	122
Tabla 16. Importancia relativa de las variables utilizadas en la estimación de la red neuronal	127
Tabla 17. Relación de Variables incorporadas al modelo	128
Tabla 18. Parámetros de la Red Neuronal	129
Tabla 19. Características del predio No 71 que procesa la red.....	130
Tabla 20. Prueba T, comparación avalúo Comercial Vs avalúo Calculado por el método de Zonas	134
Tabla 21. Prueba T, comparación avalúo comercial Vs avalúo calculado por la red neuronal	135

Lista de Ecuaciones

	Pág.
Ecuación 1. Fórmula para Calcular Valor Comercial de un Inmueble	29
Ecuación 2. Fórmula para obtener el Valor de Terreno de un Bien	30
Ecuación 3. Fórmula para obtener el Valor Comercial de un bien	31
Ecuación 4. Promedio del Valor Investigado (μ)	36
Ecuación 5. Desviación Estándar δ	37
Ecuación 6. Coeficiente de Variación (CV)	37
Ecuación 7. Cálculo de intervalo de confianza	38
Ecuación 8 . Energía de activación de la neurona.....	50
Ecuación 9. Función Logística	51
Ecuación 10. Función Hiperbólica.....	51
Ecuación 11. Distancia de Mahalanobis	57
Ecuación 12. Modelo de compensación	70

Introducción

El patrimonio inmobiliario tiene una importancia fundamental en la vida económica de cualquier país. Como expresión de riqueza, presupone una determinada capacidad económica fácil de cuantificar y difícil de ocultar; por lo tanto, avaluar una propiedad se convierte en una necesidad. Es de interés para el comprador, el vendedor y el Gobierno Nacional conocer el valor real de un predio en particular, pues la propiedad inmobiliaria está sujeta al pago de impuestos en casi todos los países del mundo. Al basarse el impuesto en el valor del avalúo catastral, se hace indispensable que la valoración sea la correcta y ajustada a los precios del mercado inmobiliario, pues la equidad del impuesto predial depende del cálculo preciso y seguro del valor del inmueble.

Inicialmente el avalúo catastral solo se relacionaba con los fines tributarios, pues por norma éste constituye la base gravable del Impuesto Predial Unificado (Ley 44, 1990); circunstancia que ha venido cambiando con el tiempo, y hoy el avalúo también es empleado en otras actividades oficiales, tales como la liquidación del costo de la libreta militar, para el ingreso a las universidades públicas, asignación de subsidios y trámites ante la Oficina de Notariado y Registro, entre otros. Esta nueva concepción de mayor visibilidad en los avalúos catastrales, ha generado los cuestionamientos cada vez más frecuentes sobre si los mismos realmente se ajustan o no a las características y condiciones reales de los inmuebles, situación que en últimas conlleva al interrogante sobre la precisión de la metodología que el IGAC viene empleado para obtenerlos.

En Colombia la entidad encargada de generar la normatividad para estimar los avalúos catastrales es el Instituto Geográfico Agustín Codazzi (IGAC). De acuerdo con el marco legislativo de esta Institución, en el proceso de Actualización de la Formación Catastral es donde se renueva la información de la Formación Catastral, en la que se encuentra el avalúo catastral, el cual no podrá ser inferior al 60 % del respectivo valor comercial del inmueble, sin llegar a superar este último (Ley 1450, 2011), lo cual está previsto en el Plan de Desarrollo del Gobierno Nacional “Prosperidad para todos, 2.010 - 2.014”.

Con esta normatividad, lo que pretende el Gobierno Nacional es que los avalúos catastrales se aproximen a la realidad del mercado inmobiliario, y de esta manera no coexistan dos

avalúos, uno que se utilice para vender, y el otro para tributar. Obviamente, lograr que los avalúos catastrales estimados sean muy cercanos al valor comercial pero sin llegar a superarlo, requiere de una metodología de valoración de mayor precisión que la empleada por el IGAC, pues actualmente, la autoridad catastral opta por aplicar bajos porcentajes sobre los valores comerciales investigados, es decir al límite de lo normado (60%), para garantizar a los ciudadanos que los avalúos catastrales en ningún momento superen los valores comerciales, y de esta manera reducir los niveles de reclamación de los propietarios y poseedores, máxime si se trata de avalúos masivos como lo son los resultantes de los procesos de Actualización Catastral.

Es precisamente en este proceso de reclamación y de revisión del avalúo en donde se hace evidente que de las variables levantadas por el Catastro, la más subjetiva es el avalúo fijado al predio. Completar la base de datos catastral estableciendo un valor al inmueble, el cual se ha identificado física y jurídicamente, es una de las actividades más complejas, y la causa que sustenta esta afirmación, es la innegable carga subjetiva que tiene todo proceso de valoración, que no puede eliminarse por muy reglado que sea el procedimiento desde el punto de vista administrativo.

La metodología de valoración masiva con la que actualmente se estiman los avalúos catastrales fue diseñada por el IGAC desde el año de 1984, es decir, que en el presente año se cumplen 30 años de su aplicación. Es lógico que cuando se elaboró dicha metodología, se tuvo en cuenta las condiciones constructivas e inmobiliarias imperantes en la época, cuando predominaba la construcción de casas individuales con terrenos propios no sometidas al régimen de Propiedad Horizontal; modelo que en la actualidad se ha reducido considerablemente (caso de los Municipios de Soacha, Chía, Cajicá y Tocancipá, entre otros), dada la escasez del suelo y la apremiante necesidad de abastecer el déficit de vivienda por el que atraviesan estos Municipios, situación que ha motivado que en la mayor parte de ellos hoy exista un predominio de la construcción en altura.

Para el caso de Colombia, la Ley señala que el avalúo catastral se obtiene mediante investigación y análisis estadístico del mercado inmobiliario, y se determina mediante la adición de los avalúos parciales practicados independientemente para los terrenos y para las edificaciones en él comprendidas (Ley 14, 1983). La estimación del valor del terreno se realiza mediante la metodología de zonas homogéneas, y para las construcciones se efectúa a través de modelos de regresión simples, ya sean lineales, exponenciales, logarítmicas o potenciales. La división del valor integral de una propiedad entre terreno y construcción, con frecuencia genera resultados poco coherentes, pues terrenos con diferente localización o con diferente forma dentro de una área homogénea, deberían tener diferentes precios, pero en la metodología de zonas homogéneas se les asigna el mismo valor, y en el caso de las

propiedades horizontales, esta metodología obliga a asignarle a la construcción el efecto de las diferencias de características particulares de cada predio dentro de la propiedad horizontal.

Ante la dificultad que se tiene para hacer una valoración masiva de los inmuebles más ajustada a la realidad inmobiliaria de hoy, se hace necesario que el IGAC ajuste o complemente la metodología que viene aplicando de tiempo atrás, explorando nuevos modelos de valoración e incorporando nuevas variables que permitan obtener avalúos de mayor exactitud y congruentes con la realidad del mercado inmobiliario.

Es así como el presente Estudio propone desarrollar una metodología basada en Redes Neuronales Artificiales (RNA) para calcular el avalúo catastral de un predio, utilizando las variables que se encuentran disponibles en el Sistema de Información Catastral. Esto, por considerar que es una adecuada y útil herramienta para el tratamiento de la información, que complementa las técnicas estadísticas tradicionales, resolviendo problemas que éstas no podían solucionar, o bien, en ciertos casos, mejorando los resultados proporcionados por ellas.

Para el caso del mercado inmobiliario, las Redes Neuronales Artificiales han mostrado que poseen una gran ventaja respecto de la Regresión Múltiple, y es la capacidad que tienen para estimar el valor de las propiedades que presentan características significativamente diferentes de las de otros bienes de su entorno (avalúos extremos, atípicos o outliers); lo anterior, si tenemos en cuenta que el mercado presenta gran heterogeneidad por la cantidad de características distintas que pueden registrar los predios en aspectos como su localización, servicios públicos y privados, comunicaciones, zonas verdes, entre otros. Lo cual obedece a que los Sistemas de Inteligencia Artificial someten las muestras a un proceso matemático mucho más complejo que el de la Regresión Múltiple, que se limita a construir una simple ecuación polinómica (Gallego, 2004).

De igual forma sucede con las variables que el Catastro actualmente aplica para la estimación del avalúo de un predio, las cuales consecuentemente también se han quedado rezagadas frente a la complejidad de lo que es el comportamiento del mercado inmobiliario, pues aspectos relacionados con la localización del predio no son tenidos en cuenta de manera cuantitativa en el actual modelo de valoración, a pesar de la importancia que las mismas tienen al momento de determinar el valor de un inmueble, pues solo a partir de estas se logra explicar el motivo por el cual los individuos deciden instalarse en una determinada zona de la geografía urbana y no en otra, es decir, que permiten analizar cuáles factores relacionados con la ubicación, como el precio del suelo o los servicios propios de una zona específica, influyen sobre el precio final del predio.

En este contexto, la Geomática como ciencia de los Sistemas de Información Geo-espacial nos proporciona métodos y procedimientos, tales como: la cartografía, sistemas de información geográfica (SIG), sistema de posicionamiento global (GPS) y fotogrametría, de los cuales se hará uso para la estimación de las variables de localización que se pretenden introducir al modelo de RNA.

Objetivos

Objetivo General

Desarrollar una metodología basada en Redes Neuronales Artificiales para calcular el avalúo catastral de un predio utilizando las variables que se encuentran disponibles en el Sistema de Información Catastral Colombiano.

Objetivos Específicos

1. Identificar variables de localización espacial a Incluir en el modelo de Redes Neuronales para calcular el avalúo de un predio.
2. Determinar de las variables catastrales que se usan actualmente tienen mayor influencia en el valor del avalúo de un inmueble.
3. Realizar la comparación entre el método actual de valoración empleado por el IGAC con los resultados obtenidos con la aplicación de Redes Neuronales.

1. El Catastro Colombiano

1.1 Definición

El Congreso Colombiano con la expedición de la Ley de 14 de 1983, definió el catastro, como un inventario de carácter multifinanciero, con lo cual pretendía darles a las autoridades locales los instrumentos adecuados para la planificación municipal. En el Decreto reglamentario de la Ley 14, se estableció que *“el catastro es el inventario o censo, debidamente actualizado y clasificado, de los bienes inmuebles pertenecientes al Estado y a los particulares, con el objeto de lograr su correcta identificación física, jurídica, fiscal y económica”*. en materia administrativa el Catastro colombiano es desconcentrado y descentralizado desde el punto de vista de la administración pública (Ley 489, 1998) ; desconcentrado, porque existe una entidad de carácter nacional, el Instituto Geográfico Agustín Codazzi-(IGAC), que a su vez, trabaja de manera desconcentrada a través de oficinas denominadas Unidades Operativas de Catastro a lo largo de todo el territorio nacional. Descentralizado, porque en la actualidad, debido a decisiones históricas de tipo político existen cuatro oficinas catastrales que administran el catastro de sus respectivas entidades territoriales de manera independiente del IGAC¹.

Teniendo en cuenta que la palabra catastro ya no significa únicamente registro de bienes raíces, se debería en el futuro utilizar la palabra catastro del suelo. Para caracterizar esta expresión aún más se vienen empleando en todos los países del mundo denominaciones tales como catastro jurídico – fiscal, catastro de límites, catastro de propietario y catastro económico, denominaciones todas que expresan la finalidad del catastro como una fuente de información de todos los datos relativos al suelo, para poner de relieve la importancia que tiene política, económica y social.

Es necesario agregar que la existencia de un catastro del suelo es una necesidad para cualquier país, pues actualmente ninguno puede prescindir de él, el cual por naturaleza

¹ Que son los Catastros de Bogotá, Medellín, Antioquia y Cali

adopta un carácter peculiar conforme a los objetivos políticos y oficiales que se le tracen al país.

1.2 Aspectos del catastro

1.2.1 Aspecto Físico.

Esta labor se realiza en el proceso de identificación predial, el cual consiste en el levantamiento de la información y la verificación de los elementos físico y jurídico del predio, mediante la práctica de la inspección catastral y demás medios probatorios para identificar en documentos cartográficos y/o catastrales su ubicación, linderos, extensión, construcciones y/o edificaciones, y precisar el derecho de propiedad o posesión (Resolución 70, 2011)

1.2.2 Aspecto Jurídico.

En este aspecto se indican y anotan en los documentos catastrales la relación entre el sujeto activo del derecho o sea el propietario o poseedor y el objeto o bien inmueble de acuerdo con el Código Civil Colombiano, mediante la identificación ciudadana o tributaria de matrícula inmobiliaria del predio respectivo (Resolución 70, 2011).

“El Código Civil Colombiano al respecto establece que los inmuebles o fincas o bienes raíces son las cosas que no pueden transportarse de un lugar a otro, como las tierras y minas y las que adhieren permanentemente a ellos, como los edificios, los árboles, etc.” La propiedad, es el derecho real a una cosa corporal, para gozar y disponer de ella arbitrariamente no siendo contra la Ley o derecho ajeno.

1.2.3 Aspecto Fiscal.

El aspecto fiscal consiste en la preparación y entrega a las Tesorerías Municipales respectivamente de los avalúos sobre los cuales ha de aplicarse la tasa correspondiente al impuesto predial (Resolución 70, 2011).

La contribución territorial o impuesto predial existirá siempre en los países donde se reúna los sistemas de propiedad privada. Tres factores desempeñan un papel preponderante en la determinación de la contribución territorial: el contribuyente, la materia imponente y la estimación de valor. La determinación de la persona sujeta al pago de impuestos y del contribuyente debe hallarse en todos los casos asociados a la identificación del propietario.

1.2.4 Aspecto Económico

El aspecto económico consistente en la determinación del avalúo catastral del predio. En ningún caso la maquinaria agrícola e industrial constituirá base para la determinación del avalúo catastral del predio. Los cultivos sí se tiene en cuenta en el estudio de zonas Homogéneas, mas particularmente.

Se define como avalúo catastral, la determinación del valor de los predios, obtenida mediante investigación y análisis estadístico del mercado inmobiliario. El avalúo catastral de cada predio se determinará por la edición de los avalúos parciales practicados independientemente para los terrenos y para las edificaciones en el comprendidos (Resolución 70, 2011).

La importancia de la valoración en Colombia, no solo se centra en los beneficios fiscales, que permite a los municipios estimar la liquidación y cobro del impuesto predial, sino que contribuye a que las entidades territoriales cuenten con ingresos que le permitan la inversión en obras de infraestructura y otros gastos territoriales para el desarrollo de estos territorios.

También se constituye en la base para conocer el comportamiento del Mercado inmobiliario, las variables que afectan sus tendencias, lo que pueden constituir herramientas importantes para formular políticas públicas inherente a la gestión del suelo, teniendo en cuenta que este último se constituye un bien limitado y escaso.

El hecho que se puedan estimar los valores de mercado, permite que los propietarios de los bienes inmuebles tengan información aproximada de su patrimonio, y así en el momento de transarlo en el mercado, lo haga con aproximación a la realidad, con la finalidad, que el intercambio se haga a un precio justo tanto para el vendedor como para el comprador.

En el tema valuatorio el Instituto Geográfico Agustín Codazzi, IGAC, es la máxima autoridad tanto en la gestión a nivel catastral (cuyo procedimiento se hace de forma masiva dentro de los procesos catastrales), como a nivel puntual (que busca estimar el valor comercial de inmuebles específicos y por demanda).

1.3 Actividades Catastrales

Dentro del marco legislativo colombiano las autoridades catastrales deben adelantar tres grandes actividades técnicas para el levantamiento de la información catastral a saber: i) la formación del catastro ii) la actualización de la formación del catastro, y, iii) la conservación del catastro. Así:

1.3.1 Formación

Este es el proceso mediante el cual se obtiene la información correspondiente a los predios de una unidad orgánica catastral o parte de ella, teniendo como base sus aspectos físico, jurídico, fiscal y económico, con el fin de lograr los objetivos generales el catastro.

La información catastral se consigue en documentos cartográficos que permitan la mensura del área, la elaboración del plano de conjunto del Municipio y de las cartas catastrales con su respectiva identificación predial y que contenga la clasificación agrológica del suelo y su uso. La información implica:

- a. Deslinde Municipal, perímetro urbano y nomenclatura general.
- b. Identificación de cada uno de los predios en asocio del propietario, poseedor o su representante.
- c. Ubicación y numeración del predio de la carta catastral Municipal.
- d. Diligenciamiento de la ficha predial, la cual constituye el acta de identificación predial, debidamente fechada y firmada por el funcionario catastral y el propietario, poseedor o representante.
- e. Determinación de las zonas homogéneas físicas y estudio de mercado inmobiliario para determinar el valor de los terrenos y edificaciones.
- f. Liquidación del avalúo catastral de cada predio.
- g. Plano del predio con indicación de sus colindantes.
- h. Elaboración de documentos gráficos, estadísticos, listas de propietarios o poseedores.
- i. Resolución que ordena la inscripción de los predios que han sido formados, con indicación de su vigencia (Resolución 70, 2011).

1.3.2 Actualización de la Formación

La actualización de la formación catastral tal como su nombre lo indica está referida a renovar o confirmar los datos derivados del proceso de formación, mediante revisión física, jurídica y económica de los elementos del Catastro. Se registraran todas las modificaciones que se detecten como consecuencia de cambios físicos, variaciones en el uso del suelo o de su productividad, elementos externos influyentes (vías, obras públicas, etc.) o condiciones del mercado. La normatividad de la actualización catastral a partir de la expedición de la Ley 14 de 1983 ha sufrido varios cambios especialmente relacionados con la temporalidad de la misma, actualmente y hasta que se modifique la Ley 223 de 1995, las actualizaciones catastrales deben adelantarse en periodos de máximo 5 años. Sin embargo, como la Ley no establece sanciones al respecto a los municipios que no lleven a cabo los procesos de actualización muchos municipios y el Gobierno nacional se escudan en la falta de recursos para adelantar dichos procesos, lo que como consecuencia ha producido y produce grandes rezagos en la información catastral.

1.3.3 Conservación Catastral

Entre las dos grandes actividades catastrales enunciadas, se encuentra el proceso de conservación catastral, el cual está regulado por la Resolución 70 de 2012, expedida por el Instituto Geográfico Agustín Codazzi, la cual define los cambios que deben ser objeto de seguimiento en este proceso que tiene como finalidad:

- a. Mantener al día los documentos catastrales de acuerdo con los cambios que experimente la propiedad inmueble.
- b. Asegurar la debida conexión entre el Notariado, el Registro y el Catastro
Designar de manera técnica los inmuebles en los documentos públicos y en los actos y contratos en genera inmobiliario.

1.4 Importancia del catastro

Actualmente, uno de los principales problemas de Colombia, está relacionado con la producción de información fundamental para el desarrollo del país. La producción, conservación, actualización y divulgación de la información catastral es uno de ellos. El catastro es el soporte de muchos proyectos nacionales, entre ellos, el fortalecimiento de las

finanzas municipales, el ordenamiento territorial, la estratificación socioeconómica, la titulación de tierras y por supuesto, el inventario de los recursos naturales. A pesar de los rezagos de la información catastral esta es muy importante para el crecimiento y desarrollo del país (Martínez et al, 2002a, 2002b) ya que de manera directa e indirecta genera impactos positivos dentro de los que destacan aspectos económicos, sociales y ambientales de los cuales se encuentran los siguientes:

- a.** Seguridad jurídica de la propiedad.
- b.** Fortalecimiento de las finanzas municipales.
- c.** Mayor información para los procesos de planeación y ordenamiento territorial.
- d.** Apoyo al Plan Colombia.
- e.** Desarrollo sostenible.
- f.** Erradicación de cultivos ilícitos.
- g.** Mayor cobertura en la estratificación socioeconómica.
- h.** Definición y coberturas de servicios públicos.
- i.** Coberturas en la formalización y la titulación predial.
- j.** Inventario de los recursos naturales para su protección y conservación.
- k.** Desarrollo de infraestructuras.
- l.** Apoyo en situaciones de emergencia por causa de desastres naturales

2. Métodos de Valoración de inmuebles

2.1 Métodos de valoración de acuerdo con la Colombiana.

Los métodos de valoración de bienes inmuebles oficialmente reconocidos en la normativa legal vigente de la mayoría de los países del mundo son cuatro (4), y en nuestro país están regulados principalmente en el Decreto 1420 de 1998 y en la resolución del IGAC N° 620 del 2008, entre otras normas aplicables. A través de estos métodos se trata de estimar el valor objetivo del bien urbano, teniendo en cuenta las características generales de su entorno y las particularidades del inmueble. En los siguientes numerales se tratan de manera sucinta estos métodos.

2.1.1 Método de Comparación de Mercado

Es la técnica valuadora que busca establecer el valor comercial del bien, a partir del estudio de las ofertas o transacciones recientes, de bienes semejantes y comparables al del objeto de avalúo. Tales ofertas o transacciones deberán ser clasificadas, analizadas e interpretadas para llegar a la estimación del valor comercial. Se comparan las características de las propiedades de venta a la propiedad de sujeto sobre la base de tiempo y características tales como la ubicación, el tamaño y el potencial urbanístico de la propiedad (Resolución 620, 2008).

Cuando para la realización del avalúo se acuda a información de ofertas y/o transacciones, es necesario que en la presentación del avalúo se haga mención explícita del medio del cual se obtuvo la información y la fecha de publicación, además de otros factores que permitan su identificación posterior.

Existen algunos factores que pudieran restar aplicabilidad a este procedimiento, entre ellos se pueden destacar los siguientes:

- a. El mercado podría no presentar suficiente casuística como para servir de base fiable, sin excesiva dispersión, para la comparación con el inmueble a valorar.

-
- b. En ocasiones las implicaciones fiscales pueden dar lugar a transacciones con escasa transparencia.
 - c. Es posible que se den condicionantes en el mercado que impidan la libre transacción entre las partes que intervienen en la compraventa, por esta razón la microeconomía transaccional puede no ser extrapolable a otro inmueble con características semejantes.

2.1.2 Método de Costo de Reposición.

Es el que busca establecer el valor comercial (V_c) del bien objeto de avalúo a partir de estimar el costo total de la construcción a precios de hoy, un bien semejante al del objeto de avalúo, y restarle la depreciación acumulada. Al valor así obtenido se le debe adicionar el valor correspondiente al terreno.

Se debe entender por costo total de la construcción la suma de los costos directos, costos indirectos, los financieros y los de gerencia del proyecto, en que debe incurrirse para la realización de la obra. Después de calculados los volúmenes y unidades requeridas para la construcción se debe tener especial atención con los costos propios del sitio donde se localiza el inmueble. Al valor definido como costo total se le debe aplicar la depreciación. Para ello se utilizará la siguiente fórmula:

$$V_c = \{C_t - D\} + V_t$$

Ecuación 1. Fórmula para Calcular Valor Comercial de un Inmueble

En donde:

V_c = Valor comercial

C_t = Costo total de la construcción

D = Depreciación

V_t = Valor del terreno

Este método se debe usar en caso que el bien objeto de avalúo no cuente con bienes comparables por su naturaleza (colegios, hospitales, estadios, etc.) o por la inexistencia de datos de mercado (ofertas o transacciones) y corresponda a una propiedad no sujeta al régimen de propiedad horizontal. Para depreciar los equipos especiales que posea el bien,

se emplea el método lineal, tomando en cuenta la vida remanente en proporción a la vida útil establecida por el fabricante.

2.1.3 Método – Técnica Residual

Este método, se basa en el principio del mayor y mejor uso del terreno y que en Términos coloquiales se expresa como que la tierra vale según lo que pueda cargar.

Es el que busca establecer el valor comercial del bien, normalmente el valor comercial del terreno (V.T.B), a partir de estimar que el valor de éste, es una porción del monto total de un proyecto de construcción, acorde con la reglamentación urbanística vigente y de conformidad con el mercado del bien final vendible, en el terreno objeto de avalúo.

El método residual permite establecer el valor del terreno a partir del desarrollo hipotético de un proyecto inmobiliario teniendo en cuenta las normas urbanísticas aplicables y estimando el mejor y óptimo aprovechamiento. Al ingreso por ventas que genera la ejecución del proyecto se deducen los costos de desarrollo y la utilidad esperada, determinando la fracción del ingreso neto la cual se capitaliza para indicar el valor de la tierra. Más concretamente, el valor de la propiedad es equivalente al valor descontado de beneficios futuros. Estos beneficios representan los flujos de efectivo anuales netos (positivos o negativos) durante el período de tiempo estimado para la ejecución de un proyecto inmobiliario. La selección de la tasa de descuento o la tasa interna de retorno se basa en la comparación de la propiedad sujeto otras oportunidades de inversión en bienes raíces, así como otras formas de inversiones.

Para la estimación del precio de un terreno en bruto, cuando por las condiciones del mercado no se pueda estimar directamente, se calculará partiendo del valor del terreno urbanizado, y se aplicará la siguiente fórmula (Resolución 620, 2008):

$$V.T.B = \%AU \left\{ \frac{V_{tu}}{1 + g} - C_u \right\}$$

Ecuación 2. Fórmula para obtener el Valor de Terreno de un Bien

En donde:

% AU	Porcentaje área útil.
Vtu	Valor del terreno urbanizado
g	Ganancia por la acción de urbanizar.
Cu	Costos de urbanismo (Debe incluir los costos financieros y no solo los de obra)

Por porcentaje de área útil se entiende el resultado de dividir el área útil de cada predio, por el área total de cada predio o predios sujetos a plan parcial; al tenor de lo establecido en el Decreto 2181 de 2006. Es necesario tener en cuenta que las obras de urbanismo guarden relación con el tipo de proyecto que la norma determine.

2.1.4 Método de Capitalización de Rentas o Ingresos

Es la técnica valuatora que busca establecer el valor comercial (VC) de un bien, a partir de las rentas o ingresos que se puedan obtener del mismo bien, o inmuebles semejantes y comparables por sus características físicas, de uso y ubicación, trayendo a valor presente la suma de los probables ingresos o rentas generadas en la vida remanente del bien objeto avalúo, con una tasa de capitalización o interés (Resolución 620, 2008).

$$VC = \frac{r}{i}$$

Ecuación 3. Fórmula para obtener el Valor Comercial de un bien

En donde:

VC= Valor comercial
r = renta mensual calculada por estadísticas de mercado
i = porcentaje de renta mensual

Es necesario realizar la investigación de los contratos que regulen la posibilidad de generar rentas o ingresos, tales como los de arrendamiento, para bienes comparables y deben tenerse en cuenta aspectos los siguientes aspectos:

- a. Que dichos contratos tengan menos de un (1) año de suscritos.
- b. Que el canon de arrendamiento no sobrepase los topes legales.
- c. Que los montos relacionados con el pago de servicios públicos y las cuotas de administración, no se incluyan en el cálculo correspondiente para la aplicación del método.

- d. Los arrendamientos a comparar deben referirse a inmuebles que tengan rentas de acuerdo con la norma de uso del terreno o de las construcciones.
- e. Las rentas a tener en cuenta para el cálculo del valor comercial de la propiedad deben estar asociadas exclusivamente al inmueble y no a la rentabilidad de la actividad económica que en él se realiza.
- f. La tasa de capitalización (i) utilizada en este método debe proceder de la relación calculada entre el canon de renta y el valor comercial de las propiedades similares al inmueble objeto de avalúo, en función del uso o usos existentes en el predio y de localización comparable.
- g. El método de capitalización de ingresos para aplicarlo a cultivos debe referirse a ingresos netos del cálculo de flujo de caja, traídos a valor presente. Debe mejorarse, haciendo claridad respecto a la inversión o costo de instalación y los costos de mantenimiento del cultivo de su etapa improductiva, lo mismo que al arrendamiento².

2.2 El Avalúo Catastral en Colombia

2.2.1 Normatividad que rige los avalúos Catastrales

Los avalúos catastrales están reglamentados principalmente por la Ley 14 de 1983 su Decreto Reglamentario 3496, del mismo año y la Resolución 70 de 2011, emanada de la Dirección General del Instituto Geográfico Agustín Codazzi.

La ley 14 de 1983 define que el avalúo catastral se determina por la adición de los avalúos parciales practicados independientemente para los terrenos y para las edificaciones en él comprendidas y establece, que las labores catastrales están sujetas en todo el territorio a las normas técnicas establecidas por el IGAC. Fundamentados en esta normatividad, la

² Tomado de Resolución 620 de 2008 del IGAC

Subdirección de Catastro, elaboró la metodología y técnicas a seguir para la determinación de los avalúos catastrales, tanto para terreno como para construcción.

La Resolución 70 de 2011, por la cual se reglamenta y se unifican disposiciones para la Formación, Actualización de la Formación y Conservación del Catastro Nacional, describe en forma general los parámetros que se deben tener en cuenta para la valoración catastral y otras disposiciones.

Definiendo que el Avalúo Catastral consiste en la determinación del valor de los predios obtenido mediante investigación y análisis estadístico del mercado inmobiliario y que en él quedarán comprendidos el valor de los terrenos y el valor de las edificaciones.

En el avalúo catastral no se tiene en Cuenta el mayor valor por la utilización futura del inmueble en relación con el momento de la identificación predial, ni los valores históricos, ni el artístico, afectivo, "*good will*", u otros valores intangibles o de paisaje natural que pueda presentar un inmueble.

Se fijan como etapas para determinar el valor de los predios: la identificación predial, la Determinación de las zonas Homogéneas Geoeconómicas, (ZHG), la determinación de valores unitarios para los tipos de edificaciones y la liquidación de avalúos.

El valor unitario del terreno, se determina con base en las zonas homogéneas geoeconómicas, espacio geográfico de una región con características similares en cuanto a su precio.

La determinación de dichas zonas se constituye en un proceso, mediante el cual se establece, a partir de puntos de investigación económica dentro de las zonas homogéneas físicas (ZHF), el valor del mercado inmobiliario para los terrenos ubicados en ellas.

Se entiende por puntos de investigación económica, aquellos seleccionados dentro del área urbana o rural del municipio para establecer valores unitarios del terreno, mediante el análisis de la información directa e indirecta de precios en el mercado inmobiliario.

Las zonas homogéneas físicas rurales se obtienen con fundamento en las condiciones agrológicas, topográficas, climatológicos de los suelos y en su capacidad y limitaciones de

uso y manejo y las zonas homogéneas físicas urbanas con las condiciones topográficas, de uso, como también el destino económico.

El valor unitario de los diferentes tipos de construcciones, se determina realizando investigaciones económicas mediante el análisis de la información directa e indirecta de precios en el mercado inmobiliario para cada uno de ellos.

Se define que los factores que inciden en el avalúo de los edificios y construcciones son los materiales de construcción propiamente dichos; los acabados, la vetustez; el estado de conservación; la ubicación; y otros factores que en un futuro deban ser considerados y que lo indiquen las normas del Instituto Geográfico Agustín Codazzi, IGAC.

Las construcciones se tipifican, agrupándolas en tipos como, residenciales, industriales, comerciales, etc., teniendo en cuenta características arquitectónicas, socioeconómicas, de uso y de servicios públicos y que se deben determinar para cada uno de ellas su valor unitario por metro cuadrado.

También, se aprueba la adopción de otros métodos técnicos, siempre y cuando cumplan la finalidad prevista en la Ley 14 de 1983 para la determinación de los avalúos, y sean autorizados por la Dirección General del Instituto Geográfico Agustín Codazzi.

2.2.2 Metodología de Zonas Homogéneas empleada por el IGAC para la determinación del Avalúo Catastral.

Como se mencionó anteriormente, el IGAC estableció como etapas para establecer el valor de los predios, la identificación predial, la determinación de las zonas homogéneas geoeconómicas, la determinación de los valores unitarios para los tipos de las edificaciones y la liquidación de avalúos.

▪ Identificación Predial

En esta etapa se recopila la información física y jurídica de los predios, determinando las áreas de terreno; de las distintas construcciones, calificándolas en la ficha predial, se obtiene la edad de la construcción; y se determina el destino económico del predio en su conjunto, precisando el derecho de propiedad o posesión. En la calificación se tiene en cuenta, el tipo

de estructura, los acabados principales y su estado de conservación), para lo cual se asigna un puntaje respectivo de acuerdo a esas características. Ver anexo A. Formato de calificación de edificaciones ficha predial.

En esta etapa se establece el valor en el mercado inmobiliario del terreno por zonas, con base en los puntos de investigación previamente: En ella se realizan las siguientes actividades:

- **Determinación de Zonas Homogéneas Físicas.**

Dependiendo si se trata de sectores urbanos o rurales, se utilizan variables que determinan en gran medida el valor de las tierras, las cuales, están acordes con la teoría económica de la renta de la tierra tanto absoluta como las rentas diferenciales.

Para las Zonas Homogéneas Urbanas las variables que se tienen en cuenta son: Reglamentación del uso del suelo, Uso actual del suelo, Vías, Topografía, Servicios Públicos y Tipificación de las construcciones:

Para las zonas rurales las variables utilizadas son: Áreas Homogéneas de Tierra, Disponibilidad de Aguas Superficiales Permanente, Influencia de las vías, Reglamentación del Uso del Suelo Rural y Uso actual del suelo.

La superposición de las variables anotadas para cada zona urbana y rural determina el plano de zonas homogéneas físicas.

- **Determinación de puntos para realizar la investigación económica.**

Sobre el plano de zonas homogéneas físicas, se determinan los puntos para hacer la investigación económica; la muestra mínima de puntos a fijar para la investigación económica, se calcula utilizando la fórmula: 15 predios más el 1% del total de los predios de la población sin incluir mejoras que se ubicarán por toda la ciudad. Es decir que para el caso de la zona urbana del municipio de Fusagasugá, al contar con un total de 43.264 predios, el número de puntos de investigación mínimo sería de $15 + 432$, lo que sería igual a 447 predios (IGAC, 2010).

Los predios que sean seleccionados como puntos de investigación deben ser representativos de cada zona física, en cuanto a tamaño y forma, teniendo en cuenta que en cada ZHF debe tener por lo menos un punto de investigación económica. La ubicación en el plano de los puntos de investigación determinados, permite verificar la cobertura territorial dada a la investigación económica.

▪ **Investigación de los valores comerciales de terreno.**

Se realiza la respectiva investigación de los valores comerciales de terreno; para la zona urbana y rural, mediante investigación directa e indirecta del mercado inmobiliario. A cada punto de investigación se le realiza el respectivo avalúo comercial.

La investigación indirecta, en la cual se obtiene y analiza las transacciones comerciales recientes, registradas en los inmuebles del municipio: ofertas, compraventas, hipotecas, arrendamientos, avalúos comerciales, avalúos administrativos ICAC y otros avalúos.

Se complementa con la investigación directa, la cual se realiza con peritos expertos de la región, concedores del municipio y del mercado inmobiliario local, con el fin de recolectar datos, procesarlos estadísticamente, corroborando la investigación indirecta. El número mínimo de encuestas requerido es de cinco (5) para cada predio seleccionado.

▪ **Cálculo del Valor Unitario por Punto.**

Una vez realizada la investigación directa e indirecta del mercado inmobiliario, la información obtenida se analiza estadísticamente, haciéndose una depuración para obtener valores comerciales por cada punto investigado.

Para ello se utiliza una serie de cálculos estadísticos utilizando los conceptos de media, desviación estándar, coeficiente de variación, límites superior e inferior (con base en intervalos de confianza determinados). Las fórmulas utilizadas son las siguientes:

$$\mu = \frac{\sum_{i=1}^n x_i}{n}$$

Ecuación 4. Promedio del Valor Investigado (μ)

$$\delta = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}}$$

Ecuación 5. Desviación Estándar δ

$$CV = \frac{\delta}{\mu} * 100$$

Ecuación 6. Coeficiente de Variación (CV)

Para los fines del estudio de zonas, el máximo valor aceptado para el coeficiente de variación, es 10 %, un valor mayor a éste, indica que se requiere reforzar la investigación, fijando nuevos puntos o consultando otros informantes.

- **Intervalo de confianza**

La investigación económica se ha hecho seleccionando una muestra de predios. Considerando una distribución normal de los datos investigados, el intervalo de confianza indica el rango de valores dentro del cual está el precio promedio de la población limitado por un valor mínimo y uno máximo, que se denominan límites inferior y superior. Conociendo el promedio, el número de datos y la desviación estándar, se calcula el rango de valores dentro de un porcentaje de error admitido, dado por el valor asignado a alfa (α), en este caso es 5%.

Límite Inferior

$$L_i = \mu - K * \delta$$

Límite superior

$$L_s = \mu + K * \delta$$

*Cálculo de la Constante **K** (Distribución Normal o t - Student)*

$$K = \frac{t}{V_n}$$

Ecuación 7. Cálculo de intervalo de confianza

En donde:

(μ) : Promedio del Valor Investigado

δ : Desviación Estándar

t : Es el valor leído en la tabla de distribución t para $n - 1$ grados de libertad y probabilidad $(1 - \alpha)$ igual al 95%..

Quien elabora el estudio, selecciona un valor para el punto investigado, sin salirse de los límites superior e inferior del intervalo de confianza calculado.

Al finalizar este proceso, se habrá obtenido el valor unitario de terreno para cada uno los puntos de investigación.

- **Diseño del Plano de Zonas Homogéneas Goeconómicas.**

Una vez adoptados en forma preliminar los valores unitarios de terreno por punto, se ordenan los puntos de investigación por rangos de valores mayor a menor, y se asigna la numeración de zonas partiendo de 1, en orden ascendente,

La base para el diseño de las zonas homogéneas goeconómicas, es el plano de zonas homogéneas físicas y los valores unitarios calculados para las zonas, explicado en el párrafo anterior.

El plano de zonas homogéneas goeconómicas representa la división de la zona del municipio de acuerdo con el valor unitario de terreno, definido para calcular posteriormente el avalúo catastral de los predios del municipio.

Contiene información del valor por m^2 / Hectárea de mayor a menor. Plano de zonas homogéneas goeconómicas municipio de Fusagasugá Zona urbana.

2.2.3 Determinación de los valores unitarios para las construcciones

En esta etapa se obtienen valores unitarios por cada tipo de construcción existente en el municipio.

En la construcción son elementos del avalúo, las cosas que contribuyen a formar un bien compuesto de manera que no puedan separarse sin que aquel se destruya, deteriore o altere.

Los factores que inciden en el avalúo de las construcciones y/ o edificaciones son:

- a. Los materiales de construcción y su calidad.
- b. Las condiciones urbanísticas y arquitectónicas.
- c. El uso de la construcción y/o edificación.
- d. La edad o vetustez.
- e. La ubicación según clasificación catastral.

El tipo de construcción y/o edificación de las mismas, teniendo en cuenta los materiales de construcción, las características arquitectónicas y su uso es la clasificación de las construcciones.

Para determinar el avalúo catastral de las construcciones, se ha implementado una metodología basada en tablas de valores unitarios relacionados con la calificación de cada unidad de construcción y determinados mediante el cálculo de regresiones con base en datos obtenidos de practicar avalúos individuales a una muestra representativa de las construcciones del municipio y mediante investigación de las condiciones del mercado inmobiliario.

La metodología para determinar valores unitarios, se requiere seleccionar una muestra de predios o puntos de investigación, que son representativos de las condiciones locales de la construcción.

Con base en estadísticas y apoyado en el conocimiento del municipio, se determinan los usos de la construcción existentes. Se agrupan en forma general en actividades comerciales, residenciales e industriales.

Las tablas de construcción se elaboran teniendo como base información de costos y precios de los inmuebles en el mercado inmobiliario.

▪ **Selección de los Puntos de Investigación.**

Una vez definidas las tablas a elaborar, se seleccionan una muestra de mínimo de quince (15) predios representativos para cada una de ellas, con el fin de hacer la investigación de valores. Este número mínimo permite hacer un cálculo de regresión confiable.

El número máximo de puntos para calcular una tabla depende, por una parte, del volumen de predios que existan con el destino y por otra, de las condiciones locales de estado de la construcción.

Los criterios para seleccionar los predios a investigar son:

- a. Predios localizados en diferentes sectores del municipio
- b. Predios que cubran los tipos de construcción existentes en el municipio, en caso de vivienda seleccionar tipos según tipo, existentes.
- c. Predios que cubran los usos contemplados para cada tabla, según se determinó, en la definición de tablas a elaborar.
- d. Predios que cubran los puntajes existentes.
- e. Predios con áreas dentro del promedio para ese destino.

▪ **Investigación de valores unitarios para cada tipo de construcción.**

Seleccionados los predios para investigar, de cada uno se obtiene una copia del anexo de calificación de la ficha predial con información vigente en cuanto a áreas, zonas, destinos y puntaje. La información contenida es base para hacer la investigación directa e indirecta del mercado inmobiliario del municipio.

Se realiza la investigación del mercado inmobiliario.

Se obtiene el valor unitario para cada tipo de construcción investigada realizándose el debido análisis estadístico a la información obtenida.

- **Cálculo de Regresiones.**

En el análisis de regresiones parte de una nube de puntos y mediante el cálculo de una ecuación se determina una línea de ajuste y el grado de correlación existente entre los datos. Con esta información se grafica los valores por m² de construcción y el puntaje, extraído de la ficha predial, obteniendo una nube de puntos, la cual describe la tendencia del comportamiento del mercado inmobiliario de la zona.

Se realizan los modelos matemáticos para hallar la ecuación de la curva que describa mejor el comportamiento del valor del m² de construcción mediante análisis de regresión simple lineal, potencial, logarítmica y exponencial, tomando como variable independiente el puntaje de calificación de la construcción y dependientemente el valor del m² investigado. Se escoge la curva que más ajusta al comportamiento de los datos, utilizando como indicador el coeficiente de determinación, R², que esté más cercano a uno.

Como conclusión y verificación de la funcionalidad del estudio de zonas homogéneas elaborado, se hacen avalúos de prueba representativos de todas las zonas geoeconómicas, aproximadamente en un 10 % del número total de predios, dependiendo también de la importancia económica. Si se encuentran inconsistencias y se realiza una nueva investigación económica en las zonas con inconsistencias y se hacen las modificaciones a que haya lugar.

Los valores se anotan en el cuadro comparativo de avalúos de prueba. De este cuadro se debe deducir cuales son los incrementos por zona y el incremento total entre el avalúo vigente y el propuesto.

Una vez obtenido los valores por m² de terreno y construcción se realiza un comité de avalúos de la Dirección Territorial, conformado por varios funcionarios, expertos en avalúos, quienes discuten el estudio y determinan los valores unitarios de terreno y construcción definitivos, teniendo en cuenta que el avalúo catastral debe ser inferior al comercial, siendo soporte de los resultados un acta firmada los participantes. El estudio se envía a la Subdirección de Catastro para obtener su concepto técnico favorable. Una vez aprobado el estudio por parte de esta Dependencia, se expide la resolución de aprobación de los estudios de zonas homogéneas y valor unitario por tipo de construcción por parte de la Dirección Territorial. Se liquidan los avalúos catastrales, producto del valor unitario del terreno y construcción por sus respectivas áreas.

Este proceso culmina con la emisión de la resolución de inscripción y vigencia catastral, si se trata de formación, para la actualización de la formación catastral con la resolución de

renovación de la inscripción de los predios actualizados y vigencia catastral, firmadas por el Director Territorial, publicándose en el diario oficial³. En la Figura 1 se muestra el Flujograma del proceso para determinar masivamente el avalúo catastral en Colombia.

³ El procedimiento de Zonas homogéneas fue extractado del Manual de Zonas Homogéneas y resolución 70 de 2011 del IGAC.

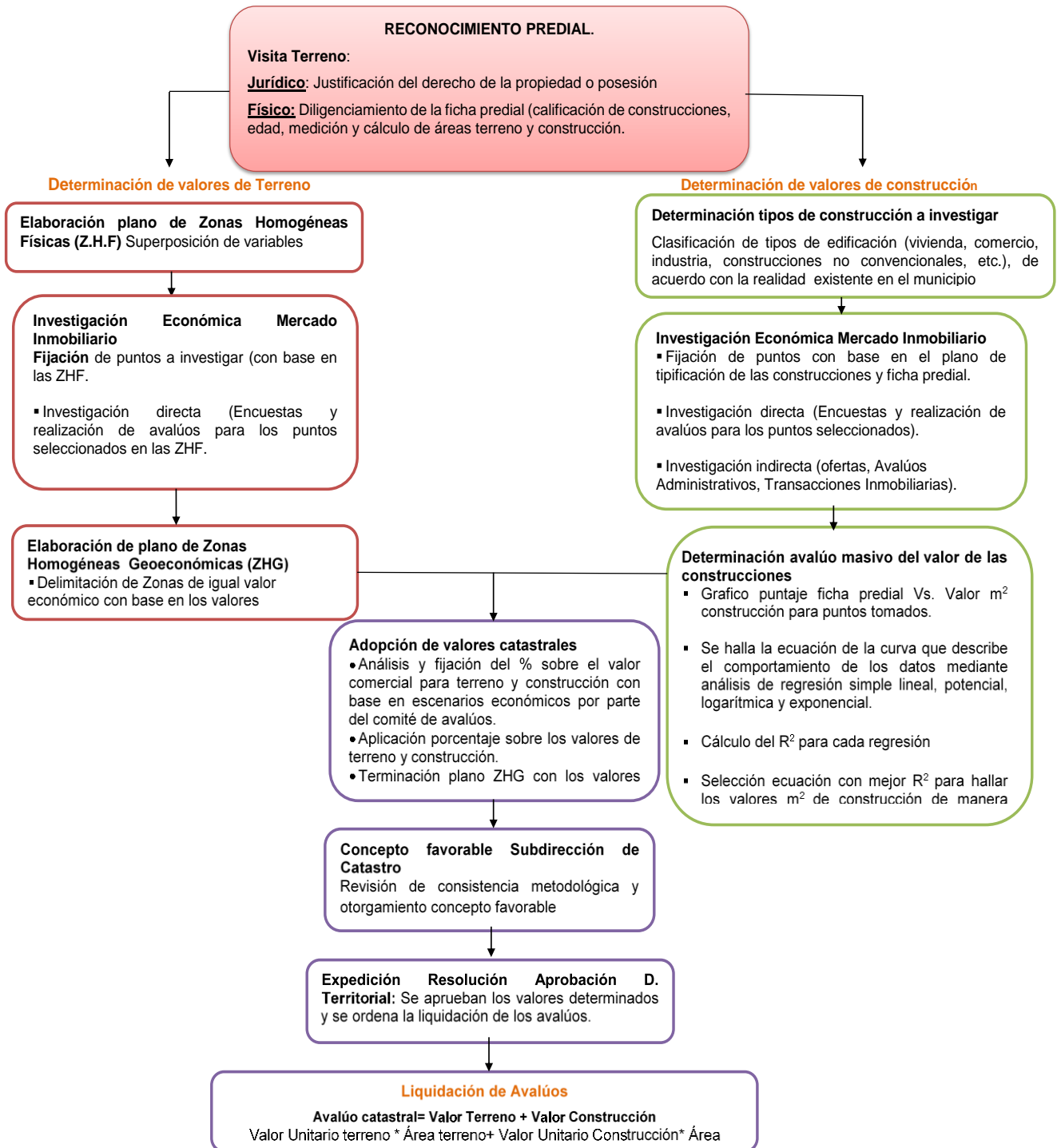


Figura 1. Flujograma del proceso para determinar el avalúo catastral de un predio en Colombia.

3. Revisión de Literatura

3.1 Redes Neuronales Artificiales

Las redes neuronales artificiales son, en esencia, modelos matemáticos diseñados a imitación del sistema nervioso biológico, centrándose fundamentalmente en los aspectos de la actividad cerebral relacionados con la formación del conocimiento humano. Si bien algunos de los modelos de redes neuronales artificiales no tienen en absoluto nada que ver con las redes neuronales biológicas, en todos ellos podemos encontrar una característica común. Esta característica es la intención de construir un sistema inteligente de procesamiento de información; entendiéndose por inteligencia aquella facultad, aptitud o factor psíquico que permite un comportamiento inteligente, es decir, un comportamiento intencional, adaptador, que resuelve problemas mediante el razonamiento, basándose en una experiencia adquirida y conservada en la memoria, capaz de prever futuros modos de conducta (Carjal, 1990).

Si tomamos un punto de vista estadístico, las redes neuronales artificiales pueden ser consideradas como una técnica multivariante de inferencia no lineal y no paramétrica. Precisamente la característica de no linealidad del modelo permite la aplicación de RNA a problemas de mayor complejidad que otras técnicas. (Rojas, 1996).

Por otro lado, el carácter no paramétrico o ausencia de hipótesis sobre la relación existente entre las variables de entrada dota a las RNA de un grado de flexibilidad no presente en las técnicas tradicionales.

Si se analiza la relación entre los métodos estadísticos más habituales y las redes neuronales artificiales, se observa cierto solapamiento, no tanto con aquella parte de la Estadística concerniente al análisis exploratorio de datos sino más bien con aquella dedicada a la inferencia estadística, es decir, a la tarea de aprender a generalizar a partir de datos ruidosos.

A continuación señalamos algunas de las tareas más importantes para las que las redes neuronales artificiales se consideran útiles (Soria & Blanco, 2001):

- a. **Clasificación:** Nos encontramos en este caso ante la tarea de decidir a qué categoría pertenece cada patrón de entrada. La red proporciona como salida, bien la probabilidad

condicionada de pertenencia a cada una de las clases o bien la clase a la que el patrón de entrada es asignado.

- b. **Asociación de patrones:** La red debe proporcionar el patrón de salida correspondiente a cada patrón de entrada presentado.
- c. **Completado de patrones:** Hace referencia a la utilización de la capacidad de memoria para proporcionar una salida cuando se ha dado parte de ella como entrada. Frecuentemente nos encontramos con conjuntos de datos incompletos. El objetivo de la red en este caso es rellenar los campos perdidos de los vectores de entrada.
- d. **Eliminación de ruido:** Como en el caso anterior, también es frecuente encontrar datos distorsionados por ruido. La salida de la red sería entonces el patrón de entrada limpio, en parte o totalmente, de ruido.
- e. **Codificación:** Se trata de codificar una entrada obteniendo como salida un dato de menor dimensión con la mínima pérdida de información posible.
- f. **Simulación:** Creación de una salida para una entrada que actúa como estímulo.
- g. **Optimización:** La entrada de la red estaría constituida por los datos iniciales de un problema de optimización y la salida sería la solución del mismo.
- h. **Control:** En este caso se tendría como entrada la situación actual de un controlador, así como la salida deseada para el mismo, proporcionando la red como salida la secuencia de acciones que se deben llevar a cabo para conseguirla.

3.1.1. Características de las Redes Neuronales Artificiales

Las redes neuronales artificiales constituyen un método para resolver problemas mediante la simulación del cerebro humano, incluyendo su comportamiento, es decir, a través del aprendizaje, cometiendo errores y haciendo descubrimientos. Son técnicas computacionales que presentan un modelo inspirado en la estructura neuronal de los organismos inteligentes y que adquieren el conocimiento mediante la experiencia, es decir, del conocimiento de situaciones pasadas. Son dirigidos a problemas con una componente de ruido fuerte y/o información incompleta. Esta es la razón principal de la ventaja obtenida en áreas tan diversas como los sistemas expertos, filtrado de datos, sistema de visión o de planteamiento.

Las RNA's son constituidas por neuronas, que son las unidades de procesamiento. Cada unidad tiene enlaces a otras unidades, en el que reciben y envían señales. Cada unidad

puede tener memoria local. Estas unidades constituyen la simulación de las células de cerebro humano, a través de la recepción y transmisión de información.

Una red neuronal artificial (RNA) puede poseer una o más capas. Por ejemplo, una red con tres capas, podríamos tener una capa de entrada, donde las unidades reciben los patrones, la capa media, donde se realiza el procesamiento y extracción de características, y la capa de salida que encuentra y muestra el resultado final. Cuanto mayor sea el número de capas, mayor será la capacidad de aprendizaje.

Un nuevo campo científico parece surgir de la utilización de las RNA's en Inteligencia Artificial como máquinas de aprendizaje, debido al gran número de aplicaciones para la extracción de conocimiento, tales como los sistemas expertos, sistemas de soporte de decisiones y análisis de sensibilidad de diagnóstico o la predicción de series de tiempo.

Participan una amplia gama de áreas que se extienden desde los sectores bancario, financiero, químico, militar, hasta la ingeniería. Se trata de la búsqueda de descripciones matemáticas o lógico-matemática, posiblemente la naturaleza compleja, sobre patrones y regularidades en conjuntos de datos (Cortez, 1997)

En particular, la aparición de las RNA's, en los años ochenta, dio un nuevo impulso al campo de la previsión con series de tiempo, multiplicando su capacidad de pronóstico de series complejas, como los mercados financieros. Su aplicabilidad en el mercado de inmuebles es más reciente pero ya tiene dos décadas de investigación.

Sin embargo, las RNA's no pueden ser vistas como una solución universal a todos los problemas. De hecho, una de las dificultades resultantes de la utilización de las RNA's para resolver los problemas es el tiempo dedicado a buscar el mejor tipo de red. Otra dificultad es acerca de la mejor manera de alimentar los datos a una red, un proceso que se denomina pre-procesamiento, en que utilizan técnicas de normalización y filtrado de datos (Cortez, 1997).

Otras características de las redes neuronales artificiales son las siguientes (Soria & Blanco, 2001):

a. Mayor potencia computacional:

Las neuronas están conectadas entre sí en paralelo y de forma masiva formando capas. Esta estructura implica que cada neurona puede estar conectada con muchas otras, de forma que la información de entrada se distribuye entre una gran cantidad de neuronas que trabajan en paralelo y, en cada momento, hay una gran cantidad de neuronas procesando la información simultáneamente. Esto hace que el proceso sea más lento que si la conexión tuviera lugar en serie (caso de los ordenadores), pero de esta forma puede trabajarse con gran cantidad de información, pudiendo manejarse información parcialmente errónea, redundante o incompleta sin que esto afecte excesivamente a los resultados y permitiendo la simulación de sistemas no lineales y caóticos.

b. Robustez y tolerancia a fallos.

Esta propiedad hace referencia al hecho de que ante una destrucción de partes aisladas del sistema, la capacidad de actuación global de la red no resulta gravemente afectada. Es sabido que el cerebro humano sufre un proceso continuo de pérdida de células nerviosas sin que su capacidad se vea mermada significativamente. La razón es que la información se representa distribuida en un número muy elevado de elementos de proceso. Este hecho aumenta su atractivo frente a las computadoras tradicionales en que, pues al ser sistemas secuenciales, el fallo en alguno de sus componentes conlleva el fallo total del sistema (Soria & Blanco, 2001).

c. Aprendizaje.

La estructura de una red está orientada al aprendizaje, de hecho para las diferentes aplicaciones la red no está programada – como puede estarlo un ordenador-, sino que “aprende”-tal y como hace el cerebro humano-, cambiando su proceso interno al comprobar los errores que va cometiendo hasta reducir el error al mínimo, es decir, este aprendizaje lo lleva a cabo mediante pruebas sucesivas.

d. Adaptabilidad.

La gran capacidad de procesamiento y su sistema de aprendizaje le hacen ser muy adaptable a la cambiante situación externa, de hecho una RNA tiene la capacidad de modificar los parámetros de los que depende su funcionamiento de acuerdo con los cambios que se producen en su entorno de trabajo (cambios en las entradas, presencia de ruido, etc.). Con respecto a la capacidad de adaptación hay que tener presente que ésta no puede ser tampoco excesivamente grande, ya que conduciría a un sistema inestable que respondería a pequeñas perturbaciones – dilema plasticidad-estabilidad-.aprendizaje lo lleva a cabo mediante pruebas sucesivas (McCulloch & Pitts, 1943).

▪ **Capacidad para procesar información inconsistente o con alta proporción de ruido.**

Esta propiedad junto con la habilidad del cerebro para tomar gran cantidad de información de entrada simultánea y generar salidas clasificadas, es fundamental para la tarea de reconocimiento de patrones.

3.1.2 Estructura de una Red Neuronal Artificial

Las redes neuronales artificiales tratan de imitar el sistema de procesamiento de la información del cerebro humano, intentando reproducir sus capacidades. Para ello, las RNA suelen basarse en el comportamiento de la sinapsis⁴.

Desde 1986 el concepto de RNA, se refiere a una red compuesta por varios operadores simples – *elementos de proceso* (PE), nodos, unidades o *neuronas*-, dotados de una pequeña cantidad de memoria. Las unidades están conectadas mediante canales de comunicación direccionales – axones -, los cuales transportan datos. Los nodos únicamente operan sobre sus datos locales y sobre las entradas que recibe a través de los axones (Rumelhart & Hinton, 1986). Por tanto, cada neurona se compone de:

- a. **Vector de entradas:** $x' = (x_1, x_2, \dots, x_i, \dots, x_n) \in R^n$, con $x_i \in R$ o $x_i \in \{0, 1\}$. Es decir, las entradas de una neurona pueden ser discretas o continuas, dependiendo del modelo de neurona considerado y de la aplicación que se le vaya a dar. Cuando las salidas pueden tomar valores continuos, se suelen limitar a un intervalo definido como $[0, 1]$ ó $[-1, +1]$.
- b. **Vector de pesos sinápticos:** $(\omega_1, \omega_2, \dots, \omega_i, \omega_n) \in R^n$. Los pesos, en valor absoluto, miden la fuerza o intensidad de la conexión, mientras que el signo correspondiente a cada peso permite distinguir el carácter de la misma. Así, si el signo de ω_i es positivo implicará una conexión excitadora, mientras que un signo negativo representará una conexión inhibitoria. De vez en cuando cada unidad examina las entradas y salida y, a partir de ellas, calcula un valor numérico llamado activación. Este valor de activación se pasa a lo largo de los enlaces que conducen a otras neuronas (Cortez, 1997).
- c. **Regla de propagación**, en general la suma ponderada o producto escalar β_{*x} , que proporciona la entrada neta.
- d. **Función de activación o transferencia** f , que proporciona el valor de salida y .
- e. **Valor de salida** $y \in R$, y $\{0, 1\}$, $\{-1, 1\}$, $\in [0, 1]$ o $y \in [-1, 1]$ o, entre otros. Las entradas de los elementos de proceso –neuronas- pueden ser las salidas de los elementos de proceso conectados a él o entradas procedentes del exterior. Así mismo, su salida puede ser una entrada a otros elementos de proceso o ser una salida de la propia RNA.

Por consiguiente, una RNA consta de varios **elementos de proceso**, conectados de alguna forma, generalmente organizados en grupos que se denominan capas. Existen dos capas típicas en toda red, que contienen conexiones con el exterior.

⁴ Al contacto entre dos neuronas se le llama sinapsis. El axón y la dendrita nunca se tocan. Siempre hay un pequeño vacío llamado hendidura sináptica.

- a. **La capa de entrada**, a través de la cual se presentan los datos a la red.
- b. **La capa de salida**, que muestra la respuesta de la red a una entrada.

La capa de entrada generalmente sirve para distribuir las entradas de la red, por lo que no se tiene en cuenta a la hora de contabilizar el número de capas de ésta.

El resto de capas existentes entre la de entrada y la de salida se denominan capas ocultas. Estas capas no poseen conexión directa con el entorno y proporcionan a la red grados de libertad adicionales, que le permitirán encontrar representaciones internas correspondientes a determinados rasgos del entorno. Pueden presentarse varios tipos de conexiones entre las neuronas (Pérez & Martín, 2003):

- a. Conexiones **intracapa o laterales**: las que se producen entre las neuronas de una misma capa.
- b. Conexiones **intercapa**: las que se producen entre neuronas de diferentes capas.
- c. Conexiones **realimentadas**: las que tienen sentido contrario al de entrada-salida (se representarían, por tanto, de derecha a izquierda).
- d. Conexiones **autorrecurrentes**: las de realimentación de una neurona consigo misma.

Asimismo, las conexiones entre las neuronas pueden ser excitatorias (en el caso de pesos positivos) o inhibitorias (en el caso de pesos negativos). Normalmente no vamos a definir de forma explícita una conexión como de un tipo u otro, sino que por medio del aprendizaje que obtiene un valor para el peso que incluirá tanto el signo como la magnitud (Pérez & Martín, 2003). Ver Figura 2.

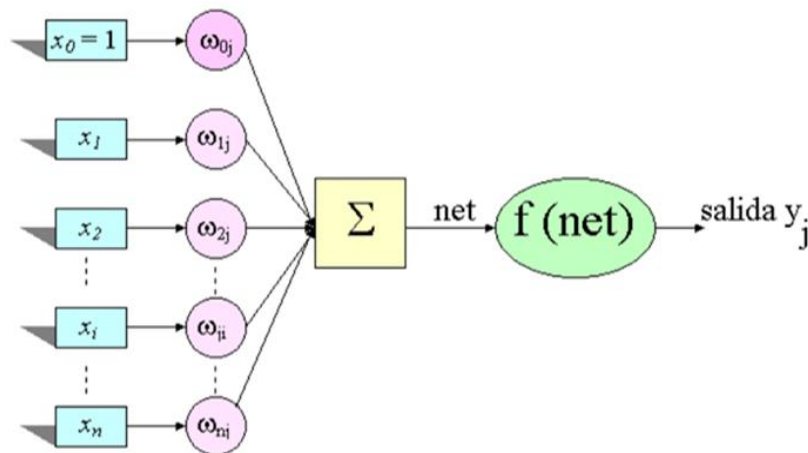


Figura 2. Diagrama de una red neuronal artificial.

En definitiva, una red diseñada para un fin específico consta de un determinado número de neuronas, conectadas en una estructura adecuada a tal fin, de unos pesos en las conexiones entre éstas y de una dinámica de funcionamiento que permite a la red obtener la salida deseada a una entrada dada.

3.1.3 La activación de las neuronas y las funciones de activación

Cada neurona calcula su activación (o energía de activación) a partir de los valores de las salidas de las células que se unen a ella y sus pesos en relación con sus conexiones. En representación de la energía de activación de la neurona i por s_i y de la neurona j por s_j se muestra en la **Ecuación 8** y **Figura 3**

$$u_i = \sum_{j=0}^n w_{i,j} \times s_j$$

$$S_i = f(u_i)$$

Ecuación 8 . Energía de activación de la neurona

Donde:

w_{ij} , representa el peso asociado a la conexión entre una neurona j y otra neurona i .

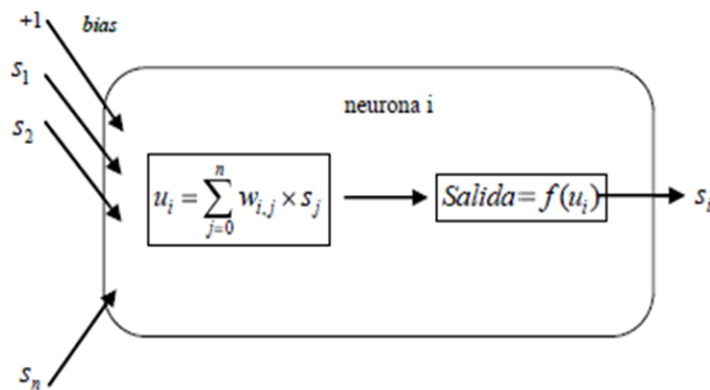


Figura 3. La activación de una neurona

La función $f(v)$ se llama función de activación. En la mayoría de modelos, es creciente y continua, como suele observarse en las neuronas biológicas (Brío & Molina, 2001). Los valores de energía de activación pueden ser discretos (por lo general tomando valores -1, 0 o 1) o continuos. En este caso, $f(v)$ suele ser una función no lineal, limitada y diferenciable por trozos.

La función sigmoidea es la función más utilizada por las RNA's. Una gran cantidad de paquetes estadístico con módulos de redes neuronales utilizan esta función. Por lo tanto, a continuación se sigue una breve presentación de esta función.

Una función $f(v)$ es Sigmoidea cuando tiene ciertas características. En primer lugar, es limitada, lo que significa que su valor no excederá nunca de un determinado valor superior o inferior. En segundo lugar, el valor de la función siempre aumenta con el valor de su argumento, es decir, es una función estrictamente creciente. En tercer lugar, una función sigmoidea es continua en su dominio.

Muchas funciones con estas características se utilizan en las RNA's. La función sigmoidea logística no necesita el uso del número de Nepper en el denominador, pudiendo ser utilizado otro valor constante mayor que 1. Cuanto mayor sea el valor de esta constante, más rápida será la función de aproximación a sus límites.

De las funciones sigmoides, existen dos que son muy utilizados, ambas con el mismo rango $[-1, 1]$, Ecuación 9 y Ecuación 10, así:

$$g(v) = 2x f(v) - 1$$

Ecuación 9. Función Logística

La otra función es la tangente hiperbólica dada por la siguiente ecuación:

$$f(v) = \frac{e^v - e^{-v}}{e^v + e^{-v}}$$

Ecuación 10. Función Hiperbólica

Esta función se aproxima más fácilmente de sus límites que la función $g(v)$. Para la implementación de una RNA, cualquiera de estas funciones es también apropiada. Las diferencias en sus límites se encuentran en la elección del intervalo de salida que se desee (0 a 1 o -1 a 1) que el espacio urbano no suele ser regular y homogéneo, sino todo lo contrario, por lo que sería una simplificación el trazado de círculos de influencia). En definitiva, la delimitación de las fronteras dentro de las cuales afecta la característica estudiada y el comportamiento del gradiente de dicha influencia en el área considerada, no son problemas que puedan calificarse como menores en la especificación del modelo de regresión, ya que

estos son los límites en los que se encuentran los valores de salida de red. Los límites de la función adoptados, no determinan, sin embargo, el rango de valores de entrada de la red. Estos pueden tomar cualquier valor, independientemente de la función sigmoidea utilizada.

3.1.4 Proceso de aprendizaje y entrenamiento de una RNA

En las redes neuronales biológicas la capacidad de aprendizaje se debe a la posibilidad de realizar cambios en la permeabilidad de las unidades sinápticas, lo que se traduce en mayor facilidad para la propagación de los impulsos nerviosos de unas unidades a otras. Esta capacidad se considera una característica esencial en toda red artificial que trate de asemejar su comportamiento al biológico. Pueden distinguirse dos fases de operación en una RNA: la de **aprendizaje** y la de **ejecución**.

La capacidad de aprender es una característica peculiar de los sistemas inteligentes. Entre las muchas propiedades interesantes de las redes neuronales, sin duda, la capacidad de aprendizaje es de relevancia primordial. Por capacidad de aprendizaje, entendemos que las redes pueden aprender de su entorno y mejorar los resultados, de acuerdo a una medida previamente fijada, a lo largo del tiempo. Las redes neuronales llevan a cabo su proceso de aprendizaje mediante la toma de información del entorno exterior y su posterior almacenamiento en los pesos de las conexiones. Los cambios que lleva consigo este proceso pretenden conseguir una mayor efectividad por parte de las RNA a la hora de realizar un trabajo específico (Basher & Hajmeer, 2000).

Una vez entrenada, una red neuronal debe ser capaz de asociar una serie de valores (insumos) a una salida específica. No se trata simplemente de una cuestión de memoria, ya que la RNA tiene capacidad de generalización y puede encontrar las respuestas correctas, incluso cuando los datos disponibles para las entradas están dañados o incompletos, o incluso cuando la relación entre entrada y salida no es correcta (Haykin, 1999).

Para que las RNA's consigan aprender con la experiencia, deben someterse a un proceso llamado de entrenamiento, para lo cual existen muchos algoritmos que determinan la forma como se adaptan los pesos o parámetros, distinguiéndose los siguientes tipos de aprendizaje (Pérez & Martín, 2003), ver **Tabla 1**.

Tabla 1. RNA según tipo de aprendizaje

Tipo Aprendizaje		Modelo de RNA
Híbrido		Funciones de Base Radial (RBF), redes de Contrapropagación
Supervisados	Realimentados	BSB, Mapa Fuzzy
	Unidireccionales	Perceptrón, Adaline (Adaptative Linear Element, Madaline (Múltiple Adaline), Perceptrón Multicapa (MLP), Red Neuronal de Regresión General (GRNN), Learning Vector Quantization (LVQ), Máquina de Boltzmann, Correlación en cascada
No Supervisados	Realimentados	Adaptative Resonance Theory (ART 1,2,3), Hopfield, Bidirectional Associative Memory (BAM)
	Unidireccionales	LAM, OLAM, Kohonen, Neocognitrón
Reforzados		Aprendizaje reforzado

Fuente: Hilera & Martínez V, 1995

- **Aprendizaje Supervisado:** En este tipo de aprendizaje existe un profesor o supervisor que dispone un conjunto de patrones de entrenamiento, formados por una entrada a la red y la salida esperada para dicha entrada. El profesor presentará a la red de forma iterativa las diferentes entradas. Para cada entrada se producirá una salida que, generalmente, no coincidirá con la salida esperada, calculándose en ese caso el error de salida. En función de este error, se actualizarán los pesos sinápticos, de manera que al volverse a presentar la misma entrada, la respuesta de la red presente un menor error. Mediante este mecanismo, es posible que la red sea capaz de estimar relaciones entrada/salida sin necesidad de proponer forma funcional de partida alguna.

- **Aprendizaje No Supervisado:** En este caso partimos de un conjunto de ejemplos para los que no existen una respuesta deseada de la red. Se determinarán las entradas de forma iterativa a fin de que la red, mediante su regla de aprendizaje, pueda descubrir las regularidades subyacentes en esos datos de entrada, organizándolos en clases no determinadas a priori. Consiste en la estimación de la función densidad de probabilidades, que describe la distribución de patrones pertenecientes al espacio de entrada, por lo que se pueden reconocer regularidades en el conjunto de entradas, extraer rasgos o agrupar patrones según su similitud (clustering), (Haykin, 1999).

- **Aprendizaje Reforzado:** Se dispone de un conjunto de entradas, para cada una de las cuales se obtiene una salida de la red. Se calcula una medida del éxito o fracaso de la red a nivel global, que permitirá determinar los pesos. Como en este caso no se puede determinar la fracción de error correspondiente a cada salida, son necesarios algoritmos de mayor complejidad, así como con conjunto de entrenamiento de un mayor tamaño.

Se trata de un mecanismo intermedio entre el supervisado y el no supervisado, ya que al igual que el primero utiliza una medida del error cometido, aunque en este caso sea un valor global indicativo del error de la red. Ahora bien tiene en común con el aprendizaje no supervisado, el no indicar explícitamente la salida deseada.

- **Aprendizaje Híbrido:** Se trata de utilizar en una misma red el aprendizaje supervisado y el no supervisado, normalmente en distintas capas de la misma (Jain, Mao, & Mohiuddin, 1996).

Normalmente, la red tarda un tiempo en aprender, incluso puede que en ocasiones no lo consiga debido a varias causas:

- a. Que el modelo de red seleccionado para resolver el problema no sea el adecuado.
- b. Que las características asociadas al modelo elegido no sean las adecuadas (número de neuronas, número de capas, funciones de activación...).
- c. Que los patrones utilizados para entrenar la red no representen bien el problema de interés, que estén mal tomados o sean pocos. Que los datos analizados no presenten ninguna regularidad, por lo que la red no puede extraer ninguna relación entre ellos.
- d. Que se hayan realizado pocas iteraciones en el entrenamiento

Una vez que el sistema ha sido entrenado viene la etapa de ejecución en la cual los pesos de las conexiones permanecen fijos (salvo en algunos modelos concretos de red) y la red generará una salida para cada entrada que se le presente. Es el momento en que la red ya está preparada para procesar datos.

En las redes unidireccionales, ante un patrón de entrada, las neuronas responden proporcionando directamente la salida del sistema, ya que al no existir realimentación no existe ningún problema respecto a su estabilidad. Por el contrario, las redes con realimentación son sistemas dinámicos no lineales, que requieren determinadas condiciones para que su respuesta acabe convergiendo a un estado estable, existiendo una serie de teoremas generales que establecen las condiciones para la estabilidad de la respuesta en una amplia gama de redes.

3.1.6 Clasificación de los modelos de RNA

Según lo visto hasta el momento, puede deducirse que dependiendo del número de capas que presente la red, de la arquitectura o topología de conexión, del tipo de respuesta, de la forma de los datos de entrada y salida o del algoritmo de aprendizaje, surgirán multitud de modelos de redes neuronales diferentes (Soria & Blanco, 2001):

- **Según número de capas**

Podemos distinguir entre redes monocapa y multicapa:

- a. **Las redes monocapa** se corresponden con la red neuronal más sencilla, pues tienen una capa de neuronas que proyectan las entradas a una capa de neuronas de salida donde se realizan diferentes cálculos. La capa de entrada por no realizar ningún cálculo no se cuenta, de ahí lo de redes neuronales con una sola capa. Una aplicación típica de este tipo de redes es como memorias asociativas.
- b. **Las redes neuronales multicapa** constituyen una generalización de las anteriores, existiendo un conjunto de capas intermedias entre la entrada y la salida (capas ocultas). Aunque no existen límites teóricos en el número de capas ocultas, normalmente hay una o dos, habiéndose comprobado experimentalmente que se pueden resolver problemas de cualquier complejidad utilizando un máximo de tres capas ocultas (Pérez & Martín, 2003).

- **Según tipo de conexiones o la topología de la red**

Según este criterio nos encontramos con redes neuronales no recurrentes (alimentadas hacia delante o unidireccionales o feedforward) y recurrentes (realimentadas o feedback).

- a. **Redes no recurrentes:** En esta red la propagación de las señales se produce en un sentido solamente, no existiendo la posibilidad de realimentaciones. No existen conexiones de una capa hacia las anteriores, ni conexiones dentro de la misma capa, ni de un elemento de proceso consigo mismo. Lógicamente estas estructuras no tienen memoria.
- b. **Redes recurrentes:** Esta red viene caracterizada por la existencia de lazos de realimentación entre neuronas de diferentes capas, entre neuronas de la misma capa o incluso entre una misma neurona. La información puede circular entre las capas en cualquier sentido, incluso el de salida-entrada. Se usa retroalimentación para dar a la red un cierto carácter temporal: la información reverberará en la red a través de las capas, o dentro de ellas, hasta que se alcance algún criterio de convergencia, momento en que la información se pasa a la salida.

- **Según tipo de respuesta**

- a. **Redes autoasociativas:** Son aquéllas que se entrenan para que asocien un patrón consigo mismo, de modo que, al presentar dicho patrón con un ruido superpuesto o de forma parcial, sean capaces de identificar el patrón. La autoasociación implica aprendizaje no supervisado.
- b. **Redes heteroasociativas:** Se entrenan para que ante la presencia de un patrón A respondan con otro diferente B. La heteroasociación supone aprendizaje no supervisado.

- **Según la forma de los datos de entrada y salida**

En algunas redes las entradas y las salidas son analógicas, es decir, toman valores continuos. Otras redes sólo aceptan datos discretos o binarios: $[0,1]$ ó $[-1,1]$ normalmente, y generan valores del mismo tipo.

Existen otras redes que pueden aceptar entradas continuas y generar salidas discretas.

- **Según el tipo de aprendizaje**

En la se recoge una clasificación de los distintos modelos de red neuronal en función del tipo de aprendizaje utilizado., ver **¡Error! No se encuentra el origen de la referencia.**

3.1.7 Las redes Perceptrón multi-capas

Entre los diferentes tipos de redes neuronales existentes, la red perceptrón multi-capas (MLP) es la más utilizada, en parte debido a su facilidad de implementación y, por otra parte, debido a su simplicidad. Estas redes se utilizan en una gran variedad de problemas especialmente en

Una serie de autores han puesto a prueba el potencial de la red perceptrón multi-capas en el ámbito de la valoración de la propiedad. Por ejemplo, (García, *et al*, 2008) han realizado varias comparaciones entre diferentes redes y llegan a la conclusión, como muchos otros autores ya habían llegado, que estas redes presentan un buen desempeño con este tipo de datos.

3.2 Análisis previo de los datos antes de ser procesados por la RNA

Es conveniente pre procesar los datos de un problema antes de que sean alimentados a la red. La idea es ajustar los datos de alguna manera tal que la red, los consiga manejar eficientemente, es decir, proceder a su normalización (Azoff, 1995). Abordamos, por tanto, a continuación la tarea relacionada con las diferentes operaciones de tratamiento previo que se le debe dar a la información antes de ser introducida y procesada por la red.

Seleccionado al menos en principio, el conjunto de variables que se van a utilizar en la aplicación, es conveniente realizar un análisis previo de los datos para lograr un conocimiento básico de los mismos, estudiando con sencillas técnicas estadísticas y

representaciones gráficas la distribución de cada una de las variables de manera aislada, así como las relaciones entre variables.

3.2.1 Detección de valores atípicos

Un aspecto importante a tratar en el primer contacto con los datos, es el relacionado con la identificación de casos atípicos, es decir, observaciones que caen fuera de los rangos de la distribución una vez establecidos los valores umbrales. Este paso es fundamental en cualquier técnica, pero especialmente en Redes Neuronales Artificiales la existencia de valores extremos hace que al normalizar o estandarizar las variables (práctica frecuente como se estudiará a continuación), las nuevas variables tendrán un rango de variación excesivamente pequeño. Esto hace que con el uso de funciones sigmoideas, la mayor parte de los valores se encuentren en la parte lineal de la misma.

La detección de estas observaciones se puede llevar a cabo mediante la estandarización de variables desde una perspectiva univariante, considerando pares de variables y realizando gráficos de dispersión e incluso tratar el aspecto desde una perspectiva multivariante utilizando, por ejemplo, la Distancia de Mahalanobis (D) o su cuadrado (D^2).

$$D^2 = (x_i - x)' S^{-1} (x_i - x)$$

Ecuación 11. Distancia de Mahalanobis

Se trata de una medida de distancia de cada observación en un espacio multidimensional respecto al centro medio de las observaciones que, además, tiene en cuenta la correlación entre las variables.

Una vez detectadas las observaciones atípicas hay que decidir si se eliminan de la muestra, evitando así la distorsión que estos datos pueden provocar en el análisis o bien optar por mantenerlas para no perder la generalidad de los resultados.

3.2.2 Tratamiento de datos ausentes

Por otra parte, con cierta frecuencia cuando nos enfrentamos a una investigación empírica nos encontramos con **datos ausentes** debido fundamentalmente a algunas de las siguientes causas: problemas en el proceso de recolección e introducción de datos, negación o desconocimiento de información por parte de la persona que atiende la visita de

reconocimiento, etc. Los efectos más importantes de la ausencia de datos son la reducción del tamaño de muestra disponible para el análisis y la posibilidad de que esas ausencias escondan sesgos.

Existen diferentes técnicas para solucionar este problema, entre las cuales se encuentra la de **Utilizar sólo aquellas observaciones con datos completos**, el cual es el método más simple y directo, de ahí que sea la solución propuesta por defecto en la mayor parte de los paquetes estadísticos, por lo tanto sólo debe utilizarse si la ausencia de datos obedece a un proceso aleatorio para evitar obtener resultados sesgados y no generalizables para la población.

3.2.3 Pre-procesamiento y post-procesamiento de los datos

Entramos, por último, en lo que propiamente se denomina **pre-procesamiento y post-procesamiento** de los datos. La Figura 4 nos muestra de manera muy simple estos pasos:



Figura 4. Fases de pre-procesamiento y post-procesamiento en una red neuronal

Fuente: Tomado de García R, 2004

El proceso de re-escalado de los datos, puede realizarse bien mediante **normalización** o bien mediante **estandarización**. Este tipo de pre-procesamiento se encuentra relacionado con las unidades de medida de las variables y, por tanto, con el rango de valores que presentan las mismas. Re-escalar un vector de datos consiste en sumar (o restar) una constante a todas las componentes del vector, y luego multiplicar (o dividir) por otra constante. Esta operación es útil, por ejemplo, cuando pretendemos cambiar las unidades en que viene medida una variable cuantitativa, para conseguir que el rango de ésta sea apropiado para ser introducida en la red neuronal.

Aunque en ocasiones se intercambian los términos normalizar y estandarizar en las operaciones de re-escalado, podríamos remarcar sus diferencias mediante las siguientes

definiciones. Entendemos por normalización el proceso de conversión del rango de una variable cuantitativa a un intervalo de tipo $[0,1]$ o $[-1,1]$. Por otra parte, consideramos como estandarización la transformación de una variable cuantitativa en otra con media nula y desviación típica igual a uno.

Sobre la conveniencia o no de llevar a cabo la normalización o estandarización de las variables de entrada, hay que analizar para qué tipo de red y con qué funciones de activación o transferencia se van a utilizar los datos. Así, por ejemplo, si las distintas variables de entrada se utilizan como inputs en una red de Kohonen y se van a calcular, por tanto, distancias euclidianas (o alguna otra medida de distancia), entonces grandes diferencias en los rangos de valores de las distintas variables influirán en la importancia relativa de cada una de ellas. En este caso no sólo es recomendable la normalización, sino que se hace imprescindible, a no ser que nos encontremos en la circunstancia de querer potenciar a propósito la importancia de alguna de las variables.

Por otra parte, cuando la red es del tipo del Perceptrón Multicapa, parece en principio poco necesario e incluso redundante, re-escalar los inputs, pues ésta puede ser la tarea de la primera capa de la red. Sin embargo, en la práctica resulta que realizar este tipo de pre-procesamiento tiene sus ventajas:

- a. Mayor rapidez en el proceso de entrenamiento.
- b. Menor probabilidad de quedar atrapados en mínimos locales.
- c. Evita el problema de saturación de los pesos. Si las variables de entrada (y las de salida en un problema de regresión) tienen rangos de valores pequeños, se podrá partir de una inicialización de los pesos aleatoria y de pequeña magnitud, evitando así la posibilidad de saturación.

Cuando las variables con las que trabajamos son de tipo cualitativo, la transformación previa de los datos se denomina **codificación**. La forma más sencilla que se nos puede ocurrir plantear es sustituir las diferentes modalidades por valores numéricos. Por ejemplo, codificación del sexo de los individuos (Hombre = 0 y Mujer = 1) o codificación de la variable Comuna (Comuna 1 = 1, Comuna 2 = 2, Comuna 3 = 3, Comuna 4 = 4, Comuna 5 = 5 y Comuna 6 = 6).

Para salvar este problema se utiliza el método de codificación conocido como uno-de-N. Este método consiste en establecer un número de variables numéricas (neuronas en la capa de entrada, en el caso que nos ocupa) igual al número de modalidades, salvo en el caso de dos modalidades, en el que una neurona será suficiente.

A la vista está el problema que plantea este método y es que en el caso en que dispongamos de un número elevado de variables cualitativas de entrada que, a su vez, presenten un número elevado de modalidades, el número de neuronas de entrada se hace demasiado grande, dando lugar al problema de la maldición de la dimensión. En cuanto a la transformación de las variables de salida cuantitativas, su conveniencia dependerá de la función de error utilizada, en particular, de la dependencia de ésta respecto de las distintas escalas de medida. Si hay una alta dependencia, como sucede con la función de error más comúnmente usada, la suma de cuadrados de los errores, al tener varias salidas con rangos distintos, el proceso de entrenamiento dará una importancia relativa mayor a las variables con mayor rango de valores. Hasta el momento hemos analizado los procedimientos de re-escalado por variables. Sin embargo, y aunque en menos ocasiones, se puede plantear la transformación por casos. Hay que señalar que los métodos de normalización o estandarización serán, en general, distintos si se trata de variables de entrada o de variables de salida.

3.2.4 Selección del conjunto final de variables

Esta fase es fundamental si se quiere evitar la pérdida de tiempo realizando aplicaciones a ciegas que darían lugar a resultados irrelevantes. Habitualmente, cuando nos enfrentamos ya sea a un problema de regresión como de clasificación, contamos con un numeroso conjunto de posibles variables explicativas a utilizar. Sin embargo, en general, no es conveniente incluir todas aquellas variables en el análisis sin más, debido a que algunas de ellas pueden resultar poco útiles para nuestro propósito. Además, el problema no es sólo que haya variables que no aporten información relevante para la reducción del error cometido sino que, debido al problema conocido como “maldición de la medición” y dado que el conjunto de datos de entrenamiento suele ser bastante limitado, es muy probable que se incurra en una situación de sobre-entrenamiento de la red, con la consecuente falta de credibilidad sobre las estimaciones de los parámetros desconocidos. Sin embargo, demasiadas variables hacen que la red se vuelva muy compleja, aumentando la contribución de la parte de la varianza al error de generalización.

La identificación de las variables significativas es una de las tareas más difíciles, y por lo general la realiza el experto en la materia de forma intuitiva. Pero esta selección intuitiva suele dar problemas. Una cosa es que se considere que una variable deba intervenir en la variación del precio, y otra cosa es que al combinarse con las demás variables pueda producir efectos distorsionantes. Las variables interactúan en conjunto y algunas de ellas, que consideradas por separado parecen necesarias, pueden presentar problemas de correlación o deformar la incidencia de otras.

Hay algoritmos matemáticos que permiten reducir el número de variables significativas, lo cual puede aplicarse como proceso previo antes de introducir los datos en el algoritmo matemático (red neuronal) que va a generar el modelo de mercado.

Las técnicas de reducción de variables más utilizadas son las siguientes:

- a. Reducción S.V.D.
- b. Análisis de Componentes Principales.
- c. Mapas Auto-organizativos.

La reducción S.V.D. (Single Value Decomposition) está basada en el cálculo matricial; El Análisis de Componentes Principales reduce las variables preservando la varianza y Los Mapas Auto-organizativos aplican una red neuronal de dos capas para llevar a cabo la reducción. Este procedimiento ya había sido utilizado para el mercado de la vivienda en Melbourne, Australia (Bourassa, 1997). De igual forma (Huawang, 2009) utilizó una técnica estadística multi-dimensiones, el Análisis de Componentes Principales para reducir la cantidad de información, en particular con características de los datos que son redundantes (Huawang, 2009). En seguida, este autor utilizó las redes neuronales para determinar el valor real del precio de la vivienda. Este nuevo modelo para establecer el precio de la vivienda se ha probado con una base de datos históricos y ha demostrado ser muy preciso.

En 2006, Zurada, Levitan y Guan, presentan un artículo donde comparan los resultados con el uso de un modelo de regresión múltiple y la RNA, de igual forma señalan algunos métodos estadísticos de pretratamiento de datos, tales como análisis de componentes principales y la selección de variables, los cuales se han utilizado para mejorar los resultados finales (Zurada et. al., 2006). Es de precisar que el análisis de componentes principales para extraer algunos factores del conjunto de variables originales, ya había sido utilizado para el mercado de la vivienda en Melbourne, Australia (Bourassa, Hamelink, M, & MacGregor., 1997).

De igual forma, Huawang y Shi (2009), también utilizó una técnica estadística multi-dimensiones, el análisis de componentes principales para reducir la cantidad de información, en particular con características de los datos que son redundantes. Posteriormente, este autor ha utilizado las redes neuronales para determinar el valor real del precio de la vivienda. Este nuevo modelo para establecer el precio de la vivienda se ha probado con una base de datos históricos y ha demostrado ser muy preciso. El conjunto de datos tenía 500 observaciones.

3.3 Uso de modelos de redes neuronales artificiales para la estimación del avalúo de inmuebles.

Como ya quedo definido las redes neuronales poseen dos características muy importantes que son la capacidad de dar respuestas rápidas a un problema y la capacidad de inferir las respuestas de los patrones desconocidos comprendidos en el dominio de entrada, aspectos que las hacen muy atractivas para realizar la valoración masiva de inmuebles. La literatura muestra que hay una mezcla de éxitos y fracasos con este método, probablemente debido a diferentes variables de entrada y condiciones de mercado diferentes.

Los intentos de aplicar la tecnología de redes neuronales a la valoración de predios datan de principios de los 90, con el estudio pionero de Borst en 1991, desde entonces los modelos con redes neuronales artificiales se han convertido en una alternativa muy atractiva a los tradicionales modelos econométricos, pues con frecuencia sus resultados son comparados con los obtenidos con los métodos estadísticos más probados.

Poco después de Borts (1991), los autores Do y Grudnitski (1993), utilizaron las RNA's demostrando que el valor de una vivienda disminuye significativamente con la edad, durante los primeros 16-20 años, esto como consecuencia del deterioro físico. En este trabajo se utilizó la información de los agentes inmobiliarios en San Diego, California, relativos a 242 viviendas vendidas durante el período de enero de 1991 a septiembre de 1991.

Shaaf (1996), utilizó las RNA para explorar el impacto en el precio medio de la vivienda en la Florida, Jacksonville, causado por el total de partículas en suspensión y el dióxido de azufre - los principales contaminantes atmosféricos - y el control de la contaminación del aire. Los resultados de este estudio confirman el efecto adverso de la contaminación del aire en los precios de la vivienda. El modelo muestra que las de medidas de control de contaminantes, incrementan el valor de la vivienda, es decir que los compradores y los propietarios tiene en consideración la contaminación del aire, y la existencia de sistemas de control de la misma, al comprar una vivienda (Shaaf, *et al*, 1996)

En 1997 Rossini mostró que los modelos con redes neuronales requieren la separación del conjunto de datos. Debe definirse un conjunto de datos para el entrenamiento del modelo y otro para la validación del mismo. Esta metodología se aplicó en varios países, utilizando conjuntos de datos con las características específicas de cada sitio. Por ejemplo, en 1991, Borst utiliza la RNA para conjuntos de datos de casas de familia en la nueva Inglaterra. Tay & Ho, (1992) examinaron los conjuntos de datos en Singapur con 833 apartamentos

residenciales para la fase de aprendizaje, y los resultados comparados contra 222 apartamentos con características similares. Do & Grudnitski (1992) han utilizado datos de un servicio de listado múltiple en la California, mientras que Evans et al (1993) trabajaron con la vivienda en el Reino Unido. (Worzala, *et al*, 1995), Borst (1995) y Borst & McCluskey, (1996) utilizan varios conjuntos de aprendizaje, y se compararon los resultados obtenidos por las RNA's con los modelos de regresión múltiple.

Según Borst (1995), la exactitud de las RNA las torna rivales de los métodos de regresión lineal múltiple. El autor cree que estos pueden ser utilizados para realizar valoración masiva, como es el caso de los avalúos catastrales, así como un control de calidad sobre los valores estimados por otros métodos.

La capacidad de precisión de las RNA, están perfectamente bien definidas en las investigaciones realizadas por McCluskey y Borst (1997).

En el mismo año, Rossini, basado en el trabajo de sus predecesores, utiliza esta técnica para los datos de ventas registrados por el *Department of Environment and Natural Resources* (DENR) en el sur de Australia. El DENR recoge los detalles de todas las ventas que han ocurrido en el sur de Australia y los pone a disposición en formato digital. Un amplio conjunto de información está disponible para cada propiedad, incluyendo detalles de la venta, la evaluación de los valores, la información sobre la ubicación y características físicas si se trata de propiedades residenciales. En su estudio, Rossini ha utilizado tres procedimientos para comparar los modelos de RNA para estimar el valor del mercado inmobiliario, con el modelo de regresión lineal múltiple (Rossini, 1997). Mientras Borst (1995), Do y Grudnitski (1993) y Tay y Ho (1992) apoyan firmemente el uso de las RNA's, los resultados de Worzala *et al*. (1995) dejan algunas dudas al respecto, pues al trabajar con las mismas variables independientes es posible obtener diferentes resultados en las RNA's, si son utilizados diferentes paquetes informáticos. Worzala *et al*. (1995) trató de replicar las metodologías utilizadas por los investigadores anteriores (Borst 1991, Do y Grudnitski 1992 y 1993) en la aplicación de RNA para el conjunto de datos, utilizando tres metodologías para el análisis de datos.

En el primer enfoque se utiliza todo el conjunto de datos, en el segundo proceso se utilizan las propiedades dentro de un rango de precios como en el estudio de Do y Grudnitski y, en el último caso, se trabajó un pequeño grupo homogéneo de casas similares, como en el estudio de Borst en 1991. También para James y Lam (1996) las dudas sobre la consistencia de las RNA's eran muchas. Ellos creen que más se debe trabajar "en los conjuntos del mundo real, a fin de validar los métodos para su uso en la valoración de la vivienda". Otro resultado curioso de James & Lam, (1996) es que para las muestras de menor tamaño, las redes

neuronales marcan muchos puntos. También Rossini (1997) llegó a la misma conclusión, de que las RNA's parecen ser una mejor herramienta para una serie pequeña de datos, mientras que la regresión es claramente superior para los grandes conjuntos de datos.

Pero este resultado no es unánime. Nguyen y Cripps (2001), han llegado a la conclusión de que una red neuronal tiene mejor rendimiento que un modelo de regresión múltiple, si se les da suficiente información.

En su trabajo de 1997, Rossini llegó a la conclusión de que el uso de modelos de regresión, es preferible en vez de las RNA, advirtiendo sin embargo que estos resultados no son completamente concluyentes. A pesar de que llegó a esta conclusión con un determinado conjunto de datos, cree que en el futuro próximo, con el aumento de las herramientas de cálculo, las RNA's se convertirán en una herramienta de gran poder.

Convencido de lo anterior, Kershaw y Rossini, en 1999, presentan un trabajo titulado *Constant Quality House Price Índices*, en el cual hacen uso técnicas econométricas y redes neuronales. En este artículo se han utilizado las ventas de los tres lugares de Australia del Sur como en las investigaciones previas de Rossini en 1997. Los datos corresponden a todas las transacciones de mercado entre julio de 1980 y junio de 1998. Estos datos han sido analizados con un método hedónico estándar, con variables binarias para indicar los periodos de tiempo. También se aplicó un modelo de redes neuronales con algoritmos genéticos para optimizar la estructura de la red. Este trabajo demostró que las RNA's pueden ser una seria alternativa a los métodos econométricos, confirmando el credo de Rossini.

Sin embargo, Zhang *et al*, en 1998, presentan el estado actual de aplicación de las RNA's en la predicción. El objetivo de estos autores es sintetizar la investigación en esta área, el conocimiento profundo de las técnicas utilizadas a modelar las RNA's y sugerir el camino futuro de la investigación. Hace diez años, los investigadores todavía, no estaban seguros sobre el efecto de los factores clave en el rendimiento de las RNA's en la predicción. Así mismo, estos autores llegan a la conclusión de que las redes neuronales tienen un desempeño satisfactorio en la predicción, en todos los ámbitos. Su capacidad de adaptación, no linealidad y su capacidad de —mapear la función arbitraria (*arbitrary function mapping ability*), son características únicas de las RNA's que hacen con que estas sean totalmente adecuadas y útiles en la tarea de previsión. Los resultados de esta investigación no han sido concluyentes, con respecto a dónde y cuándo es la RNA mejor que lo método de previsión tradicional. Un número considerable de investigadores trabajaron para tratar de llegar a una conclusión. Hay varios factores que pueden afectar el desempeño de la RNA. Sin embargo, no existe una investigación sistemática sobre este tema.

González y Formoso 2000, mostraron como el análisis de regresión lineal múltiple, presenta serias dificultades para hacer frente a la complejidad del mercado inmobiliario, especialmente teniendo en cuenta la correlación espacial y la falta de forma funcional. Estos autores consideran que de todos los atributos, la ubicación es la variable más importante al momento de definir el valor de un predio y es porque se relaciona con la configuración espacial del producto (inmovilidad). El valor de la ubicación de un inmueble está relacionado con las accesibilidades (oferta y calidad de las carreteras y del transporte), con las características del barrio y con el uso del suelo en entorno de la propiedad.

En resumen, desde los años 90 en que las redes neuronales se aplicaran por la primera vez en el sector inmobiliario se aprecia una creciente demanda y la creación de nuevos y mejores algoritmos para el funcionamiento de las redes neuronales. Son también numerosos los estudios que establecen una comparación entre sistemas de inteligencia artificial y métodos tradicionales de evaluación de la propiedad, especialmente con la regresión múltiple. Con este fin, en general, se estima una tasa de error del sistema de inteligencia artificial y otra para el modelo de regresión múltiple, a través de sus aplicaciones a una muestra representativa del mercado, para el que se conoce el precio de venta de la propiedad. Las ventajas de los sistemas de inteligencia artificial, en respecto a los métodos tradicionales, se pueden resumir principalmente en dos: Los sistemas de inteligencia artificial tienen, en las pruebas, errores medios entre el 5 y el 10%, mientras que los modelos de regresión múltiple tienen errores entre el 10 y el 15%. Hay, sin embargo, que destacar que en algunos experimentos se obtienen resultados similares, cuando se trata de muestras homogéneas (Couto, 2007); La segunda ventaja de un sistema de IA es su capacidad para estimar el valor de las viviendas que tienen características significativamente diferentes, es decir, valores extremos, o *outliers*, ya que este tipo de sistemas somete las muestras a procesos matemáticos más complejos que el modelo de regresión múltiple. No obstante, en algunos estudios, los sistemas de inteligencia artificial tienen dificultades para estimar con precisión los valores de las propiedades con características especiales.

Hay sistemas de inteligencia artificial que operan en España, en ciertas áreas, tales como el sistema desarrollado por la Agencia Tributaria para la detección de fraude y evasión fiscal en el impuesto sobre el valor añadido, IVA. En España, y en el reconocimiento del mercado inmobiliario, destacamos las aportaciones de Ceular y Caridad (2000), García Rubio (2004), Gallego (2004) y Lara (2005), con aplicación a Córdoba, Albacete, Madrid y Jaén, respectivamente. En la actualidad, la Dirección General de Registro está desarrollando un proyecto para la estimación del valor de cada propiedad a precios de mercado, con el fin de luchar contra el fraude de bienes inmobiliarios usando redes neuronales. En cualquier aplicación econométrica a la realidad portuguesa, la obtención de datos fiables representa una parte sustancial del trabajo a realizar. En el caso del mercado de la vivienda, las dificultades encontradas se incrementan, ya que no existe una serie temporal de precios de

“venta” suficientemente larga, y que incluye a los distintos atributos residenciales (Carvalho, 1999).

Paula Couto, en su tesis doctoral, entre otras, presenta una aplicación con redes neuronales en una base de datos recogidos por el INE con el apoyo del software Java NNS 10 donde se utilizó una red neuronal backpropagation, formada por tres capas, y cuya función de activación elegida para las neuronas ha sido la sigmoidea. Todavía, en este trabajo no se presenta el resultado de la red. Sin embargo, la autora concluyó que los tipos de evaluación obtenidos por regresión lineal múltiple, y que resultaran en un monomio multiplicativo de coeficientes, representa un modelo bueno para la evaluación en masa para Portugal Continental, con vista a la obtención de valores para tributación de impuestos (Couto, 2007). Este es, sin embargo, el único trabajo que conocemos a la fecha, donde se hace la aplicación de redes neuronales en los datos de evaluación de inmuebles en Portugal, pero sin ningún retorno. Creemos que la ausencia de tales trabajos en nuestro país, se debe principalmente a la dificultad de acceso a los datos. La necesidad de una gran cantidad de información sobre las características de la vivienda, así como el precio de venta real, hace casi imposible hacer este tipo de análisis en Portugal.

En resumen, desde los años 90 en que las redes neuronales se aplicaran por la primera vez en el sector inmobiliario, y como hemos visto a lo largo de estos años han aparecido diversos modelos para la evaluación del valor del mercado inmobiliario en varias regiones del planeta. Estos resultados subrayan la creciente demanda a lo largo del tiempo de nuevos y mejores algoritmos para el funcionamiento de las redes neuronales. Son también numerosos los estudios que establecen una comparación entre sistemas de inteligencia artificial y métodos tradicionales de evaluación de la propiedad, especialmente con la regresión múltiple. Con este fin, en general, se estima una tasa de error del sistema de inteligencia artificial y otra para el modelo de regresión múltiple, a través de sus aplicaciones a una muestra representativa del mercado, para el que se conoce el precio de venta de la propiedad. Las ventajas de los sistemas de inteligencia artificial, en respecto a los métodos tradicionales, se pueden resumir principalmente en dos: Los sistemas de inteligencia artificial tienen, en las pruebas, errores medios entre el 5 y el 10%, mientras que los modelos de regresión múltiple tienen errores entre el 10 y el 15%.

Hay, sin embargo, que destacar que en algunos experimentos se obtienen resultados similares, cuando se trata de muestras homogéneas (Couto, 2007); La segunda ventaja de un sistema de IA es su capacidad para estimar el valor de las viviendas que tienen características significativamente diferentes, es decir, valores extremos, o outliers, ya que este tipo de sistemas somete las muestras a procesos matemáticos más complejos que el modelo de regresión múltiple. No obstante, en algunos estudios, los sistemas de inteligencia

artificial tienen dificultades para estimar con precisión los valores de las propiedades con características especiales.

Es de resaltar que hay sistemas de inteligencia artificial que hoy operan en España, en ciertas áreas, tales como el sistema desarrollado por la Agencia Tributaria para la detección de fraude y evasión fiscal en el impuesto sobre el valor añadido, IVA en España, y en el reconocimiento del mercado inmobiliario, destacamos las aportaciones de Ceular y Caridad (2000), García Rubio (2004), Gallego (2004) y Lara (2005), con aplicación a Córdoba, Albacete, Madrid y Jaén, respectivamente. En la actualidad, la Dirección General de Registro está desarrollando un proyecto para desarrollar la estimación del valor de cada propiedad a precios de mercado, con el fin de luchar contra el fraude de bienes inmobiliarios usando redes neuronales.

Para el caso de Colombia creemos que la ausencia de trabajos en esta materia, obedecen principalmente a la dificultad que se tiene a la consecución de los datos. La necesidad de contar con una gran cantidad de información sobre las características de la vivienda, así como el precio de venta real, hace casi imposible hacer este tipo de análisis en nuestro país.

3.4 Modelos de ubicación de los Inmuebles

En Colombia de acuerdo con la normatividad que rige el catastro, para la estimación del avalúo de un predio debe tenerse en cuenta su localización, pues dentro de la valoración está implícita la identificación del predio, la cual en últimas busca llegar a la localización del mismo en la cartografía oficial (Resolución 70, 2011). No obstante, en la metodología que hoy se tiene definida en Colombia para la estimación del avalúo catastral, en el procesamiento estadístico no se incorpora de manera cuantificable variables de localización del predio, esto a pesar de contar con la disponibilidad de la información, pues el IGAC también tiene la función misional de elaborar la cartografía oficial del país y adicionalmente, cuenta con toda la base predial de los inmuebles de la mayor parte del territorio Colombiano.

De esta forma se está desconociendo que las variables de localización residencial tienen un papel muy importante al momento de determinar el valor de un inmueble, pues a partir de ellas se pretende explicar el motivo por el cual los individuos deciden instalarse en una determinada zona de la geografía urbana y no en otra, es decir, tratan de analizar qué tipo de factores relacionados con la ubicación, tales como el precio del suelo o los servicios propios de una zona específica influyen sobre el precio de la vivienda.

El valor de mercado de cualquier inmueble susceptible de entrar en una transacción, no permanece constante a lo largo del tiempo, sino que experimenta alzas o bajas con el paso del mismo. Este cambio en el valor no sólo es debido a su envejecimiento o estado de conservación, sino principalmente a los cambios que se producen en el entorno económico, parámetros de ubicación en el espacio o características constructivas.

La importancia de cada uno de los factores explicativos depende de los gustos de los compradores y no siempre está estrictamente relacionada con su capacidad económica⁵. En las zonas céntricas la vivienda alcanza mayores precios de los esperados debido a la escasez de suelo y al incremento de la demanda para usos no residenciales⁶ (uso comercial y de oficinas).

La literatura económica ha tratado desde muy diversos enfoques el crecimiento y la composición de los núcleos urbanos, especialmente en países desarrollados. Diversos autores han realizado su aporte a este campo, teniendo en cuenta las diversas fuerzas que intervienen en la localización de un individuo (Follain & Jiménez, 1985 a).

Desde el enfoque clásico (*modelo monocéntrico*) hasta los enfoques actuales se han sucedido los cambios en la composición de las ciudades, por ello los más antiguos han ido perdiendo capacidad explicativa de la realidad urbana para dar paso a las aportaciones más recientes.

De manera general, los modelos planteados giran en torno a dos teorías principalmente: *La teoría de la accesibilidad* y *la teoría de la externalidad*. La primera de ellas basa su fundamento en la facilidad de acceso y su coste, a los centros de decisión y consumo de la ciudad, siendo su punto de arranque el *modelo monocéntrico – disyuntiva o de compensación* -. Por su parte, la teoría de la externalidad aglutina todo aspecto influyente en

⁵ De tal forma que el que busca el centro de la ciudad no tiene en cuenta las razones de entorno, ni las posibles incomodidades de esta localización; el que busca una zona histórica está dispuesto a soportar los problemas derivados de la falta de aparcamiento, los problemas derivados del turismo (ruido y congestión) o la incomodidad en los accesos; el que busca un barrio tranquilo y rodeado de un

⁶ La competencia por las mejores localizaciones se encarga de que los terrenos sean ocupados por los usos más competitivos, esto provoca la estructura espacial del suelo y de los valores. Berry (1978) observa como el precio del suelo está relacionado con el número de peatones que transitan por las avenidas. La competencia por estas localizaciones provoca el alza de los precios de los locales comerciales en los enclaves privilegiados.

la reordenación de los valores de los bienes inmuebles relacionado con la jerarquización social, urbanística, psicológica y económica. Uno de los modelos pioneros y exponente importante en la contemplación de esta teoría es el *modelo de Tiebout*. (Azqueta, 1994).

Al momento de analizar la localización de un predio se destaca el interés constante existente por la estimación de las externalidades tanto positivas como negativas, sobre todo las segundas. Quizás este último es el punto más débil de todas las aplicaciones realizadas, pues la medición de cualquiera de ellas es compleja y desafortunadamente en Colombia poco se ha estudiado sobre como cuantificar el impacto que tiene el deterioro medioambiental sobre el precio de los inmuebles. Lo cual indiscutiblemente sirve de base para la mejora de los modelos de ubicación.

En este capítulo se pretende efectuar un repaso a las principales características de los modelos más significativos aparecidos en el ámbito de la ubicación residencial.

3.4.1 Modelo Monocéntrico

También llamado *modelo disyuntiva o modelo de compensación*. En éste la idea principal es la elección que realiza el comprador entre el lugar de ubicación del predio y su cercanía al centro de la ciudad (llamado Centro Comercial de Negocios, CCN; en inglés, Central Business District, CBD) (Wilkinson, 1973). Bajo la hipótesis de que la superficie de vivienda, medida en metros cuadrados totales es más barata en la periferia que en el C.B.D⁷ y que a su vez desplazarse desde la periferia al mismo conlleva costes de transporte para ir al trabajo, de compras, etc. (medidos en dinero y tiempo), el comprador tiene que elegir, para una renta dada, entre espacio y cercanía, ver Figura 5.

Desde esta óptica, los modelos Monocéntrico ofrecían como única variable explicativa la minimización de los costes, en lo que se denominaba hipótesis de compensación *Espacio /acceso*.

⁷ También existen otros estudios que demuestran que no se trata de una elección tan simple, ya que su estudio de la ciudad de Leeds revela que los servicios ofrecidos por las viviendas son mayores conforme se aleja uno del centro de la ciudad (Wilkinson, 1973),

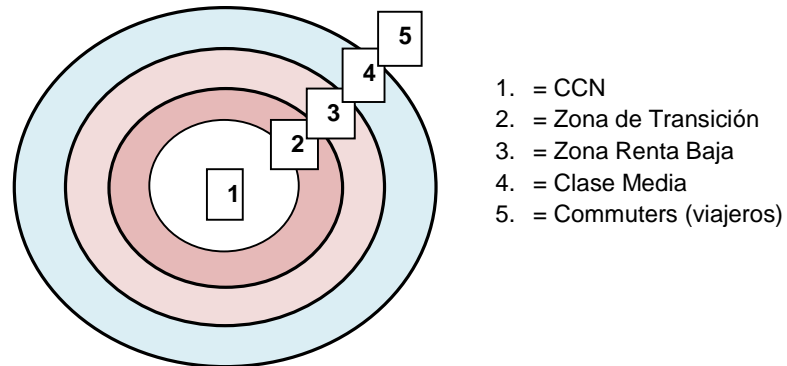


Figura 5. Teoría de las Zonas Concéntrica

Fuente: Chica Olmo, (1994). Elaboración propia

El enfoque Monocéntrico, así mismo, ofrecía un marco teórico⁸ en el que se encuadraban necesariamente dichos trabajos, en el que los supuestos básicos eran (Solow, 1972):

- a. No externalidad en el consumo de la vivienda.
- b. Incremento del coste del transporte conforme nos alejamos del C.B.D.
- c. Elasticidad precio constante y elasticidad renta unitaria en la demanda de vivienda.
- d. Por implicación, que vivienda y coste del transporte sean parte significativa del presupuesto familiar.
- e. Supuesto adicional, los puestos de trabajo se hallan localizados en el C.B.D.

En forma sintética, y aceptando dichos supuestos, el modelo de compensación quedaba explicado como una función de densidad exponencial negativa en la forma, (Goldstein & Moses, 1973):

$$D(U) = D_0 e^{-\alpha u}$$

Ecuación 12. Modelo de compensación

⁸ Revisado y actualizado desde la Nueva Economía Urbana (N.E.U.), encabezado por (Solow, 1972).

Donde:

$D(U)$ es la densidad residencial, α el gradiente de densidad, U la distancia al C.B.D., y D_0 , es la densidad junto al C.B.D.

El modelo predice que al incrementar la población metropolitana, D_0 también aumenta, aunque no existe predicción alguna sobre el valor de α . Sin embargo, se puede demostrar empíricamente que las ciudades mayores están más descentralizadas y que α será menor en tamaño relativo, a medida que las áreas metropolitanas mantengan sub-centros de compra y empleo (Henderson J. , 1985).

Teniendo en cuenta el Modelo Monocéntrico, se decidió incorporar al modelo de RNA la variable de localización, Distancia al Centro, que como su nombre lo indica, es la distancia que existe entre el lugar de ubicación del predio y el centro de la ciudad (CCN), la cual será evaluada para estimar el modelo.

3.4.2 El Stock de Capital No Maleable

En las aportaciones de Harrison y Klain (1974) y Anas (1978), se realiza una revisión del modelo monocéntrico y se parte del fundamento de que el stock de capital no es maleable, y además no es muy duradero. Los motivos para la variación de tal supuesto son bastante obvios. Por un lado, la vivienda es un bien intrínsecamente duradero, por ello las entidades financieras acceden a su financiación a mayor plazo que cualquier otro bien. Por otro, no se observa por lo general capacidad en las autoridades locales para ir regenerando continuamente las ciudades, intentando minimizar el deterioro sufrido en las mismas por el paso del tiempo (Alonso, 1964).

El crecimiento urbanístico tiene lugar por medio de círculos: el centro histórico, el anillo a su alrededor, el primer ensanche, el segundo, etc. Cada uno de ellos responde a distintos períodos de tiempo y por eso ofrece distintos servicios, aunque es muy probable que los últimos los ofrezcan de mucha mayor calidad que los primeros.

Los nuevos ensanches pueden ir mermando la importancia del CCN como única fuerza determinante del precio, ya que las nuevas zonas tienen mejores servicios que las antiguas, y también tiran de la demanda de localización¹⁸.

Por último, conviene hacer referencia al deterioro y abandono que están sufriendo los centros de las ciudades, como consecuencia de la emigración a zonas con mejores equipamientos

e infraestructuras¹⁹ y la falta de actuación de las autoridades en general, ofrece una evidencia empírica

3.4.3. El modelo de Tiebout

Tiebout expuso que la elección de los compradores está basada en las preferencias por determinadas regiones, zonas o entornos, por ejemplo, la demanda de localización viene definida por las condiciones en que se encuentra el entorno en que se ubica la vivienda, hecho que puede o no verse influenciado por la distancia al C.B.D.

Esta postura implica un concepto mucho más amplio de accesibilidad, teniendo en cuenta otra serie de factores no contemplados en los modelos clásicos. Por ello, en este enfoque "heterodoxo" se integran los modelos basados en economías de escala, o modelos de aglomeraciones, herencia del análisis del suelo de uso industrial.

Las sucesivas ampliaciones que ha sufrido este modelo ponen de relieve la existencia de un número notable de factores que pueden incidir en las decisiones de los compradores.

Mellado, (2002) clasifica las externalidades en tres grupos:

- **Externalidades Físicas**

Aúna toda una serie de elementos de carácter urbanístico – ambiental que influyen directamente en el valor del inmueble entre los que puede destacarse la existencia de servicios públicos (tanto en cantidad como en calidad) en la zona donde se ubique el bien inmueble a valorar, tales como parques, zonas ajardinadas, energía eléctrica, telefonía, gas, alcantarillado, parqueaderos, etc. A ellos añadir otros elementos de carácter asistencial, educacional, comercial, de administración, financiero, e incluso ocio, que prestan un servicio muy apreciado por la ciudadanía, incluso la situación de proximidad a de servicios de trasportes (Mellado, 2002).

- **Externalidades Sociales**

Parte del presupuesto básico de que el suelo no es sólo un medio de producción, sino que se manifiesta en gran número de ocasiones como símbolo de estatus social, por lo que según el barrio en el que se encuentre situado un inmueble, la sociedad interpreta una situación del estado personal de sus ocupantes – ya sea en cuanto a renta

económica, nivel cultural, situación profesional -, repercutiendo este aspecto en la aceptación en un determinado estamento social (Mellado, 2002).

En efecto, el ciudadano tiene un interés notable por demostrar al resto su ascenso social (Boleat, 1973), y para ello nada mejor que la compra de una vivienda en un barrio de mayor prestigio que el de procedencia, puesto que la vivienda es el activo más visible de todos los que puede exponer una familia, (Caridad & Brañas, 1998).

▪ **Externalidades Económicas**

Hay veces en que la motivación inversora se fundamenta en razones de tipo psicológico, entre los que se hallarían los motivos de previsión de aceptación social de los valores del mercado inmobiliario que se alcanzarían a lo largo del desarrollo de una celebración a nivel internacional o nacional (con carácter cultural, deportivo o comercial, por ejemplo) en un momento determinado del tiempo y en un lugar concreto (Boleat, 1973).

Otros factores de índole económica a los que puede efectuarse mención son la zonificación, las especulaciones y acciones monopolistas sobre el mercado o los factores históricos que impliquen cierto grado de protección para una determinada localización.

Una de las causas del elevado precio de las urbanizaciones de lujo alejadas del centro urbano⁹ hay que buscarla en la existencia de demandantes de mayor renta que están dispuestos a ofrecer más, porque desean separarse de los demás y, por tanto, hacen aumentar los precios de esas viviendas. Es decir, las razones de entorno que motivan estas demandas de localización y, por tanto, los aumentos de precios no son estrictamente naturales (menor ruido, menor contaminación...), sino que responden a factores sociales: el entorno que se busca no es verde, sino de éxito social. Además, se trata de un proceso circular en el sentido de que si se concentran habitantes de prestigio cada vez habrá más y cada vez las viviendas serán más caras (Caridad & Brañas, 1998).

⁹ Ya que las instalaciones tienen un coste, pero que al dividirse entre toda la promoción no será tan elevado, y probablemente se verá compensado por un suelo mucho más barato.

Teniendo en cuenta el Modelo de Tiebout, se decidió incorporar al modelo de RNA las variables de localización, Usos del Plan de Ordenamiento Territorial (USOPOT) y Comuna, las cuales son evaluadas por el IGAC en el proceso del estudio de zonas homogéneas que se adelantó y cuyas características se detallaran en el capítulo de metodología, esto debido a que las mismas se corresponden a una forma de zonificación del territorio.

3.5 Aplicación de la Geomática para el cálculo de las variables de localización.

La medición de los efectos de la ubicación espacial en el valor de los inmuebles, es difícil, porque no son directamente cuantificables. Se usan variables *proxy* como la distancia al centro de comercio o urbano. Es aquí donde la Geomática como ciencia de información Geoespacial se constituye en una herramienta fundamental para la identificación de las variables de localización, porque a través de los Sistemas de Información Geográfica se puede manejar datos espaciales geográficos caracterizados por aportar información relativa a la posición absoluta y relativa con otros aspectos y detalles de características no-espaciales.

Los SIG llevan en su nombre la íntima relación de uno de sus orígenes, la Geografía, la cual, junto con la Cartografía y otros avances científicos y tecnológicos, que en paralelo se lograron, permitieron la formalización y desarrollo de la Geomática.

La localización afecta al valor de las viviendas a causa de la inmovilidad, (Kiel & Zabel, 2008). Basados en esta afirmación realizó un estudio para determinar el precio de una propiedad a través de un software diseñado para este propósito, el *Metropolitan Statistical Area – MSA*. En este software se especifican condiciones de la vivienda tales como la temperatura, la proximidad a las zonas con agua y lugares de interés cultural, son atributos que explican que viviendas similares, tienen precios diferentes en el MSA. En este programa, ha sido establecida una base de información única, con base en el *American Housing Survey (AHS)*, constituida por pequeños grupos de unidades de viviendas, los “*clusters*”, con información sobre la estructura y características residenciales que han sido fundidos con los atributos resultantes de los niveles de regiones. Los autores llegaron a la conclusión que los tres niveles de localización son importantes cuando se estima el precio de la vivienda a través de un modelo.

Kauko, (2003) publicó un artículo con el fin de evaluar los beneficios y desventajas de los modelos de redes neuronales para evaluar el valor del mercado inmobiliario. En particular, ha hecho una comparación con los mapas de Kohonen (1997), con los modelos hedónicos. Tiene particular interés para estudiar cómo los factores de localización, ya sea ambiental o

social, pueden afectar tanto la demanda del mercado, como el nivel de precios en estos lugares. En este estudio, el autor analizó el mercado de la vivienda en Helsinki - Finlandia, con un total de 6200 observaciones y 10 variables de entrada. Así mismo, Kauko considera que las redes neuronales son una poderosa herramienta para la evaluación de propiedades en masa.

García, *et al* (2008), presentaran un sistema automatizado para la evaluación de la propiedad, que combina el uso de redes neuronales artificiales con un Sistema de Información Geográfica. La elección de los autores se basa en que ambos instrumentos han demostrado ya su potencialidad en el campo de la investigación económica. En 2002, Thurston dijo que una RNA vinculada a un SIG puede ser utilizada para simular el funcionamiento del cerebro humano a procesar los problemas de datos espaciales. Existen varias aplicaciones en que la unión de las RNA's con el SIG ha llegado a ser muy útil. Por ejemplo, podemos mencionar el uso del suelo, la oceanografía, la deforestación, la evaluación del ruido generado por el aeropuerto, entre otros. Thurston, ha mostrado como algunos modelos de redes neuronales y un sistema de información geográfica puede ser combinado para constituirse en una herramienta poderosa en la economía (Thurston, 2002).

Noelia García y otros investigadores en 2008, han dicho que cualquiera que sea el enfoque, el análisis puede ser mejorado mediante la integración de un Sistema de Información Geográfica (García, Gámez, & Alfaro, 2008). En su trabajo, estos investigadores utilizaron los modelos de RNA's Perceptrón Multi-capas, las funciones de bases radiales y los mapas de Kohonen. Los dos primeros modelos son una interesante alternativa a los modelos de regresión tradicionales, mientras que los mapas de Kohonen (SOM), son especialmente destinados a tareas de clustering. Así que los dos primeros modelos han sido utilizados para estimar el precio de los inmuebles, mientras que los mapas de Kohonen se utilizaran para tareas intermedias relacionadas con la estimación de los valores en falta, para diversas variables cualitativas, como la calidad de la propiedad.

En este interesante estudio, los autores combinaron la RNA con un SIG, con un sistema automático para estimar el precio de una casa en Albacete, España. Mediante un simple clic, sobre el mapa de esta ciudad, y una vez atendidas las características del sistema de la vivienda deseada, el sistema devuelve la estimación para el precio de la casa. Los resultados de rendimiento de los modelos se compararon y se obtuvo una mayor precisión con la RNA estimada en la determinación del precio total, con un R^2 de 92% y un error relativo medio de 5,56%. Los autores sospechan que el motivo de estos resultados está relacionado con el tamaño de la muestra disponible (591), pequeña para las necesidades de la red de funciones de base radial. El análisis de sensibilidad demostró que la variable más importante es la distancia al centro de negocios, con una pendiente negativa en el supuesto monocéntrico.

Otro resultado del estudio de Noelia y otros, de 2008, ha sido la no-linealidad existente en la relación entre el efecto edad en el precio de la vivienda (García et al., 2008). En este sentido, es importante mencionar la capacidad de los modelos neuronales para la detección de relaciones no lineales, que no pueden ser detectadas por modelos más tradicionales.

Existen otros estudios realizados en España, como los de Caridad & Ceular, (2001), Gallego, (2004) y Lara Cabeza, (2005), los cuales realizaron sus investigaciones en diferentes zonas urbanas. Caridad, *et al* (2009), utilizó una muestra con más de 10.000 operaciones, recogidas en estudios anteriores entre 2002 y 2006, lo que permite también hacer comparaciones en el tiempo. En el contexto que estamos viviendo actualmente, es urgente encontrar maneras objetivas para determinar el verdadero valor de las propiedades. En España, el precio de las viviendas es recogido por el INE y los municipios, a efectos fiscales, y no se concentran en la evaluación real de la propiedad individual. Así, ha sido a través de encuestas y entrevistas con agentes de la propiedad inmobiliaria que los datos de este estudio han sido recogidos. Usaron una red neuronal multi-capa para modelar el precio, con seis variables de entrada, de las cuales la más explicativa de los precios ha sido la superficie, seguido de la *ubicación* y el costo de los gastos mensuales (Valor de la administración). Los años de la construcción, la existencia de depósito y garaje, y el índice de calidad, han demostrado menos importancia relativa, pero, sin embargo si se retiraban del modelo, el resultado final es menos bueno. Caridad considera que el uso de las RNA es más flexible que los tradicionales modelos econométricos, cuando si dispone de un conjunto de datos suficiente, (Caridad et al., 2009)

Kauko, (2009) presentó un estudio que trata de los modelos de *localización* residencial en Budapest, con un enfoque con redes neuronales, conocida como los mapas de Kohonen (SOM – *Self – Organising - Map*). Se trata de un tipo de clasificación de redes neuronales. Una vez identificados algunos elementos urbanos, el análisis se centra en la dinámica de dos barrios del interior de la ciudad de Budapest.

A lo largo de la revisión de literatura se ha mostrado como hoy los Sistemas de Información Geográfica se han convertido en una herramienta fundamental para agregar el componente espacial dentro de las variables que se tienen en cuenta al momento de definir el avalúo de un predio. La característica de inmovilidad de una vivienda, hace que, por definición, su mercado sea espacial, razón por la cual le son intrínsecas, de forma natural, las asociaciones y tendencias en cuanto al espacio. No tener en cuenta los efectos espaciales y no incorporarlos en la modelación de precios de la vivienda, puede llevar a que los resultados obtenidos presenten problemas.

4. Metodología

4.1. Municipio de Fusagasugá – Caso de Estudio

4.1.1 Entorno físico y geográfico

El Municipio de Fusagasugá es la capital de la provincia del Sumapaz en el departamento de Cundinamarca, fundada el 15 de enero de 1592, con una superficie de 239 km². Es conocido como "*Ciudad Jardín de Colombia*" o "*Tierra Grata*". Actualmente es una de las ciudades con más desarrollo de la región central, centro económico y de servicios del sur del departamento. Sus principales sectores productivos son el comercio, la construcción y el mercadeo agropecuario, ver Figura 6.

Fusagasugá se encuentra ubicada entre los 4° 20' latitud norte y los 74° 21' 00" longitud oeste, la ciudad ocupa una verde meseta enmarcada por los ríos Cuja y Chocho y los cerros Quininí y Fusacatán que conforman el valle de los Sutagaos. Dista 64 kilómetros (39,76 millas), hacia el Sur Oeste, de la capital del país, Bogotá., En la Tabla 2. se muestran los municipios con los cuales limita.

Tabla 2. Límites Municipio Fusagasugá

Límites Municipio Fusagasugá	Municipio
Al Norte	Silvania y Sibaté
Al Sur	Arbeláez
Al Oriente	Pasca y Sibaté
Al Occidente	Tibacuy, Silvania e Iconzo

Fuente: Oficina Asesora de Planeación, 2010.

El territorio fusagasugueño se encuentra entre los 550 a los 3.050 metros sobre el nivel del mar. El perímetro urbano se encuentra en una altura promedio de 1.726 msnm con una temperatura promedio de 20 °C.

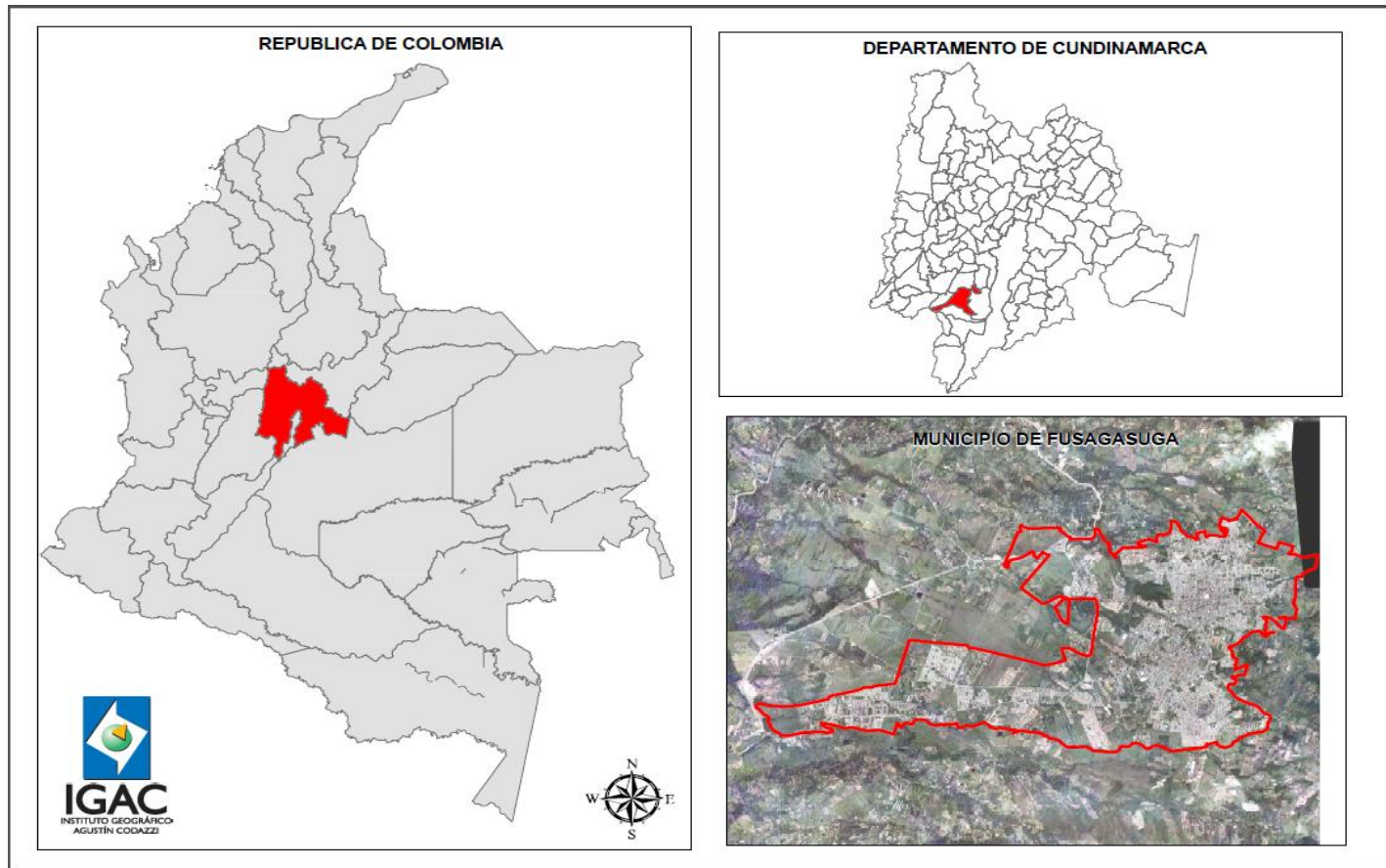


Figura 6. Ubicación general municipio de Fusagasugá

Fuente: Ortofoto y Cartografía oficial IGAC

4.1.2 División político administrativa

En desarrollo del artículo 318 de la C.N. y la ley 136 de 1994 y mediante el Acuerdo Municipal No. 64 de julio de 1995, se crearon en Fusagasugá las Juntas Administradoras Locales y su división en comunas y corregimientos. Mediante el Decreto 120 de 2007 se establece la nueva delimitación de comunas y corregimientos sin variar el número de las mismas, sólo el área, quedando definidas como se muestra en la Tabla 3.

Tabla 3 Distribución de Comunas Municipio de Fusagasugá

Zona Rural		Zona Urbana	
Corregimiento	Área KM ²	Comuna	Área KM ²
Nororiental	34,79	Norte	1,48
Oriental	17,94	Centro	0,63
Suroriental	59,91	Oriental	1,54
Suroccidental	48,76	Suroriental	1,49
Occidental	29,59	Occidental	4,08
TOTAL	190,980	Suroccidental	3,81
TOTAL		13,019	

Fuente: Oficina Asesora de Planeación, 2010.

4.1.3 Información Predial

- **Distribución por rangos de avalúo y superficie**

En la actualidad la zona urbana del Municipio cuenta con 44.951 predios urbanos, y un avalúo catastral global urbano de \$2.977.017.785.500. Del análisis del comportamiento de la distribución por rango de avalúos catastrales, se observa que el 36% de la totalidad de los predios registran un avalúo menor a 25 millones de pesos, valor que resulta distante del valor comercial que hoy tiene la propiedad en este municipio, pues difícilmente se puede adquirir un predio de medianas características, por un valor de \$25.000.000. Ver Figura 7

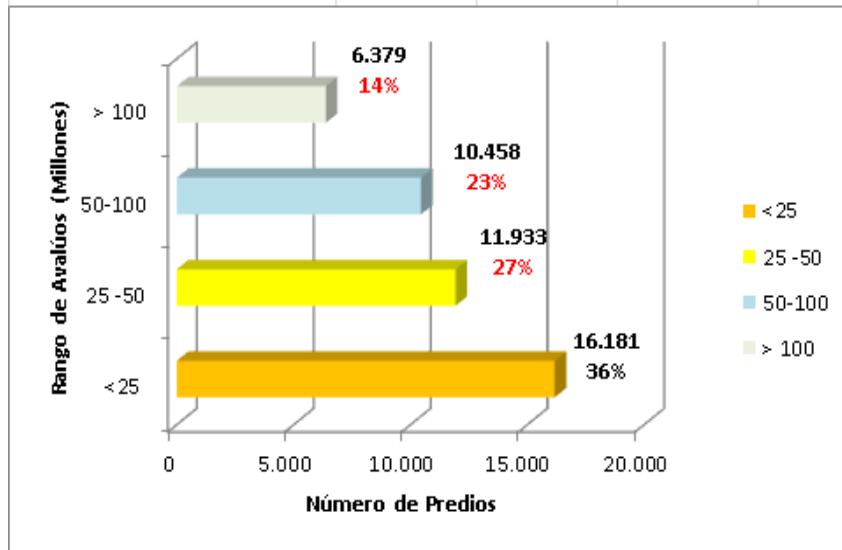


Figura 7. Distribución de predios por rango de avalúos zona urbana municipio de Fusagasugá

Fuente: Instituto Geográfico Agustín Codazzi 2014, elaboración propia

4.2 Materiales

En la Tabla 4 se relacionan los insumos empleados para la elaboración del presente trabajo de investigación.

Tabla 4. Relación de los materiales empleados en la investigación

Insumo	Escala	Fuente	Año	Características
Ortofoto	1:10.000	IGAC	2005	Escala 1:10.000, año 2009. Ver anexo B
Cartografía Digital	1:2.000	IGAC	2014	La información cartográfica digital de la vigencia 2005 está referida al sistema de coordenadas planas (Transversa de Mercator) "MAGNA Colombia Bogotá", lo que es adecuado solo para la zona rural ya que la zona urbana requiere de mayor precisión.
Registros	NA	IGAC	2014	Registros 1 y 2 del Sistema de Información Catastral del IGAC
Avalúos	NA	IGAC	2014	Puntos de investigación e investigación indirecta del estudio de Zonas Homogéneas Goeconómicas de la zona urbana del municipio de Fusagasugá.
POT	NA	MPIO	2001	Acuerdo 029 de 2001, usos del POT, Comunas
Software	NA	NA	NA	SPSS V 21. Para el procesamiento de la Red Neuronal Artificial

4.3 Métodos

4.3.1 Constitución de la muestra

Los datos fueron obtenidos de las Memorias del Estudio Zonas Homogéneas Geoeconómicas de la Zona Urbana del Municipio de Fusagasugá, realizado por el Instituto Geográfico Agustín Codazzi – Territorial Cundinamarca, obtenidos durante la fase de campo durante el segundo semestre de 2012. Los avalúos catastrales resultantes del citado estudio fueron puestos en vigencia para efectos fiscales el primero de enero de 2013.

La información económica inherente a los avalúos comerciales con los que se alimenta la RNA, fueron obtenidos de los puntos de investigación y de la investigación de mercado indirecta (avalúos). De igual forma, los demás datos relacionados con los aspectos físicos de cada predio, fueron obtenidos de los registros 1 y 2 del Sistema de Información Catastral y de la cartografía digital con su respectiva geodatabase.

.A continuación se describe las variables directamente relacionadas con los atributos intrínsecos o propios de la vivienda objeto de estudio, distinguiéndose dos agrupaciones:

- a. **Variables físicas del predio.** Estas variables corresponden a la información del predio levantada en campo por el IGAC en la labor de identificación predial, y fueron tomadas de la base de datos del Sistema de Información Catastral del IGAC.
- b. **Variables de ubicación espacial:** La construcción de un modelo exige que también haya variables que indiquen la mejor o peor situación de los inmuebles. Estas variables de situación hay que construirlas, y aportárselas a la muestra. Son responsables, en buena parte, del éxito de la valoración.

En esta categoría incluimos dos variables que se obtuvieron del Plan de Ordenamiento Territorial del municipio, las cuales fueron: el Uso permitido del suelo y Comuna, la primera de ellas es una variable que es empleada por el IGAC, en sus estudios de zonas homogéneas físicas, y la segunda hasta el año 2013 no se tenía en cuenta en ningún proceso catastral. No obstante lo anterior, es preciso mencionar que las dos revisten importancia pues son variables en las cuales se zonifica el territorio, lo cual tiene connotaciones de tipo socioeconómico que inciden directamente en el valor del predio.

En cuanto a la forma de obtener variables de localización, esta se realizó zonificando el territorio. La zonificación es el sistema más extendido para aportar variables de situación a la muestra. Para el cálculo de las variables de localización resulta indispensable establecer la relación entre el dato, que para este caso es el predio, con su ubicación geográfica. Es aquí donde los Sistemas

de Información Geográfica tienen un papel fundamental - SIG, pues son la integración del elemento geográfico y alfanumérico, en donde el mapa viene a ser una fuente de datos, así como también, una estructura para almacenarlos y un instrumento de análisis y de despliegue. Por ello, para realizar la estimación de las variables de localización de este proyecto resulta indispensable el uso del plano de conjunto urbano, la ortofoto y la base alfanumérica del Municipio de Fusagasugá. Las variables que fueron evaluadas en la RNA, se muestran en la

Tabla Relación de variables empleadas en la RNA.

Variable	Convención	Unidades	Descripción
Variables Físicas del Predio			
Área Terreno	ARTERR	M ²	Medida en metros cuadrados de terreno
Área Construida	ARCONST	M ²	Medida en metros cuadrados de construcción
Puntaje Estructura	PEST	Cantidad	Calificación de valoración que da el IGAC a este ítem
Puntaje Cocina	PCOON	Cantidad	Calificación de valoración que da el IGAC a este ítem
Puntaje Acabados	PACAB	Cantidad	Calificación de valoración que da el IGAC a este ítem
Puntaje Baños	PBAÑO	Cantidad	Calificación de valoración que da el IGAC a este ítem
Puntaje Total	PTOTAL	Cantidad	Calificación de valoración que da el IGAC a este ítem
Nº Pisos	PISOS	Cantidad	Número de niveles que tiene un predio
Nº Baños	BAÑOS	Cantidad	Número de baños que tiene un predio
Edad de la Construcción	EDAD	Años	Tiempo transcurrido en años desde la fecha de construcción
Variables de Localización			
Uso POT	USOPOT	NA	Es el uso permitido por norma según POT
Comuna	COMUNA	NA	División político administrativa aprobada en la zona urbana
Coordenada X	XCOORD	NA	Representa la distancia del punto sobre el eje X
Coordenada Y	YCOORD	NA	Representa la distancia del punto sobre el eje Y
Distancia al Centro	DISTAN	M ²	Recorrido en metros del punto de ubicación del predio al Centro

NA: No Aplica

4.3.2 Análisis previo de los datos antes de ser procesados por la RNA

- **Detección de valores atípicos y ausentes de la muestra**

El estudio se inició con una muestra con 991 predios urbanos, con uso residencial, los cuales fueron sometidos a un cuidadoso proceso de depuración, que consistió en la eliminación de datos atípicos y extremos. Esta labor de filtrado se realizó, a partir de la experticia del investigador y con el apoyo de la estadística exploratoria y el estudio de los gráficos dispersión que fue realizado para cada variable, así como las relaciones existentes entre ellas.

Con relación a los predios con datos ausentes, estos fueron eliminados, debido a que se contaba con una muestra de tamaño suficiente, por lo tanto solo fueron utilizadas observaciones con datos completos. Fue así, como se eliminó o descartó 306 predios, lográndose finalmente una muestra efectiva de **685 predios**.

Los predios que fueron eliminados correspondieron a observaciones que tenían información incompleta o simplemente porque no se logró su ubicación en la cartografía, por no conocer su número predial, situación que se presenta en los casos que corresponden a los avalúos suministrados por bancos, pues en la mayoría de ellos no se relaciona el número predial, situación que también se presenta en las ofertas, en donde difícilmente un propietario conoce este dato de su predio, y no siempre quien manifiesta ser el propietario del predio en oferta es el que figura en el catastro, por lo tanto muchas veces no es posible hacer su conexión exacta con la base de datos.

- **Pre-procesamiento y pos-procesamiento de la muestra**

Una vez depurada la muestra esta fue sometida a un proceso de pre-procesamiento y pos-procesamiento de los datos para cada una de las variables que consistió normalización de las variables cuantitativas y codificación de las cualitativas, para que de esta forma pudieran ser analizadas e incluida en el modelo. La idea es ajustar los datos de alguna manera tal que la red, los consiga manejar eficientemente, es decir, proceder a su normalización (Azoff, 1995). A continuación se detalla cada una de ellas:

- Área Terreno:** Esta variable cuantitativa representa el número de metros cuadrados de terreno que posee la vivienda, la cual es calculada directamente del plano digital del municipio. Este testigo es fundamental por la capacidad explicativa que tiene. Dado que se encuentra medida en metros, el rango de valores es bastante alto, por tanto fue necesario normalizarla de manera que tome valores en el rango [0,1].
- Área Construida (C Privada):** Variable cuantitativa medida en metros cuadrados de construcción y al igual que la anterior fue necesario normalizarla.
- Puntaje Estructura, Puntaje Cocina, Puntaje Acabados, Puntaje Baños y Puntaje Total:** Variables cuantitativas que corresponden a la calificación de valoración de la construcción que da el IGAC a estos

ítems. Como en los anteriores casos fue necesario, por tanto, normalizarlas, de manera que tomen valores en el rango [0,1].

- d. N° Pisos:** Esta variable cuantitativa aunque es de recorrido corto, se normalizó en el rango de [0,1]
 - e. N° Habitaciones:** Al igual que la variable anterior, es de recorrido corto, pero también se normalizó.
 - f. N° Baños:** Variable cuantitativa que se normalizó en el rango de [0,1].
 - g. Edad:** Esta variable Cuantitativa que corresponde a los años que posee la construcción. También en este caso se lleva a cabo la normalización de la variable.
 - h. Coordenada X (X Centroide) y Coordenada Y (Y Centroide):** Estas dos variables fueron obtenidas con el SIG del municipio de Fusagasugá, con el uso del software de ArcGIS. Originalmente las dos variables cuantitativas vienen medidas en metros, por lo que el rango de valores es bastante alto, por lo tanto, fue necesario normalizar ambas variables de manera que tomen valores en el rango [0,1].
 - i. Distancia:** Esta variable al igual que las coordenadas, fueron obtenidas con el SIG del municipio de Fusagasugá, con el uso del software de ArcGIS, y corresponde a la distancia euclidiana.¹⁰ Que es medida en metros en línea recta desde el punto donde se ubique el predio hasta la plaza central del municipio Centro Comercial de Negocios (CCN). Los valores obtenidos en la mayor parte de los casos son altos, por lo tanto también fue necesario normalizarla. Ver Figura 8.
-

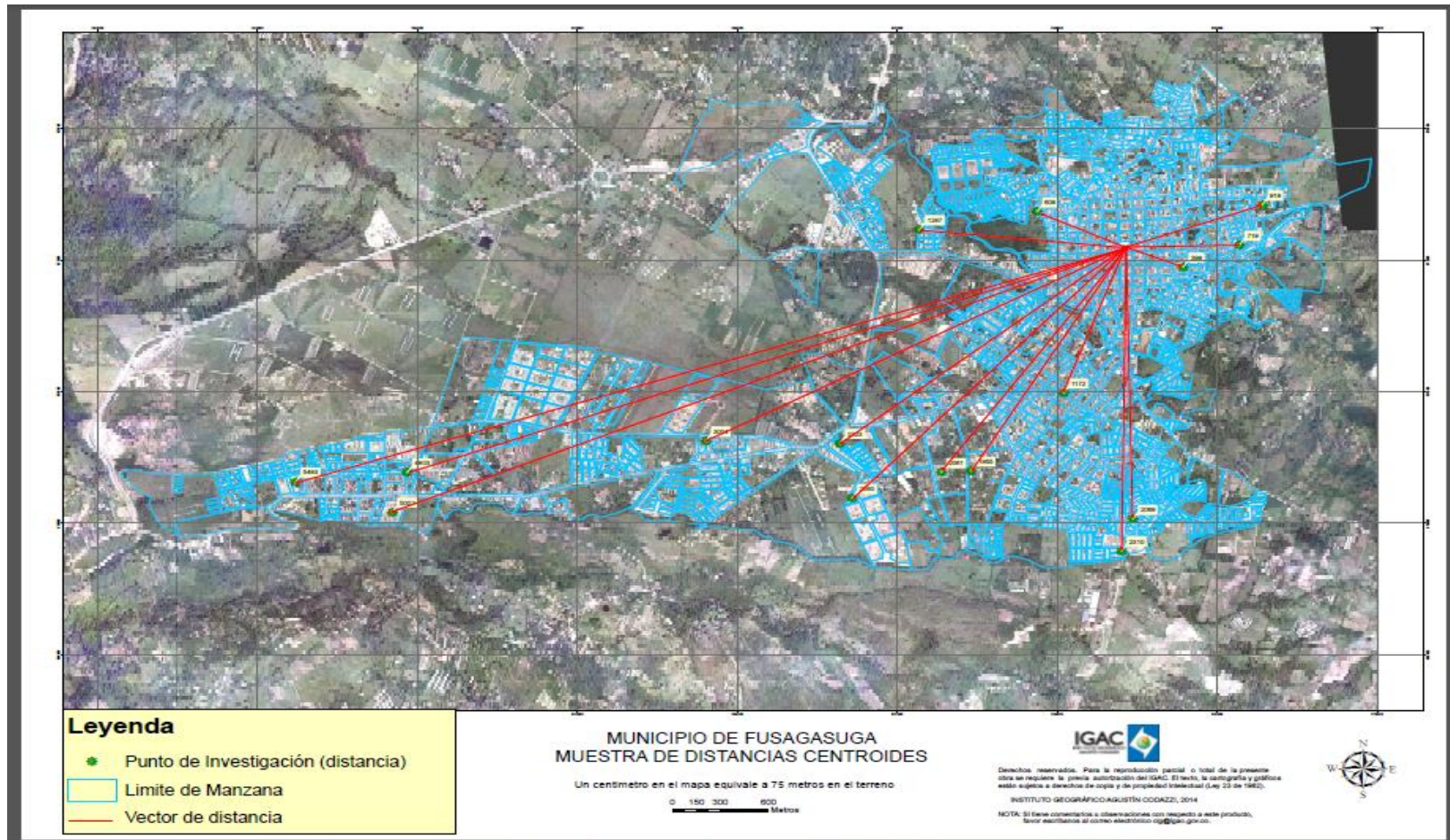


Figura 8. Plano de conjunto urbano Municipio de Fusagasugá cálculo de distancia al centro de la ciudad

Fuente: Ortofoto y Cartografía oficial IGAC

- j. **Uso Plan de Ordenamiento Territorial (USOPOT):** Esta variable cualitativa se obtuvo del plano digital de zonas homogéneas físicas del municipio y corresponde al uso permitido del suelo que está aprobado en el Plan de Ordenamiento Territorial. En la **Tabla 5** usos encontrados en la muestra.

Tabla 5. Relación de Usos del suelo encontrados con codificación

Usos del Suelo POT- cualitativa	Codificación Variable
Área de Desarrollo Habitacional de Densidad Alta	1
Área de Desarrollo Habitacional de Densidad Media	2
Zona de Comercio Tipo 3 (C-3) Especial	3
Zona de Comercio Tipo	4
Área de Actividad Múltiple	5
Ronda Hídrica	6
Área de Actividad Institucional	7
Área de Desarrollo Habitacional de Densidad Baja	8
Área de Desarrollo Habitacional	10

En la Figura 9 se muestra la espacialización de los usos permitidos dentro de la muestra.

- k. **Comuna:** Esta variable cualitativa corresponde a la división político administrativa que tiene la zona urbana, la cual está conformada por 6 comunas, como se muestra en la Tabla 6.

Tabla 6. Relación de Comunas del Municipio de Fusagasugá y la conversión a variable cuantitativa

Comuna	Conversión a variable Cualitativa
Norte	1
Centro	2
Occidental	3
Oriental	4
Suroccidental	5
Suroriental	6

La ley 136 de 1994, dio vida a las comunas que son unidades administrativas del área urbana, que agrupa sectores o barrios. La división Comunal se crea con el fin de mejorar la prestación de los servicios y asegurar la participación de la ciudadanía en el manejo de los asuntos públicos de carácter local. Ver Figura 11.

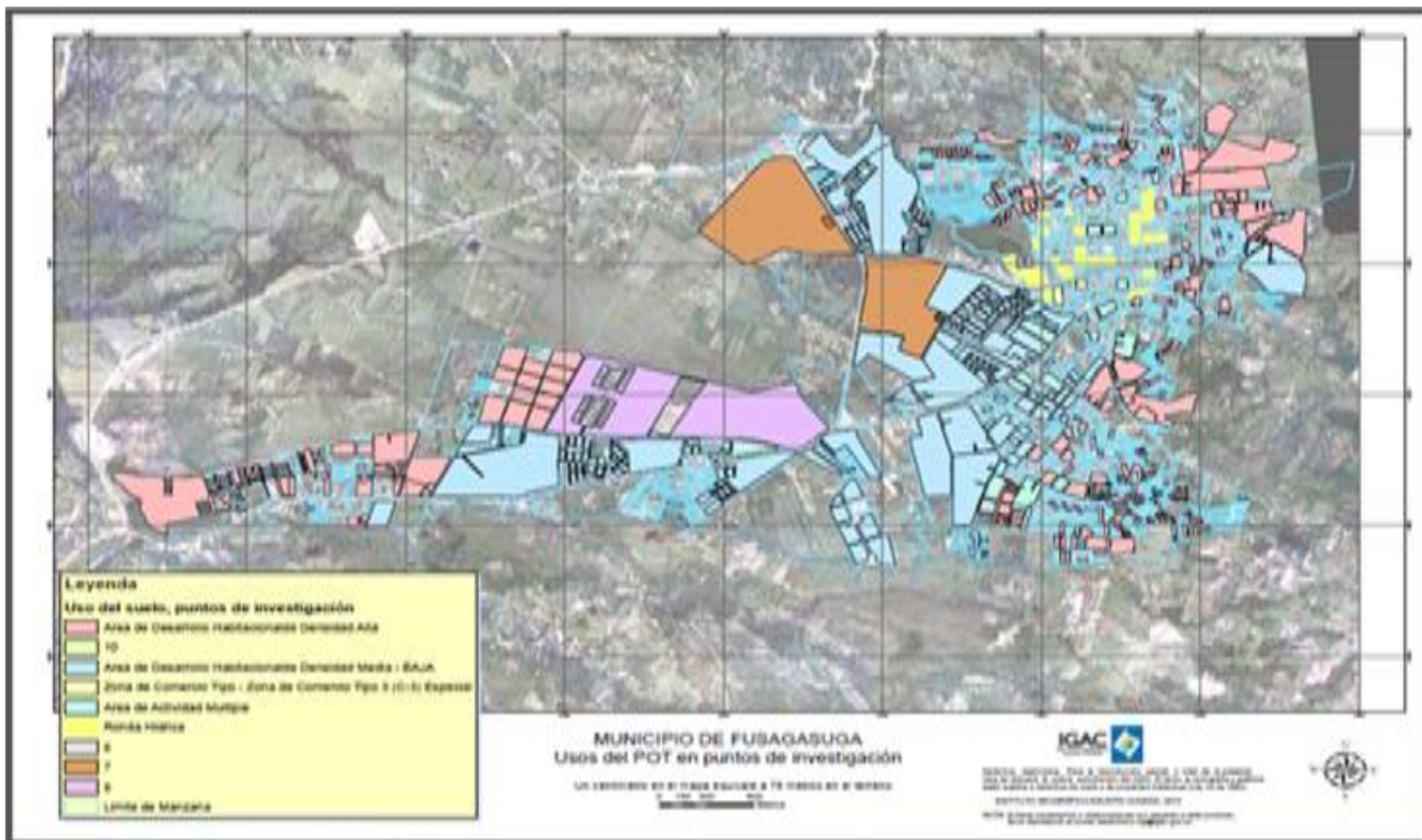


Figura 9. Plano de conjunto urbano Municipio de Fusagasugá con espacialización de Usos permitidos POT

Fuente: POT, Ortofoto y Cartografía oficial IGAC

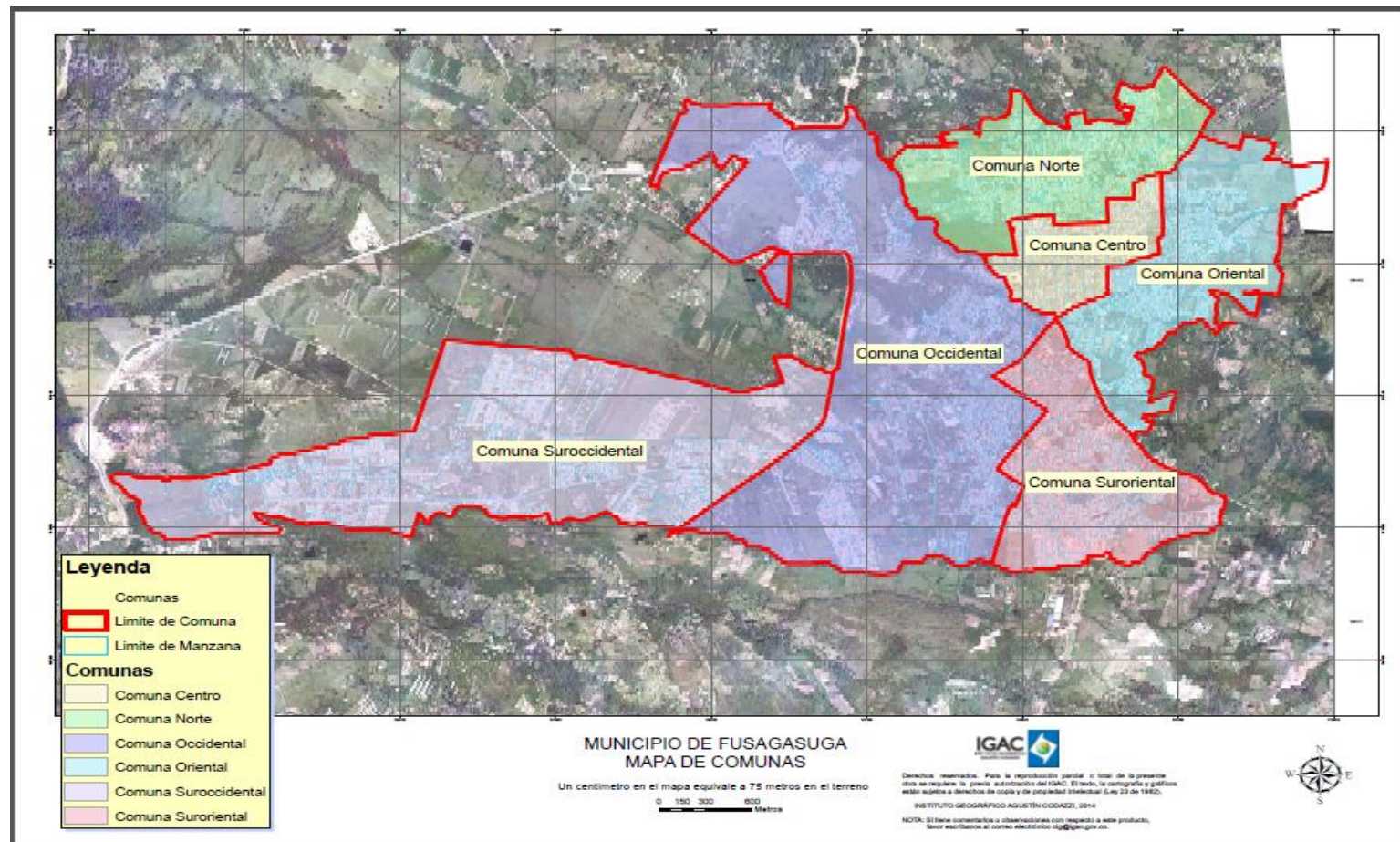


Figura 10. Plano de conjunto urbano Municipio de Fusagasugá con División político Administrativa de comunas Municipio Fusagasugá

Fuente: POT, Ortofotografía y Cartografía oficial IGAC

- I. Avalúo comercial.* Esta variable corresponde al valor de mercado del predio, pues no corresponde a transacción alguna y dada su magnitud fue normalizada. Este valor fue obtenido, como se mencionó anteriormente de los puntos de investigación y de la investigación indirecta (ofertas y avalúos) realizada en el marco del estudio de Zonas.

En la Figura 11, se muestra la espacialización de la muestra en el plano de conjunto de la zona urbana del Municipio.

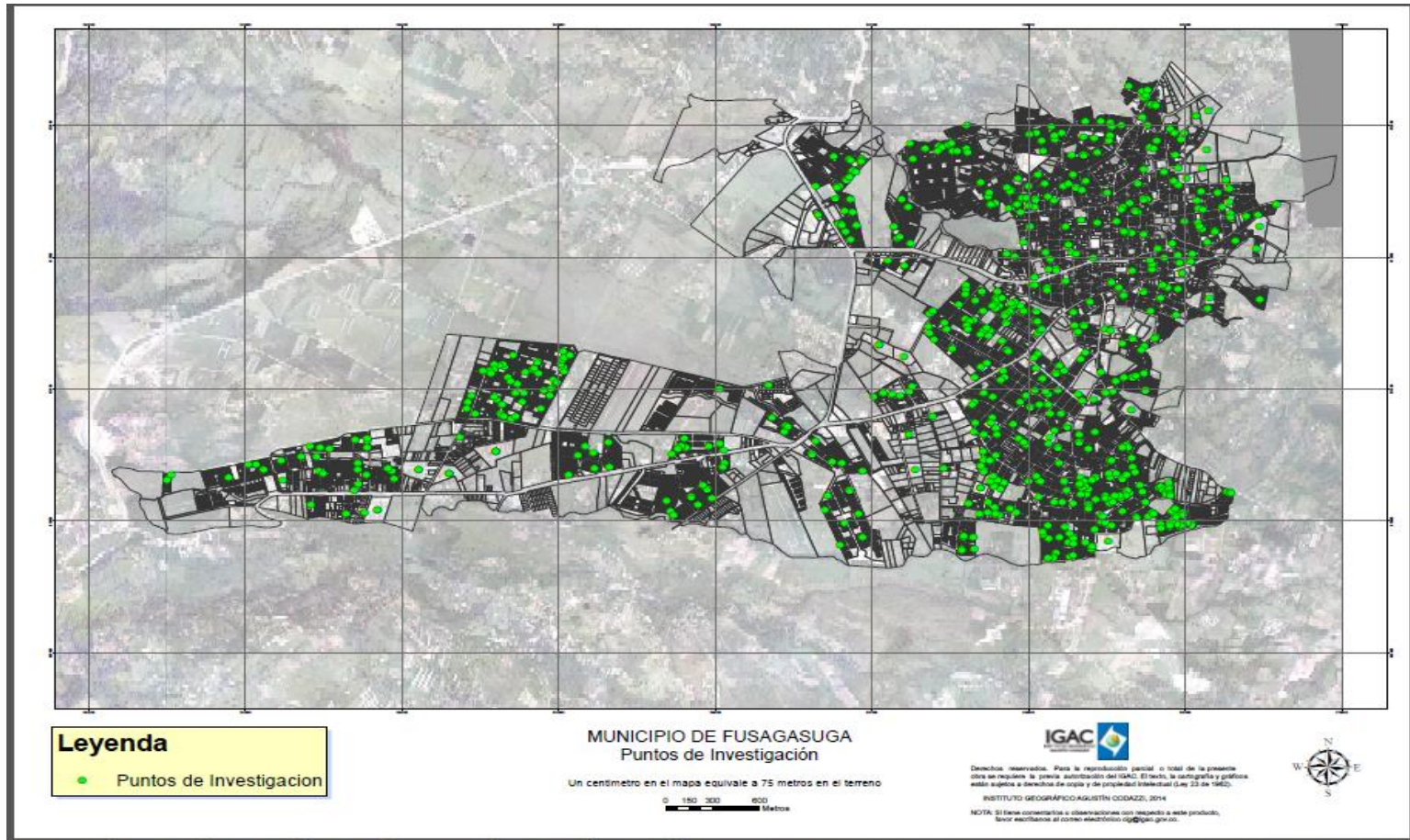


Figura 11. Plano urbano con distribución espacial de predios de la muestra, fuente IGAC

- **Selección del conjunto final de variables.**

Para la selección del conjunto final de variables, se empleó la técnica de Análisis de Componentes Principales ACI, para sintetizar la información o reducir la dimensión (número de variables), procurando perder la menor cantidad de información posible, para extraer algunos factores del conjunto de variables originales, buscando con esto mejorar los resultados finales de la red.

Es de precisar que a pesar de la ordenación teórica que, por razones de exposición, se ha realizado, hay que destacar que en la práctica las etapas señaladas anteriormente para el tratamiento de los datos no están perfectamente delimitadas en el tiempo, sino que se solapan unas con otras. Es el caso, por ejemplo en Redes Neuronales Artificiales, de la selección de variables de entrada, esta tarea no puede darse por finalizada hasta que, entrenada la red, se realice un análisis de sensibilidad del modelo respecto a cada input.

4.3.3 Estimación de la Red Neuronal

Para la estimación de la red se desarrollaron básicamente dos fases:

La primera de ellas denominada de **operación**, que es donde se da el proceso de aprendizaje o entrenamiento de la red a partir del conjunto de los patrones de entrenamiento (*avalúos*) introducidos. Para ejecutar esta fase se seleccionó el *entrenamiento supervisado* con el algoritmo de *backpropagation*.

En esta fase de entrenamiento se presentan a la red un conjunto de patrones de entrenamiento de forma iterativa. La red adaptará los pesos de las conexiones de forma que la respuesta a cada uno de los patrones tenga un error cada vez menor. Generalmente el entrenamiento concluye cuando se alcanza un error predeterminado

Antes de empezar el entrenamiento de una red, se debe proceder a la elección de los valores iniciales de los pesos asociados a las conexiones entre las neuronas. Éstos se iniciaron con valores pequeños. Una iteración termina cuando todos los casos disponibles se han considerado. El proceso finaliza cuando los cambios en los pesos y en el error cuadrático medio sean mínimos.

Luego se procedió a la segunda fase que corresponde a la ejecución, la cual se da una vez el sistema ha sido entrenado, los pesos de las conexiones permanecen fijos, la red generará una salida para cada entrada que se le presente. Este es el momento en que la red ya está preparada para procesar datos. En la

Figura 12 se muestra el diagrama de flujo del proceso para estimación de la red.

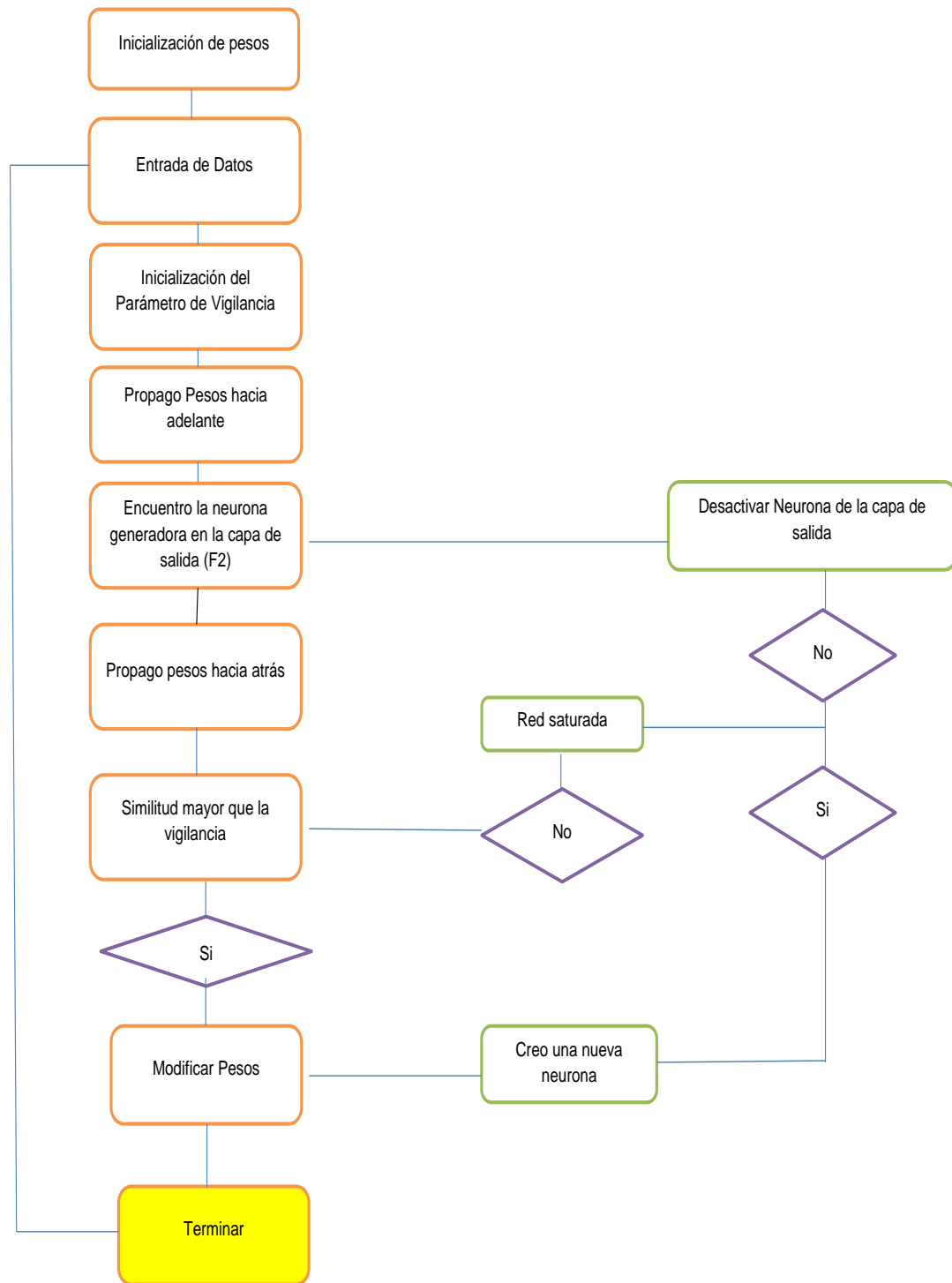


Figura 12. Flujograma de procesamiento de redes neuronales de la red Perceptrón Multicapa.

Para estimar la red neuronal se ha utilizado el módulo de Redes Neuronales del programa estadístico SPSS Versión 21, el cual ha demostrado ser muy eficaz en la estimación de la red neuronal, es sencillo de manejar y facilita de una manera importante estimación de las redes, adicionalmente el proceso de estimación de la red no lo se realiza en un paso único, pues es necesario evaluar constantemente las soluciones del modelo con el objetivo de mejorar su grado de ajuste. Con este software se analizaron cuáles eran las opciones que funcionaban mejor con este conjunto de datos, por lo tanto se generaron numerosas redes para comprender cuáles eran las opciones de programación disponibles en SPSS para la estimación de la mejor red.

Se debe tener en cuenta que como no existe una regla que dicte la elección de las opciones de la red, pues todo depende de los datos y de las interrelaciones entre ellos. Esta tarea, ha sido sin lugar a la duda, una de las que ha consumido el mayor tiempo en la investigación.

El SPSS ofrece la opción de la selección aleatoria de estos conjuntos de datos en diferentes porcentajes. También permite al usuario elegir un conjunto de prueba, además del conjunto de aprendizaje y de validación. Sin embargo, esta opción tiene un inconveniente: la imposibilidad de conocer qué datos se utilizaron en el conjunto de entrenamiento y qué datos se utilizaron en el conjunto de validación. El programa hace internamente esa separación de conjuntos, presentado al final, solamente el error relativo de cada conjunto generado. Para evitar esta situación, tenemos otra opción, donde el usuario crea una partición a priori de la muestra, aleatoria, con datos que estarán en el conjunto de entrenamiento y en el conjunto de validación. Con esta opción es posible calcular otras medidas de error, porque conocemos cuáles son los datos utilizados para validar la red.

Este trabajo se inició con muchas pruebas, en diferentes porcentajes, para los diferentes conjuntos. Esta tarea que consume principalmente tiempo, tenía dos propósitos. El primero ha sido entender con qué proporción en el conjunto de entrenamiento, la red funcionó mejor. Después de esto, el segundo objetivo fue percibir que padrón de resultados podríamos esperar en términos de la precisión de la red creada.

Se señala que sin el conjunto de test, tendríamos una evaluación optimista del error, sesgada. Con su uso, se tiene una estimación no sesgada del error que proporciona una verdadera generalización, desde que el conjunto de entrenamiento y de test sean aleatorios y puesto que el conjunto de test sea eficaz (siempre mayor que 70%) y cubriendo todo el espacio de datos. Sin embargo, todavía tenemos el pre-procesamiento de datos de entrada, que consiste en el cambio de la escala de datos. Esta tarea puede hacerse a través de la tipificación – conversión de una variable cuantitativa en una variable con media cero y desvío

típico de 1 – o a través de un proceso de normalización, en el que la variable cuantitativa se transforma en una variable en el rango de valores [0, 1] o [-1, 1]. Sobre la necesidad de hacer algún pre-procesamiento a los datos de entrada, hay que analizar qué tipo de red y que funciones de activación serán utilizadas.

En el caso de SPSS, existe la posibilidad de elegir para las variables en el campo opción cambio de escala de Covariables (variables cuantitativas) a través de las opciones *tipificados*, *normalización*, o *Corregida normalizada*.

La transformación de las variables de salida cuantitativas dependerá de la función de error utilizada, en concreto, de la suya dependencia para las diferentes escalas de medición de las diferentes variables de salida. Si hay una alta dependencia – como en el caso más usual con la suma del error cuadrático –, al tener varias salidas con diferentes intervalos, el proceso de aprendizaje da más peso a las variables con un rango más amplio de valores. En este caso concreto, sólo hay una variable de salida, el avalúo del predio.

Otra cuestión importante es la selección del número de capas ocultas de la red y el número de neuronas por capa. No hay reglas específicas para determinar el número óptimo de neuronas y el número de capas ocultas de un problema en concreto. Hay que tener en cuenta que un número excesivo de capas puede generar ruido, sin embargo, se puede lograr una mejor tolerancia a fallos (Bonilla & Puertas, 1997). Un gran número de capas ocultas lleva a una disminución en el margen de error de aprendizaje, pero no necesariamente conduce a un pequeño error en el test. Además, el número de neuronas ocultas intervienen en la efectividad del aprendizaje y generalización de la red (Hilera & Martínez, 1995). Esto significa que, en general, pocas neuronas ocultas hacen con que la red no distingue bien las características del problema, haciendo que en la fase de aprendizaje no se alcanzan los límites deseados, mientras que con demasiadas neuronas ocultas, aunque no conduzca a resultados erróneos, genera procesos de aprendizaje muy demorados. Una capa oculta con muchas neuronas requiere menos iteraciones para la red aprender, pero cada iteración se alargará en el tiempo debido a que tiene de calcular más peso. Cada capa no deberá tener el mismo número de neuronas como padrones, ya que esto favorece la memorización de estos padrones, pero se recomienda que, en general, se usen menos neuronas que variables (Pérez & Q., 2003). En definitiva, determinar el número de capas no es una tarea sencilla; normalmente se basa en la experiencia y se llevará en función del problema en estudio y la exactitud esperada de la red.

Teniendo en cuenta todas estas cuestiones previas, la red está diseñada y pronta para “trabajar”. Todo comenzó con la elección de un vector con las variables de entrada, y la correspondiente salida, en este caso el avalúo del predio. Esta información se propagará a

la última capa de la red, dando como resultado un vector de salida. Este proceso se lleva a cabo a través de la función de activación, que produce la transformación de las entradas de cada neurona de una misma capa, en salidas de neuronas de esta capa a la siguiente y que constituyen las entradas de la última capa. Una vez que haya completado este proceso con la n vector de entrada, la red calcula el valor de salida global. A través de la diferencia con la salida deseada, se obtiene el error global por unidad de salida. A continuación, se determina la contribución relativa de las neuronas en ese error y al través de un algoritmo de aprendizaje, los errores serán modificados, repetidamente, para cada vector del conjunto de padrones de aprendizaje, hasta que el error global obtenido sea mínimo. A través del desarrollo de este comportamiento, se consigue el auto-ajuste de la red, que permite que al presentar una nueva entrada, consiga proporcionar una salida apropiada – esto es lo que se llama la capacidad de generalización de la red

El módulo de redes neuronales del SPSS, tiene una opción para seleccionar automáticamente la arquitectura. Esta opción permite seleccionar la “*mejor*” estructura de la red de forma automática.

A continuación se presenta un flujograma de proceso en el cual se resumen las diferentes etapas que se deben surtir para obtener el modelo final de la Red Neuronal Artificial con la cual se espera estimar el avalúo de un predio, ver Figura 13.

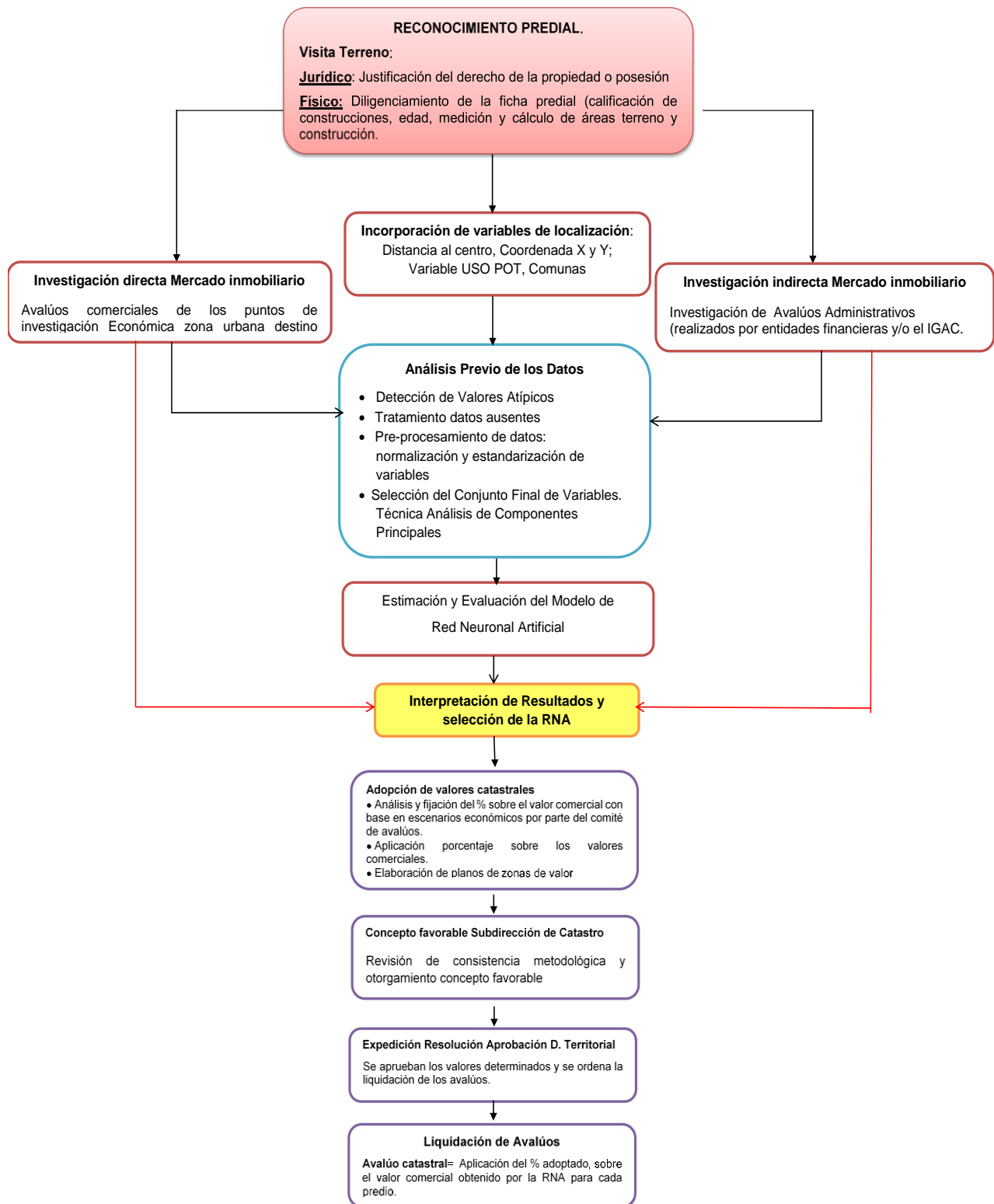


Figura 13. Flujograma de la metodología para calcular el avalúo catastral de un predio utilizando RNA

5. Análisis de Resultados

5.1 Análisis Exploratorio de las Variables

Una vez realizado el proceso de depuración de la información, eliminando datos atípicos y faltantes se procedió a realizar la estadística descriptiva para las variables evaluadas en el modelo:

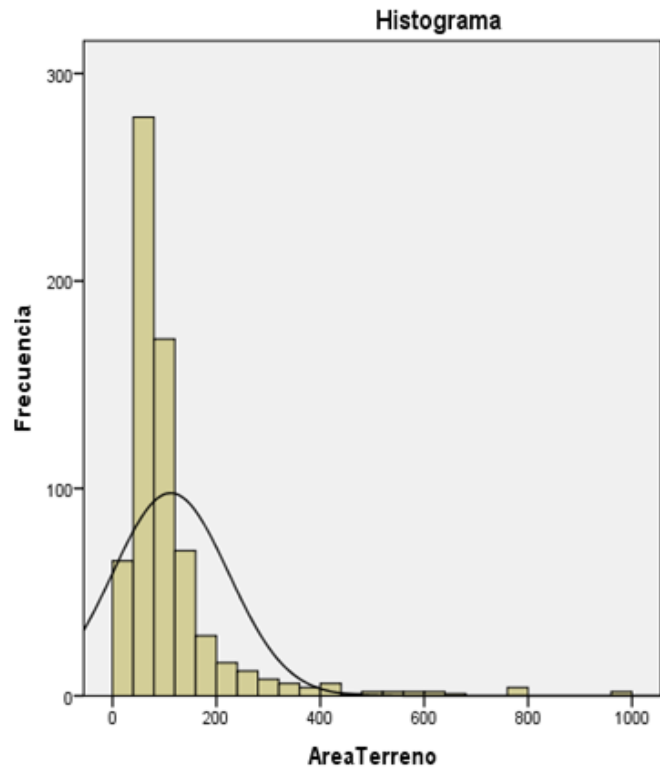
5.1.1 Variables Físicas del Predio

- **Área Terreno**

Estadísticos		
AreaTerreno		
N	Válidos	685
	Perdidos	0
Media		111,92
Error típ. de la media		4,265
Mediana		78,67 ^a
Moda		72
Desv. típ.		111,463
Varianza		12423,932
Asimetría		3,884
Error típ. de asimetría		,094
Curtosis		19,426
Error típ. de curtosis		,187
Rango		956
Mínimo		5
Máximo		961
Suma		76442
Percentiles	25	63,90 ^b
	33	71,62
	50	78,67
	66	101,47
	75	118,75

a. Calculado a partir de los datos agrupados.

b. Los percentiles se calcularán a partir de los



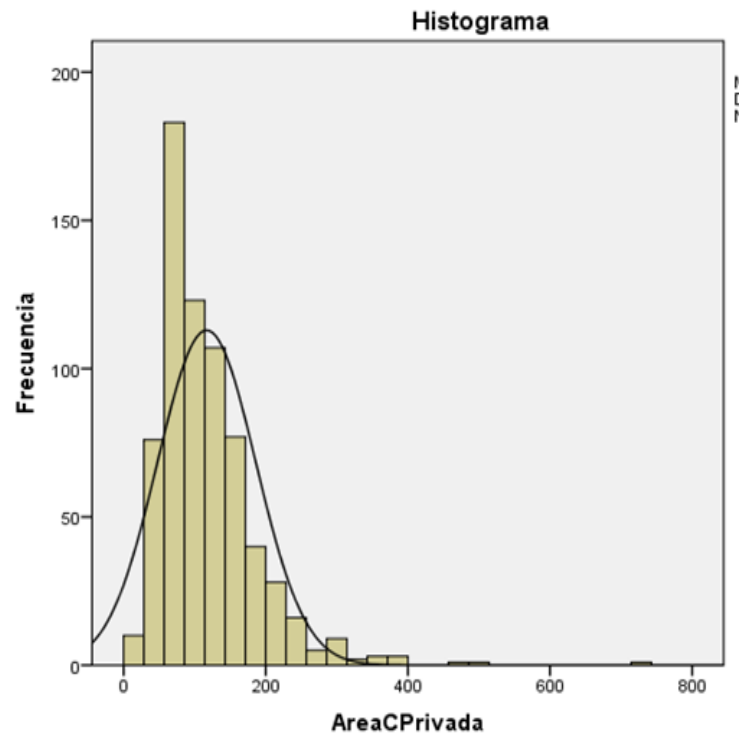
Rango Área T	Predios
5- 60	164
61-90	239
91-120	122
121-200	92
201- 300	35
Mayor a 300	33
Total	685

En general se observa que el 59% del total de los predios tienen un área de terreno inferior a 90 M² el predio más frecuente en términos de superficie útil de 72 M², siendo la media general 111 M².

▪ **Área Construida o Privada (Área C Privada)**

Estadísticos

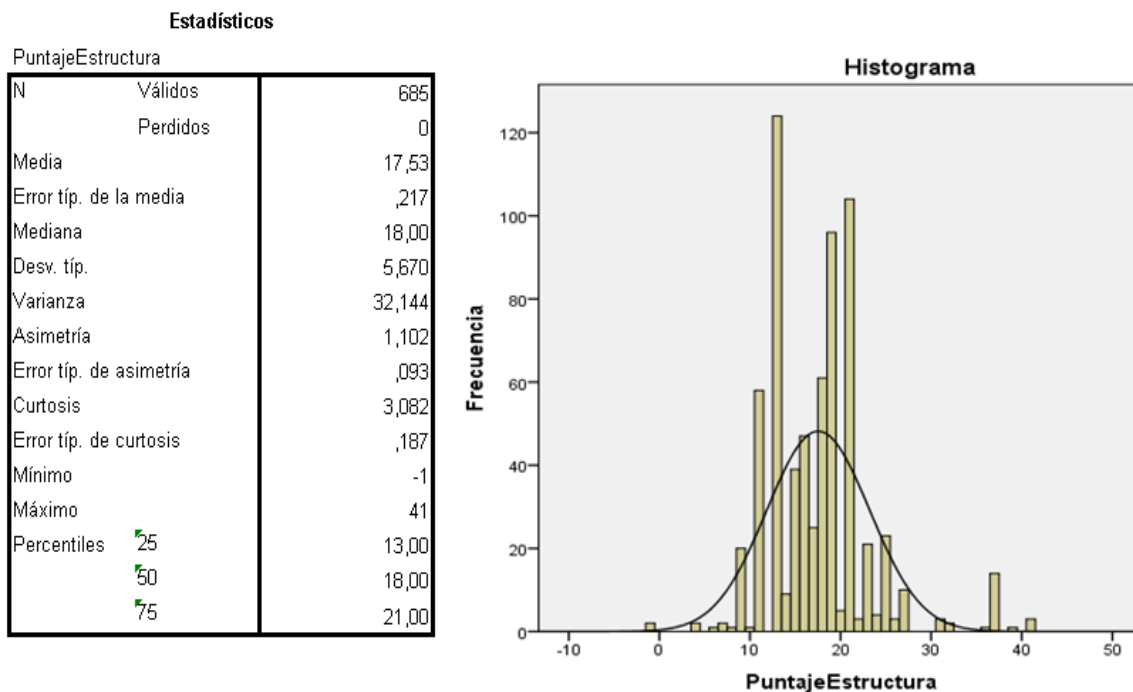
AreaCPrivada		
N	Válidos	685
	Perdidos	0
Media		116,84
Error típ. de la media		2,641
Mediana		104,00
Desv. típ.		69,117
Varianza		4777,156
Asimetría		2,382
Error típ. de asimetría		,093
Curtosis		12,177
Error típ. de curtosis		,187
Mínimo		5
Máximo		736
Percentiles	25	67,50
	50	104,00
	75	146,00



Rango Área C	Predios
5- 60	119
61-90	177
91-120	122
121-200	198
201- 300	55
Mayor a 300	14
Total	685

En general se observa que hay una distribución proporcional en las áreas, pues 119 son menores a 60 M² y menores a 200 M² tenemos 198 predios, estando la media en 116,84. También encontramos que la mayor parte de los predios tiene un área inferior a 200 M².

▪ Puntaje de la Estructura



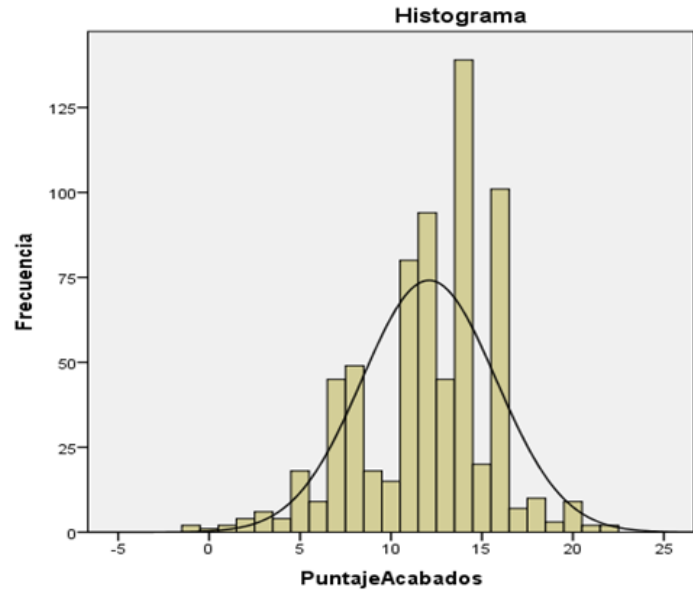
En general se observa que la media del puntaje de la estructura se encuentre en 17,53 puntos y la mayor frecuencia se da para 13 y 21 puntos, con 124 y 104 predios, respectivamente.

▪ **Puntaje Acabados**

Estadísticos

PuntajeAcabados

N	Válidos	685
	Perdidos	0
Media		12,09
Error típ. de la media		,141
Mediana		12,00
Desv. típ.		3,686
Varianza		13,588
Asimetría		-,556
Error típ. de asimetría		,093
Curtosis		,429
Error típ. de curtosis		,187
Mínimo		-1
Máximo		22
Percentiles	25	10,00
	50	12,00
	75	14,00



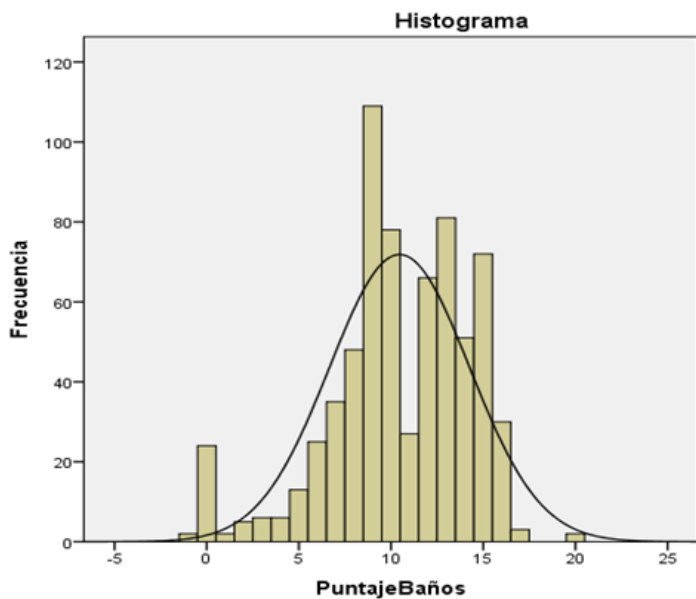
En general se observa que la media del puntaje de acabados se encuentre en 12 puntos y la mayor frecuencia se da para 14 y 16 puntos, con 139 y 101 predios, respectivamente.

▪ **Puntaje de Baños**

Estadísticos

PuntajeBaños

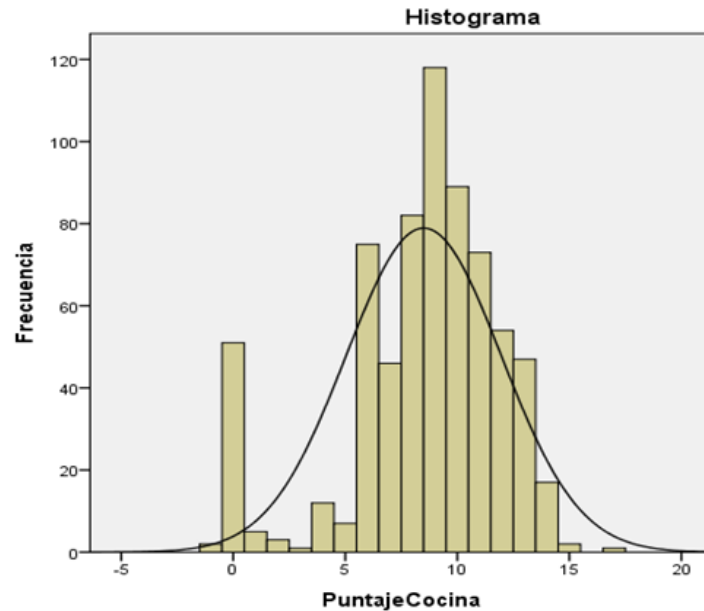
N	Válidos	685
	Perdidos	0
Media		10,46
Error típ. de la media		,145
Mediana		10,00
Desv. típ.		3,803
Varianza		14,465
Asimetría		-,742
Error típ. de asimetría		,093
Curtosis		,607
Error típ. de curtosis		,187
Mínimo		-1
Máximo		20
Percentiles	25	9,00
	50	10,00
	75	13,00



Respecto de esta variable se puede decir que el promedio es de 10,46 puntos, la mayor frecuencia es para 9 puntos, con 109 predios, seguida de 15 puntos con 72 predios.

▪ Puntaje de Cocina

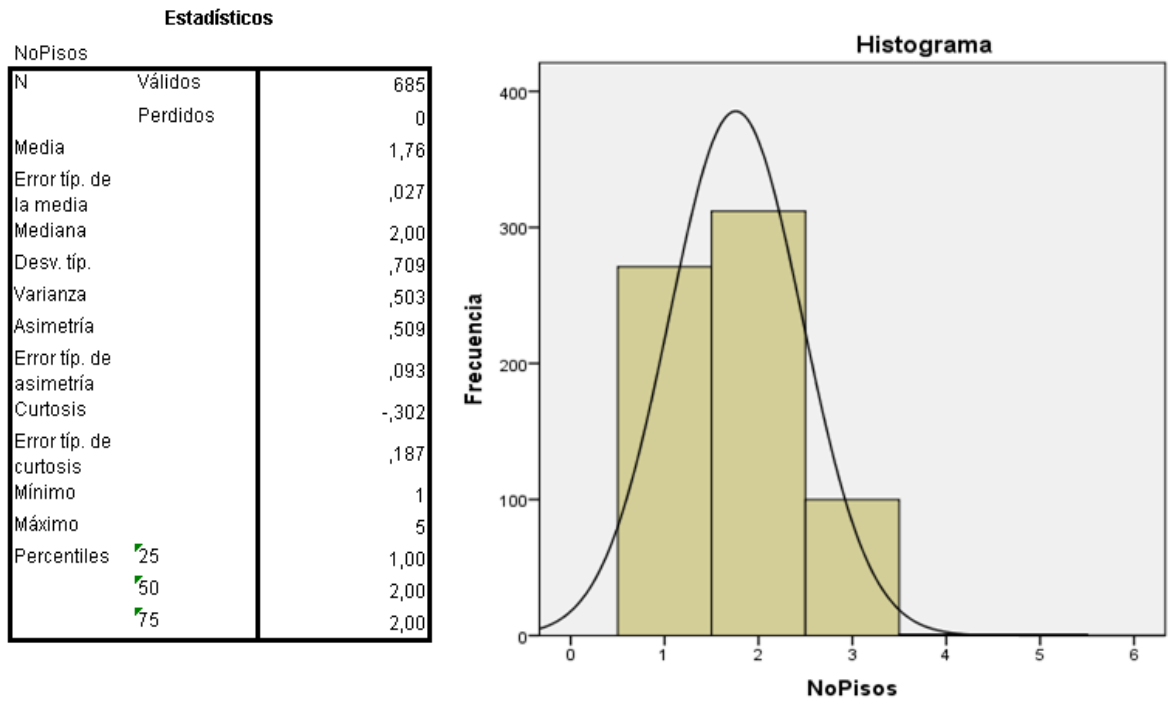
Estadísticos		
PuntajeCocina		
N	Válidos	685
	Perdidos	0
Media		8,50
Error típ. de la media		,132
Mediana		9,00
Desv. típ.		3,461
Varianza		11,981
Asimetría		-,948
Error típ. de asimetría		,093
Curtosis		,775
Error típ. de curtosis		,187
Mínimo		-1
Máximo		17
Percentiles	25	7,00
	50	9,00
	75	11,00



El comportamiento de esta variable es un poco más de homogéneo, pues como puede observarse en la gráfica están distribuidos entre los 6 y los 14 puntos, siendo la más alta frecuencia para 9 puntos con 118 predios, el cual es cercana a la media que es de 8,5 puntos.

▪ Número de Pisos

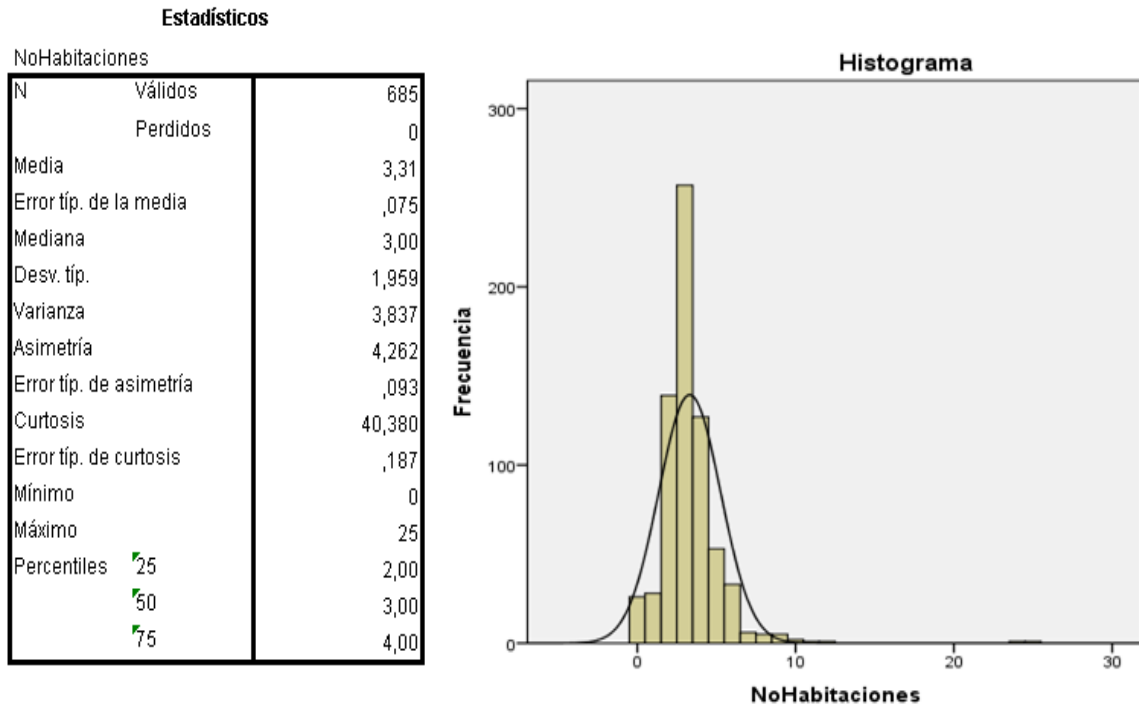
NoPisos				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos 1	271	39,6	39,6	39,6
2	312	45,5	45,5	85,1
3	100	14,6	14,6	99,7
4	1	,1	,1	99,9
5	1	,1	,1	100,0
Total	685	100,0	100,0	



En el anterior cuadro puede observarse que en el municipio predominan los predios de dos pisos, existiendo en la muestra 312 de ellos, seguidos por un piso con 271 predios. Nótese que en muestra solo encontramos una casa para 4 y 5 pisos.

- **Número de Habitaciones (Nº Habitaciones)**

		NoHabitaciones			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	0	26	3,8	3,8	3,8
	1	28	4,1	4,1	7,9
	2	139	20,3	20,3	28,2
	3	257	37,5	37,5	65,7
	4	127	18,5	18,5	84,2
	5	53	7,7	7,7	92,0
	6	33	4,8	4,8	96,8
	7	6	,9	,9	97,7
	8	5	,7	,7	98,4
	9	5	,7	,7	99,1
	10	2	,3	,3	99,4
	11	1	,1	,1	99,6
	12	1	,1	,1	99,7
	24	1	,1	,1	99,9
	25	1	,1	,1	100,0
	Total	685	100,0	100,0	



La mayor parte de los predios tiene entre 2 y 4 habitaciones, existiendo solo un caso para predios con 10, 11,12, 24 y 25 habitaciones.

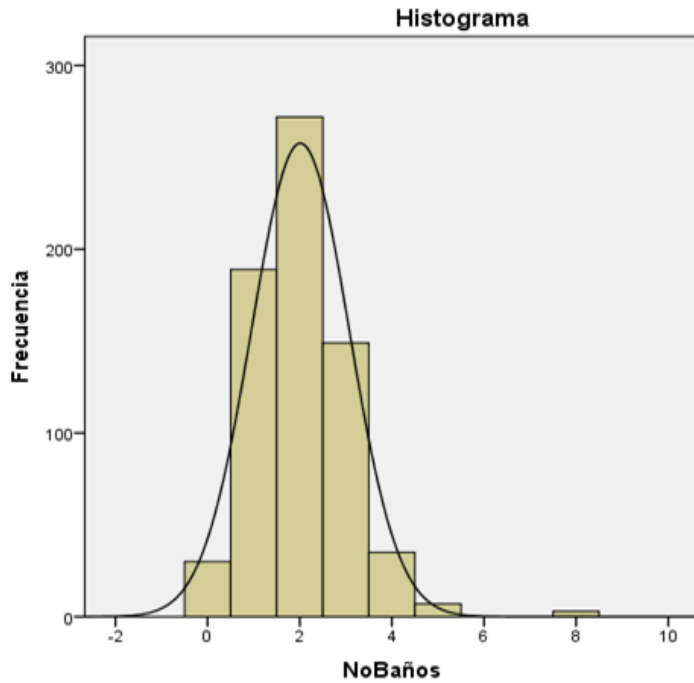
▪ Número de Baños

		NoBaños			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	0	30	4,4	4,4	4,4
	1	189	27,6	27,6	32,0
	2	272	39,7	39,7	71,7
	3	149	21,8	21,8	93,4
	4	35	5,1	5,1	98,5
	5	7	1,0	1,0	99,6
	8	3	,4	,4	100,0
	Total	685	100,0	100,0	

Estadísticos

NoBaños

N	Válidos	685
	Perdidos	0
Media		2,01
Error típ. de la media		,041
Mediana		2,00
Desv. típ.		1,060
Varianza		1,124
Asimetría		,992
Error típ. de asimetría		,093
Curtosis		3,745
Error típ. de curtosis		,187
Mínimo		0
Máximo		8
Percentiles	25	1,00
	50	2,00
	75	3,00



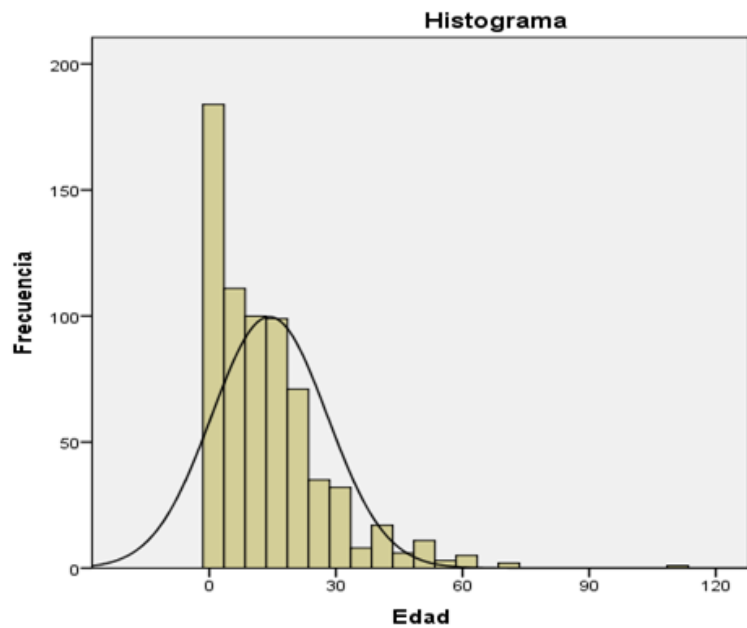
En el anterior cuadro se puede observar que el 71,6% de los predios tienen entre 1 y 2 baños, aunque también hay una buena cantidad correspondiente a 149 predios que poseen tres baños.

▪ **Edad del Predio**

Estadísticos

Edad

N	Válidos	685
	Perdidos	0
Media		14,26
Error típ. de la media		,523
Mediana		11,00
Desv. típ.		13,701
Varianza		187,709
Asimetría		1,790
Error típ. de asimetría		,093
Curtosis		5,223
Error típ. de curtosis		,187
Mínimo		1
Máximo		112
Percentiles	25	3,00
	50	11,00
	75	20,00



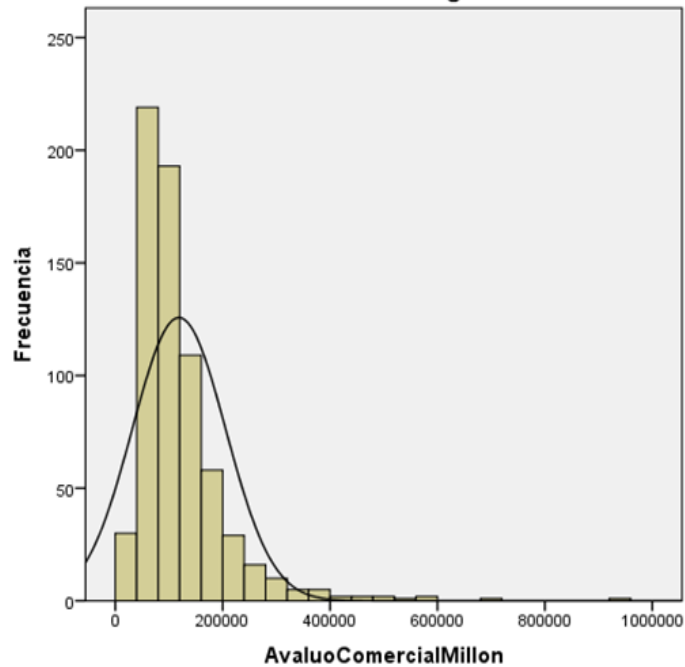
De las anteriores cifras se puede deducir que las construcciones del Municipio de Fusagasugá, son relativamente recientes, pues el promedio de edad esta apenas en 14 años, siendo la más antigua con 112 años y la más reciente con 1 año.

Variable Dependiente Avalúo Comercial

Estadísticos

AvaluoComercialMillon		
N	Válidos	685
	Perdidos	0
Media		118518,10
Error típ. de la media		3321,265
Mediana		98140,64
Desv. típ.		86925,821
Varianza		7556098362,497
Asimetría		3,359
Error típ. de asimetría		,093
Curtosis		19,285
Error típ. de curtosis		,187
Mínimo		5189
Máximo		950000
Percentiles	25	66047,00
	50	98140,64
	75	140000,00

Histograma



Rango Avalúo Comercial (Millones de \$)	Nº Predios
0-70	212
71-100	152
101-150	177
151_200	73
201-300	47
Mayor 300	24
Total	685

De la muestra avalúos que se emplearán en la elaboración del modelo, podemos decir que el máximo valor es de \$950.000.000 y el mínimo de \$5.189.000, con una media de

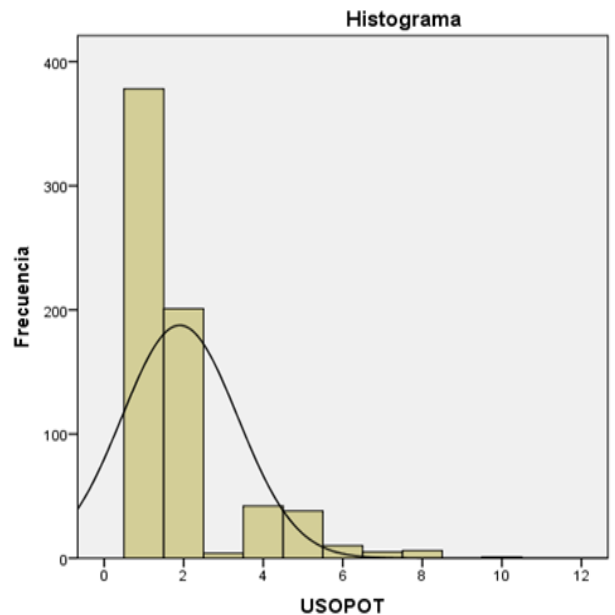
\$118.518.000 millones, siendo más frecuente avalúos por debajo de los \$70 millones, 212 predios.

Respecto a esta información, es importante precisar que la recolección y análisis de la investigación mercado es la labor más compleja y dispendiosa en la elaboración los Estudios de Zonas Homogéneas que se realiza en el marco de los procesos de Actualización de la Formación Catastral, pues en la mayoría de los casos estos datos se capturan e investigan directamente en campo, pues las inmobiliarias que también poseen esta información, son reacias a suministrar esta información para fines catastrales.

5.1.2 Variables de Localización

- **Uso del suelo Plan de Ordenamiento Territorial (USOPOT)**

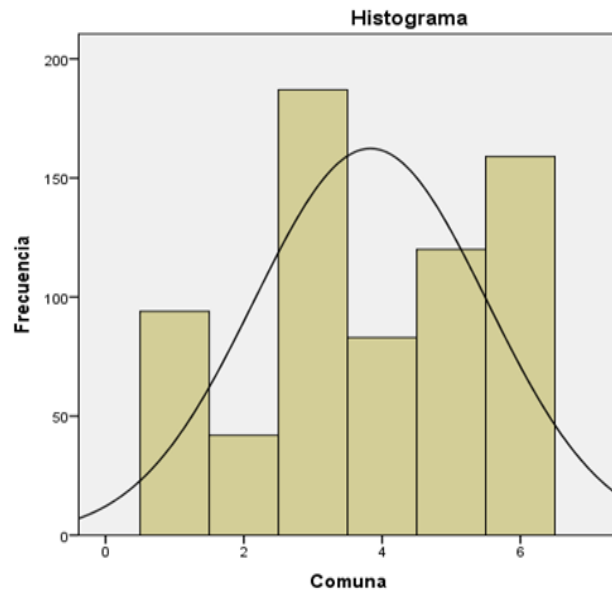
USOPOT				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos 1	378	55,2	55,2	55,2
2	201	29,3	29,3	84,5
3	4	,6	,6	85,1
4	42	6,1	6,1	91,2
5	38	5,5	5,5	96,8
6	10	1,5	1,5	98,2
7	5	,7	,7	99,0
8	6	,9	,9	99,9
10	1	,1	,1	100,0
Total	685	100,0	100,0	



Vemos que el 55,2% de la totalidad de los predios tienen en el POT un uso permitido 1, seguido del uso 2 con un 29,3%, los cuales corresponden a los Usos Área de Desarrollo Habitacional de Densidad Alta y Media, respectivamente. Los demás usos tienen una participación menor dentro de la muestra.

Variable Comuna

Comuna				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos 1	94	13,7	13,7	13,7
2	42	6,1	6,1	19,9
3	187	27,3	27,3	47,2
4	83	12,1	12,1	59,3
5	120	17,5	17,5	76,8
6	159	23,2	23,2	100,0
Total	685	100,0	100,0	



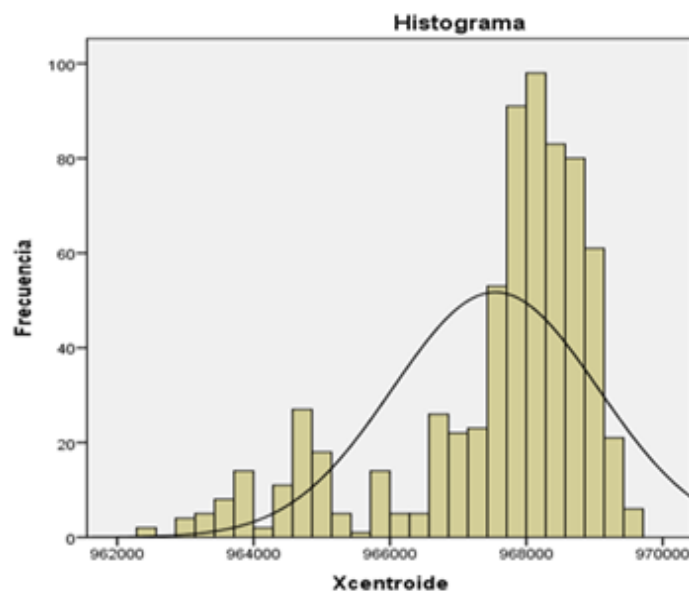
Como se puede observar la muestra tiene representación proporcional en cada una de las comunas, siendo la de mayor cantidad de predios la Comuna 3 (Comuna Occidental) con 187 predios, seguida de la Comuna 6 (Comuna Suroriental) con 159 predios.

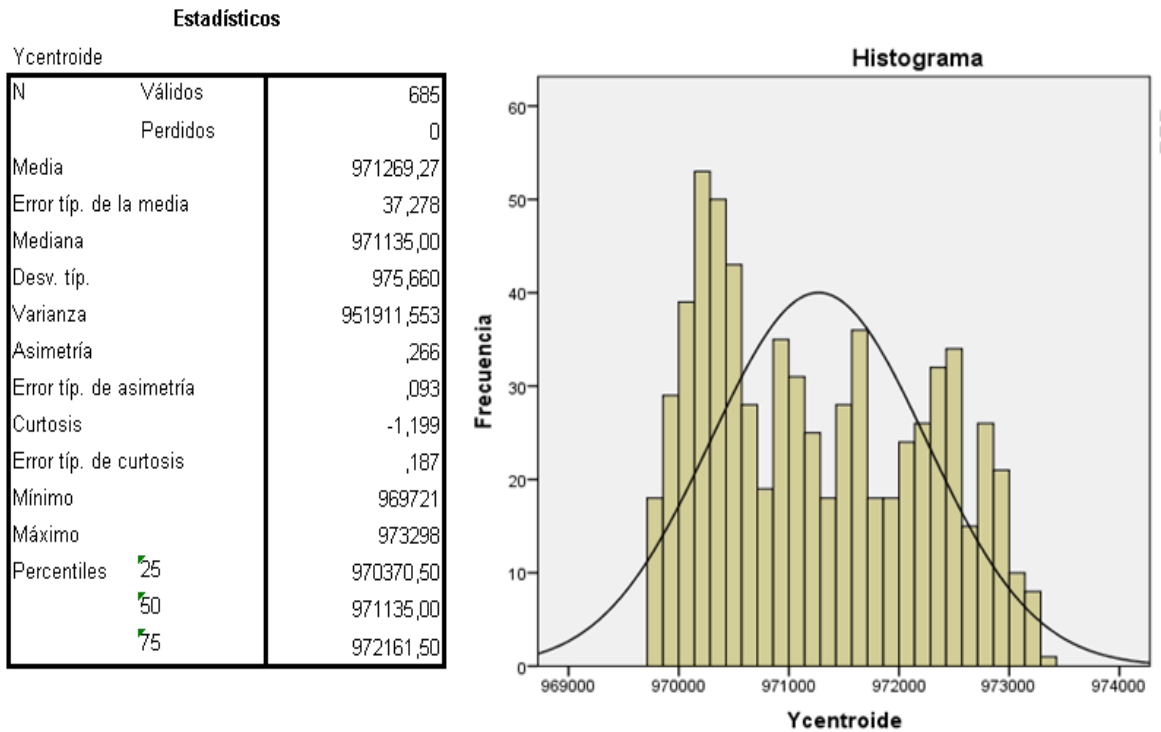
Coordenada X y Y (X Centroide – Y Centroide)

Coordenada X.

Estadísticos

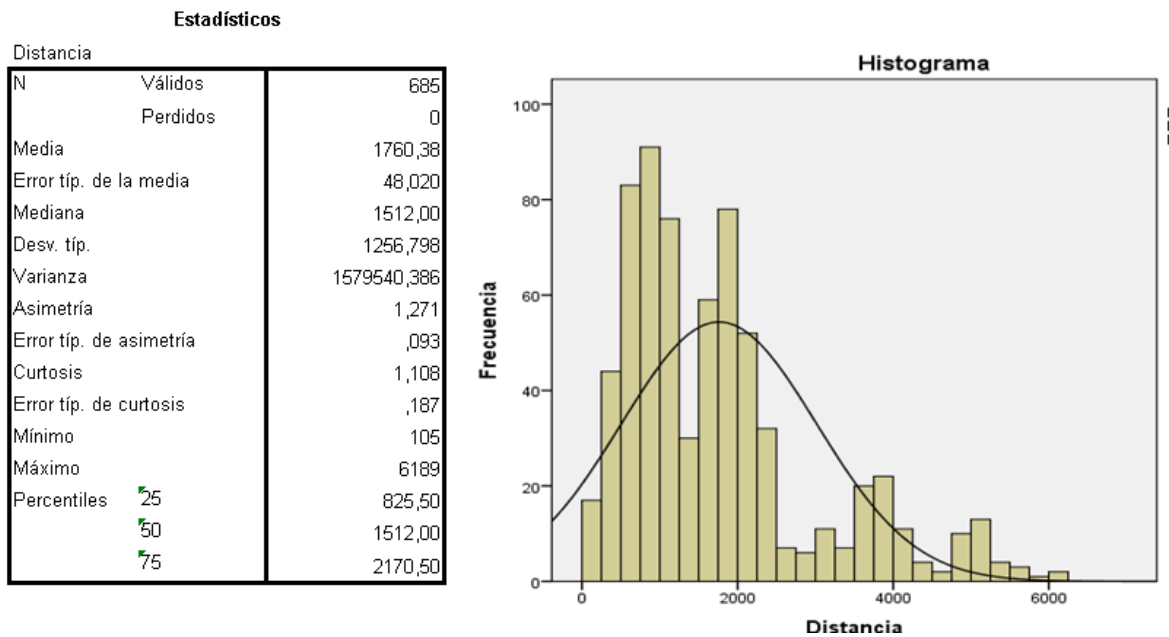
Xcentroide		
N	Válidos	685
	Perdidos	0
Media		967546,91
Error típ. de la media		57,671
Mediana		968011,00
Desv. típ.		1509,406
Varianza		2278307,684
Asimetría		-1,385
Error típ. de asimetría		,093
Curtosis		1,066
Error típ. de curtosis		,187
Mínimo		962496
Máximo		969579
Percentiles	25	967177,50
	50	968011,00
	75	968551,50





Las variables Coordenadas X y Y muestran que existe una adecuada distribución de la muestra por toda la zona urbana, ver Figura 11.

▪ **Distancia al Centro de la ciudad (Distancia CCN)**



La anterior gráfica muestra predios que se encuentran ubicados a diferentes distancias con relación al centro de la ciudad, siendo la distancia mínima al Centro de la ciudad (Plaza Central) de 105 Mt y la máxima 6.169 Mt, estando la distancia media en 1.760 Mt.

5.2 Análisis de Componentes Principales

Con el empleo de esta técnica se buscó reducir las dimensiones del conjunto de variables sin pérdida excesiva de información. En la Tabla 7 se presenta la estadística descriptiva básica de la prueba de análisis de componentes principales y la prueba del Test de esfericidad de Barlett, con el cual comprobamos que las correlaciones entre las variables son distintas de cero de modo significativo, se comprueba si el determinante de la matriz es distinto de uno, es decir, si la matriz de correlaciones es distinta de la matriz unidad. Por otra parte, la medida de la adecuación muestral de *Kaiser Meyer Olkin* contrasta si las correlaciones parciales entre las variables son suficientemente pequeñas, para nuestro caso es de 0.679, lo que indicaría adecuación mediana de los datos a un modelo de análisis factorial.

Tabla 7. Estadísticos Descriptivos Análisis de Componentes Principales

Estadísticos descriptivos				KMO y prueba de Bartlett		
VARIABLES	Media	Desviación típica	N del análisis	Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,679
XCOORD	967546,907	1509,4064	685	Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	6395,886
YCOORD	971269,273	975,6595	685		gl	91
DISTAN	1760,38	1256,798	685	Sig.	0,000	
ARTERR	111,60	111,455	685			
ARCONST	116,84	69,117	685			
PTOTAL	48,921	12,5411	685			
PESTR	17,53	5,670	685			
PACAB	12,09	3,686	685			
PBAÑO	10,46	3,803	685			
PCOCN	8,50	3,461	685			
PISOS	1,76	,709	685			
HABIT	3,31	1,959	685			
BAÑOS	2,02	1,062	685			
EDAD	14,263	13,7007	685			

En la Tabla 8 se muestran los resultados de la Matriz de las correlaciones con la significación de cada componente. Para que se pueda realizar el ACP, es necesario que las variables presenten factores comunes. Es decir, que estén muy correlacionadas entre sí. Los coeficientes de la matriz de las correlaciones deben de ser grandes en valor absoluto. Se obtienen los componentes principales a partir de la matriz de correlaciones para emplear las mismas escalas en todas las variables.

Tabla 8. Matriz de Correlaciones

Coorelación	XCOORD	YCOORD	DISTAN	ARTERR	ARCONST	PTOTAL	PESTR	PACAB	PBAÑO	PCOCN	PISOS	HABIT	BAÑOS	EDAD
XCOORD	1,000	,266	-,861	,055	,027	-,138	-,080	-,103	-,109	-,100	-,029	,044	-,097	,241
YCOORD	,266	1,000	-,580	,116	,032	-,122	-,064	-,084	-,120	-,139	-,132	,009	-,107	,231
DISTAN	-,861	-,580	1,000	-,139	-,105	,091	-,002	,040	,118	,140	,059	-,063	,053	-,352
ARTERR	,055	,116	-,139	1,000	,411	-,190	-,183	-,168	-,140	-,148	-,045	,270	,116	,425
ARCONST	,027	,032	-,105	,411	1,000	,028	-,082	,005	,027	,046	,383	,494	,443	,193
PTOTAL	-,138	-,122	,091	-,190	,028	1,000	,654	,832	,734	,657	,194	,034	,361	-,413
PESTR	-,080	-,064	-,002	-,183	-,082	,654	1,000	,638	,195	,068	,072	-,128	,094	-,278
PACAB	-,103	-,084	,040	-,168	,005	,832	,638	1,000	,592	,514	,182	,085	,348	-,333
PBAÑO	-,109	-,120	,118	-,140	,027	,734	,195	,592	1,000	,854	,210	,159	,440	-,293
PCOCN	-,100	-,139	,140	-,148	,046	,657	,068	,514	,854	1,000	,227	,182	,435	-,311
PISOS	-,029	-,132	,059	-,045	,383	,194	,072	,182	,210	,227	1,000	,384	,524	-,192
HABIT	,044	,009	-,063	,270	,494	,034	-,128	,085	,159	,182	,384	1,000	,682	,210
BAÑOS	-,097	-,107	,053	,116	,443	,361	,094	,348	,440	,435	,524	,682	1,000	-,042
EDAD	,241	,231	-,352	,425	,193	-,413	-,278	-,333	-,293	-,311	-,192	,210	-,042	1,000

La **Tabla 9** presenta las comunalidades obtenidas para cada una de las variables. Al considerar tan sólo un número reducido de factores entre todas las variables observadas, la varianza total no queda totalmente explicada. Por tanto, sus valores oscilarán entre cero y uno, es decir, entre la posibilidad de que los factores comunes no expliquen nada de la variabilidad de una variable o que por el contrario ésta quede totalmente explicada por los factores comunes.

Una vez realizada la extracción, se puede observar que variables tales como DISTAN, PTOTAL, PCOCN Y PBAÑO registran una comunalidad de 0.961, 0.927, 0.912 y 0.904, respectivamente, es decir que son los que explican en mayor proporción la varianza según su participación en los factores o componentes resultantes en el análisis, mientras que variables como YCOORD Y EDAD participan en menor medida con un 0,471 y 0.645, respectivamente en la explicación de dicha varianza.

Tabla 9. Comunalidades de Análisis de Componentes Principales

Variable	Inicial	Extracción
XCOORD	1,000	,813
YCOORD	1,000	,471
DISTAN	1,000	,961
ARTERR	1,000	,733
ARCONST	1,000	,659
PTOTAL	1,000	,927
PESTR	1,000	,913
PACAB	1,000	,840
PBAÑO	1,000	,904
PCOCN	1,000	,912
PISOS	1,000	,770
HABIT	1,000	,708
BAÑOS	1,000	,779
EDAD	1,000	,645

Método de extracción: Análisis de Componentes principales.

Considerando la Tabla 10 podemos decidir con cuantos componentes o factores nos vamos a quedar. Existen reglas para saber el número más adecuado a conservar, por ejemplo, la que se conoce como Criterio de Kaiser que indica que hay que conservar los componentes principales cuyos valores propios son mayores que la unidad, aunque el criterio más utilizado es el de observar el porcentaje de varianza total explicada por cada componente o factor, y cuando éste llega a un porcentaje acumulado considerado alto normalmente cerca del ochenta por ciento, significa que el número de factores es suficiente.

En la Tabla 10 se puede observar que a partir de la componente número cinco el autovalor comienza a ser inferior a la unidad aunque en esta componente todavía mantiene un valor alto, y además el porcentaje de la varianza explicada acumulada asciende al 78,84%, por lo que podemos considerar que éste puede ser un valor lo suficientemente alto para estimar que cinco es un número de factores suficiente.

Los valores propios de cada componente que se muestran en la Tabla 11 en donde nos indican la cantidad de varianza, la cantidad de información que tiene cada componente. Como podemos ver en esta tabla la primera componente tiene un 29,5928% de información, la segunda un 19.117%, la tercera 14,549%, la cuarta 7.510% y en la quinta componente ya tenemos un 78.834% acumulado de la varianza. Por lo tanto, con las cinco primera componentes perdemos un 21.16 % de información únicamente

Tabla 10. Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	4,143	29,592	29,592	4,143	29,592	29,592
2	2,676	19,117	48,709	2,676	19,117	48,709
3	2,037	14,549	63,258	2,037	14,549	63,258
4	1,129	8,066	71,324	1,129	8,066	71,324
5	1,051	7,510	78,834	1,051	7,510	78,834
6	,735	5,250	84,084			
7	,629	4,495	88,579			

Método de extracción: Análisis de Componentes principales.

Finalmente optamos por extraer cinco componentes, A continuación pasamos a renombrar las componentes en función de las variables iniciales que incorporan, ver en la Tabla 11.

Tabla 11. Tabla Matriz de Componentes ACP

	Componente				
	PUNTAJE	ESPACIO	LOCALIZACION	ESTRUCTURA	ALTURA
XCOORD	-,307	,349	,695	-,204	-,269
YCOORD	-,300	,257	,550	-,021	,112
DISTAN	,320	-,442	-,803	,057	,129
ARTERR	-,253	,542	-,147	,278	,526
ARCONST	,100	,730	-,230	,251	-,030
PTOTAL	,884	-,075	,311	,151	,141
PESTR	,505	-,236	,375	,676	-,072
PACAB	,803	-,043	,336	,263	,105
PBAÑO	,815	,080	,149	-,376	,264
PCOCN	,774	,099	,078	-,495	,226
PISOS	,416	,429	-,226	,030	-,601
HABIT	,221	,773	-,242	-,017	-,057
BAÑOS	,581	,625	-,202	-,011	-,097
EDAD	-,526	,465	,057	,076	,378

Método de extracción: Análisis de componentes principales.

a. 5 componentes extraídos

Componente 1. Esta componente engloba al conjunto de atributos que se concentran en el puntaje de la construcción, siendo las más destacadas las variables: PTOTAL, PACAB, PBAÑO, PCOC y EDAD, teniendo en cuenta que son en buena parte las responsables de definir la calidad de la construcción que se encuentra levantada en el predio, esta

componente se denominará **PUNTAJE**, la cual explica por sí sola el 29,52|%, de la varianza total, ocupando con esta cifra el primer lugar frente al resto de componentes que vamos a presentar.

Componente 2. Se destacan en ella cuatro variables que se pueden considerar como indicadoras de la calidad del espacio de un predio. Siendo las variables más destacadas: ATERR, ARCONST, HABIT y BAÑOS. Llamaremos a esta componente **ESPACIO**, la cual explica un 19.117% de la varianza total.

Componente 3. En esta componente resaltan las variables XCOORD, YCOORD y DISTANCIA, registrando esta última un valor de varianza alto, pero inversamente proporcional (-,803). Como puede observarse las tres variables revelan aspectos relacionados con la ubicación del predio, por eso la denominamos **LOCALIZACIÓN**, y explica un total 14.549% de la varianza total.

Componente 4. De esta componente solo se destaca la variable ESTRUC, la cual hace referencia exclusivamente al puntaje de la estructura de la construcción, variable que desempeña un papel muy importante en cualquier edificación, pues es el esqueleto que la sostiene y gracias a ella se puede levantar y detener, es por eso que la denominamos **ESTRUCTURA**, y explica el 8.066% de la varianza.

Componente 5. De esta componente solo sobresalió la variable PISOS, con una varianza inversamente proporcional de -, 601, es decir que a mayor número de pisos menor valor del predio, lo cual se ratifica con el hecho de que en el municipio de Fusagasugá se encontró que los ciudadanos no frecuentaban centros comerciales u oficinas de más de dos pisos, de la misma manera se observó que el predominio de la propiedad horizontal era en casas, no en apartamentos en altura. Por eso denominaremos esta componente **PISOS**, la cual explicó el 7.510% de la varianza.

Teniendo en cuenta el análisis anteriormente realizado al Análisis de Componentes Principales, en donde se evidencia desde los tres primeros componentes que prácticamente todas las variables cuantitativas tienen una participación importante en la explicación del avalúo, por lo tanto se decidió emplear la totalidad de ellas para la estimación de la red neuronal.

5.3 Estimación de la Red Neuronal

El objetivo principal de este trabajo es el desarrollar una metodología basada en Redes Neuronales Artificiales para calcular el avalúo catastral de un predio utilizando las variables

que el IGAC tiene disponibles en el Sistema de Información Catastral. Anteriormente ya se abordó aspectos como la selección de los inputs o variables de entrada, transformación de los mismos antes de su presentación a la red –preprocesamiento- y transformación de los datos de salida de la red – postprocesamiento- y, por último, en este apartado se mostrará la arquitectura de la red con todas las decisiones que ello implica: número de capas ocultas, número de nodos o neuronas por capa oculta, conexiones entre capas, algoritmo de aprendizaje, función de activación, etc.

Este trabajo se inició con muchas pruebas, en diferentes porcentajes, para los diferentes conjuntos. Esta tarea que consume principalmente tiempo, tenía dos propósitos. El primero ha sido entender con qué proporción en el conjunto de entrenamiento, la red funcionó mejor. Después de esto, el segundo objetivo fue percibir que padrón de resultados podríamos esperar en términos de la precisión de la red creada.

Así, después de muchos experimentos, se han constituido finalmente de forma aleatoria dos particiones, con **90,1%** para el conjunto de entrenamiento y **9,9%** para el conjunto de validación. Estas particiones se probaron varias veces, con el objetivo de encontrar resultados similares al anterior, en términos de precisión. La literatura ofrece un breve guía sobre cómo elegir las dimensiones de estos conjuntos (véase, por ejemplo, Nguyen y Cripps, 2001). Muchos autores hacen su selección basados en reglas empíricas de 90% vs 10%, 80% vs 20% o 70% vs 30%, para el aprendizaje y el test, respectivamente. No obstante, en la mayoría de los casos, esta selección debe basarse en el problema en concreto y, una vez más, la experimentación es la que puede guiarnos en esta elección (Zhang, 1998)

Se señala que sin el conjunto de test, tendríamos una evaluación optimista del error, sesgada. Con su uso, se tiene una estimación no sesgada del error que proporciona una verdadera generalización, desde que el conjunto de entrenamiento y de test sean aleatorios y puesto que el conjunto de test sea eficaz (siempre mayor que 70%) y cubriendo todo el espacio de datos. Sin embargo, todavía tenemos el pre-procesamiento de datos de entrada, que consiste en el cambio de la escala de datos. Esta tarea puede hacerse a través de la tipificación – conversión de una variable cuantitativa en una variable con media cero y desvío típico de 1 – o a través de un proceso de normalización, en el que la variable cuantitativa se transforma en una variable en el rango de valores [0, 1] o [-1, 1]. Sobre la necesidad de hacer algún pre-procesamiento a los datos de entrada, hay que analizar qué tipo de red y que funciones de activación serán utilizadas.

En el caso de SPSS, existe la posibilidad de elegir para las variables en el campo opción cambio de escala de Covariables (variables cuantitativas) a través de las opciones tipificados, normalización, o Corregida normalizada. También para el tipo de pre-

procesamiento, hemos hecho varios experimentos, incluyendo sin ninguna transformación. No obstante, siempre los mejores resultados se obtuvieron con la opción **tipificación** de las variables.

La transformación de las variables de salida cuantitativas dependerá de la función de error utilizada, en concreto, de la suya dependencia para las diferentes escalas de medición de las diferentes variables de salida. Si hay una alta dependencia – como en el caso más usual con la suma del error cuadrático –, al tener varias salidas con diferentes intervalos, el proceso de aprendizaje da más peso a las variables con un rango más amplio de valores. En este caso concreto, sólo hay una variable de salida, el avalúo del predio.

Otra cuestión importante es la selección del número de capas ocultas de la red y el número de neuronas por capa. No hay reglas específicas para determinar el número óptimo de neuronas y el número de capas ocultas de un problema en concreto. Hay que tener en cuenta que un número excesivo de capas puede generar ruido, sin embargo, se puede lograr una mejor tolerancia a fallos (Bonilla & Puertas, 1997). Un gran número de capas ocultas lleva a una disminución en el margen de error de aprendizaje, pero no necesariamente conduce a un pequeño error en el test. Además, el número de neuronas ocultas intervienen en la efectividad del aprendizaje y generalización de la red (Hilera & Martínez, 1995). Esto significa que, en general, pocas neuronas ocultas hacen con que la red no distingue bien las características del problema, haciendo que en la fase de aprendizaje no se alcanzan los límites deseados, mientras que con demasiadas neuronas ocultas, aunque no conduzca a resultados erróneos, genera procesos de aprendizaje muy demorados. Una capa oculta con muchas neuronas requiere menos iteraciones para la red aprender, pero cada iteración se alargará en el tiempo debido a que tiene de calcular más peso. Cada capa no deberá tener el mismo número de neuronas como padrones, ya que esto favorece la memorización de estos padrones, pero se recomienda que, en general, se usen menos neuronas que variables (Pérez & Q., 2003). En definitiva, determinar el número de capas no es una tarea sencilla; normalmente se basa en la experiencia y se llevará en función del problema en estudio y la exactitud esperada de la red.

La última consideración a tener en cuenta es la determinación de las funciones de activación de las neuronas de las capas ocultas y de salida, así como el algoritmo de aprendizaje. Las Funciones Tangente Hiperbólica e Identidad están disponibles en SPSS para la capa oculta, para nuestro caso la mejor red se obtuvo con la **Función Tangente Hiperbólica**. En el caso de la función de activación de la capa de salida (en este caso sólo con una neurona), además de estas dos funciones, tenemos también la **Función de Identidad**, con la cual también se obtuvo la mejor Red Neuronal.

Acerca de los algoritmos de aprendizaje hay que tener en cuenta que constantemente aparecen nuevas propuestas en la literatura y, a menudo, un método ideal para acelerar el aprendizaje puede producir un pobre rendimiento en una aplicación diferente (Freeman & Skapura, 1993). El SPSS, tiene dos algoritmos de aprendizaje, el *Gradiente conjugado Escalado*¹¹ y el *Pendiente de Gradiente*¹², utilizados para calcular los pesos entre las conexiones.

En cuanto al tipo de entrenamiento el mejor modelo se obtuvo con *mini-lote*¹³ y con el algoritmo de optimización ***Pendiente de Gradiente*** con el cual se estimó los pesos de las conexiones entre las neuronas, con los parámetros descritos en la Tabla 12. **Opciones de Entrenamiento de la Red**

Tabla 12. Opciones de Entrenamiento de la Red

Opciones de Entrenamiento	Valor
Tasa de Aprendizaje Inicial	0.1
Límite Inferior de la tasa de Aprendizaje	0.001
Reducción de la Tasa de Aprendizaje en Épocas	10
Impulso	0.9

11 Los supuestos que justifican este método sólo se aplican al tipo de entrenamiento **Lote**, por lo que este método no está disponible en el entrenamiento En línea o Mini lote. El algoritmo se puede encontrar en la ayuda de SPSS.

12 Este es el algoritmo de optimización necesario para el entrenamiento En línea, y Mini-lote, y es opcional para el entrenamiento batch. Cuando este algoritmo se utiliza con el mini-batch, debe especificarse el valor de algunos parámetros como la velocidad inicial de aprendizaje, el límite inferior de esta tasa, el momentum y el número de pasos (epochs). El algoritmo se puede encontrar en la ayuda de SPSS.

13 El entrenamiento mini-batch divide los datos en K grupos de tamaño aproximadamente igual, a continuación, actualiza los pesos sinápticos, después de haber pasado por un grupo. Es decir, la formación mini-batch utiliza la información de grupos de registros. El proceso también recicla los K grupos, si es necesario. El número de registros de entrenamiento para cada mini-batch es determinado por la palabra clave MINIBATCHSIZE. El entrenamiento mini-batch ofrece un compromiso entre el entrenamiento batch y el on-line, y puede ser considerado el mejor entrenamiento para todos los conjuntos de tamaño medio.

Para el algoritmo Pendiente de Gradiente la *Tasa Inicial de Aprendizaje* utilizada ha sido de **0.1** Cuando este valor es alto, significa que la red aprende rápidamente, pero puede tener el costo de la inestabilidad de la red. El *límite inferior de la tasa de aprendizaje* ha sido de **0.001**. El parámetro inicial del *impulso* utilizado fue de **0.9**.

Este valor ayuda a prevenir la inestabilidad causada por valores elevados de la tasa de aprendizaje. El número de pasadas utilizadas para reducir la velocidad inicial de aprendizaje hasta su límite inferior, ha sido igual al 10 (Épocas).

El siguiente gráfico muestra que los errores del entrenamiento y del test, disminuyen al inicio del proceso de aprendizaje Sin embargo, desde un determinado número de iteraciones, el error de entrenamiento sigue disminuyendo pero el error del test aumenta. A partir de ese punto, los parámetros de la red están a ser sobre-ajustados y la red está a modelar esencialmente ruido. Por lo tanto, este procedimiento indica que el mínimo de la curva de validación sea usado como criterio de finalización del entrenamiento.

Teniendo en cuenta todas estas cuestiones previas, la red está diseñada y pronta para “trabajar”. Todo comenzó con la elección de un vector con las variables de entrada, y la correspondiente salida, en este caso el avalúo del predio. Esta información se propagará a la última capa de la red, dando como resultado un vector de salida. Este proceso se lleva a cabo a través de la función de activación, que produce la transformación de las entradas de cada neurona de una misma capa, en salidas de neuronas de esta capa a la siguiente y que constituyen las entradas de la última capa. Una vez que haya completado este proceso con los n vectores de entrada, la red calcula el valor de salida global. A través de la diferencia con la salida deseada, se obtiene el error global por unidad de salida. A continuación, se determina la contribución relativa de las neuronas en ese error y al través de un algoritmo de aprendizaje, los errores serán modificados, repetidamente, para cada vector del conjunto de padrones de aprendizaje, hasta que el error global obtenido sea mínimo. A través del desarrollo de este comportamiento, se consigue el auto-ajuste de la red, que permite que al presentar una nueva entrada, consiga proporcionar una salida apropiada – esto es lo que se llama la capacidad de generalización de la red

El módulo de redes neuronales del SPSS, tiene una opción para seleccionar automáticamente la arquitectura. Esta opción permite seleccionar la “*mejor*” estructura de la red de forma automática.

A través de un proceso iterativo para encontrar la solución óptima global, muchas redes se han generado mediante la combinación de números diferentes de neuronas en la capa oculta

y diferentes funciones de activación. De las diversas opciones, se optó por una red Perceptrón Multicapa (MLP) con una capa oculta con la siguiente estructura **16:29-9-1:1**. Esto significa que el número de variables de entrada es de diez y seis (16)¹⁴ y tras el preprocesamiento se requieren **29** neuronas de entrada, esto se debe a que las variables USOPOT y COMUNA se codificaron según el esquema 1-de-N de manera que se necesita una neurona por clase; hay nueve neuronas o nodos en la capa oculta, y el número de unidades en la capa de salida es una sola, ya que propone llevar a cabo una regresión con una variable dependiente, el precio de venta de la vivienda. En total, resultan **280** parámetros estimados, de los cuales **diez** (10) son sesgo (*bias*).

En la capa oculta, tenemos la **tangente hiperbólica** como función de activación, en la capa de salida se utiliza la **función identidad**. Estas funciones fueron seleccionadas en el SPSS cuando se opta por la selección de la arquitectura de la red de forma personalizada. Con estas funciones se mostraron los mejores resultados, ver Tabla 13.

¹⁴ 16:29, significa que el número de neuronas de entrada es de 6 y que tras el preprocesamiento de los datos esta cifra se mantiene ya que no se han utilizado codificaciones especiales como el esquema 1 de N-inculadas a las variables cualitativas ante la ausencia de éstas

Tabla 13. Información sobre la red

Información sobre la red			
Capa de entrada	Factores	1	Comuna
		2	USOPOT
	Covariables	1	Distancia
		2	AreaCPrivada
		3	AreaTerreno
		4	Xcentroide
		5	NoHabitaciones
		6	PuntajeTotal
		7	Ycentroide
		8	Edad
		9	PuntajeEstructura
		10	NoBaños
		11	PuntajeCocina
		12	NoPisos
		13	PuntajeAcabados
14		PuntajeBaños	
Número de unidades ^a		29	
Método de cambio de escala para las		Normalizada	
Capas ocultas	Número de capas ocultas		1
	Número de unidades de la capa oculta 1 ^a		9
	Función de activación		Tangente hiperbólica
Capa de salida	Variables	1	AvaluoComercialMillon
	Número de unidades		1
	Método de cambio de escala para las		Tipificados
	Función de activación		Identidad
	Función de error		Suma de cuadrados

a. Sin incluir la unidad de sesgo

Fuente: Elaboración propia en el SPSS

Para evaluar el grado de ajuste de la red, el SPSS calcula dos medidas: la suma de cuadrados del error (SSE) y el error relativo. La mejor red, es decir, la más eficaz, será la que tiene el **menor error relativo en el conjunto de prueba**, cuya fórmula se puede ver a continuación:

$$\frac{\sum_{m=1}^M y^{(m)} - \tilde{a}^{(m)2}}{\sum_{m=1}^M y^{(m)} - \tilde{a}^2}$$

Donde:

- $y^{(m)}$ Es el precio observado para el caso m
- $\hat{y}^{(m)}$ Es el precio estimado para el caso m
- \bar{y} Es la media de los valores observados
- M Es el conjunto de casos de test

Desde el error relativo, podemos calcular la eficiencia de la red utilizada por (Pulido, Montesinos, Roldan, & Ruiz - Navarro, 2007; Pulido, Montesinos, Roldan, & Ruiz - Navarro, 2007).

$$E = 1 - \frac{\sum_{m=1}^M y^{(m)} - \tilde{a}^{(m)2}}{\sum_{m=1}^M y^{(m)} - \tilde{a}^2}$$

Por ejemplo, (Kusan, Aytakin, & Ozdemir, 2010) utilizan otras medidas, como el MAPE y el RMSE (ver fórmula en la tabla siguiente), para evaluar la capacidad de generalización de la red, medida mediante el ajuste del conjunto de Prueba, ver Tabla 14.

Tabla 14. MAPE y RMSE

MAPE Error Porcentual Absoluto Medio Mean Absolute Percentage Error	RMSE Raíz cuadrada del error cuadrado medio
$MAPE = \frac{1}{M} \sum_{m=1}^M \left \frac{\tilde{a}^{(m)} - y^{(m)}}{y^{(m)}} \right $	$RMSE = \sqrt{\frac{1}{M} \sum_{m=1}^M y^m - y^{m2}}$

A continuación, en la Tabla 15 se presentan los resultados en SPSS de “la mejor” red neuronal obtenida, la cual corresponde al modelo en el que se obtuvo el menor error relativo (0,136), en donde podemos observar que el error relativo del conjunto de Entrenamiento es casi el doble que el obtenido en el de Prueba.

Tabla 15. Resultado de la red neuronal en términos de precisión (*salida de SPSS*)

Resumen del modelo		
Entrenamiento	Suma de errores cuadráticos	80,935
	Error relativo	,263
	Regla de parada utilizada	1 pasos consecutivos sin disminución del error ^a
	Tiempo de entrenamiento	0:00:00.68
Prueba	Suma de errores cuadráticos	6,172
	Error relativo	,136
Variable dependiente: avalúo Comercial Millón		
a. Los cálculos del error se basan en la muestra de prueba.		

Fuente: Elaboración propia en el SPSS

En la Figura 14, también producida por SPSS, podemos ver la red neuronal estimada, así como la identificación de los parámetros o pesos y su señal. De igual manera se muestra de manera fragmentada, para dar mejor visibilidad, en la Figura 15, se observa las variables cualitativas COMUNA y USOPOT y en la Figura 16, todas las demás variables cuantitativas.

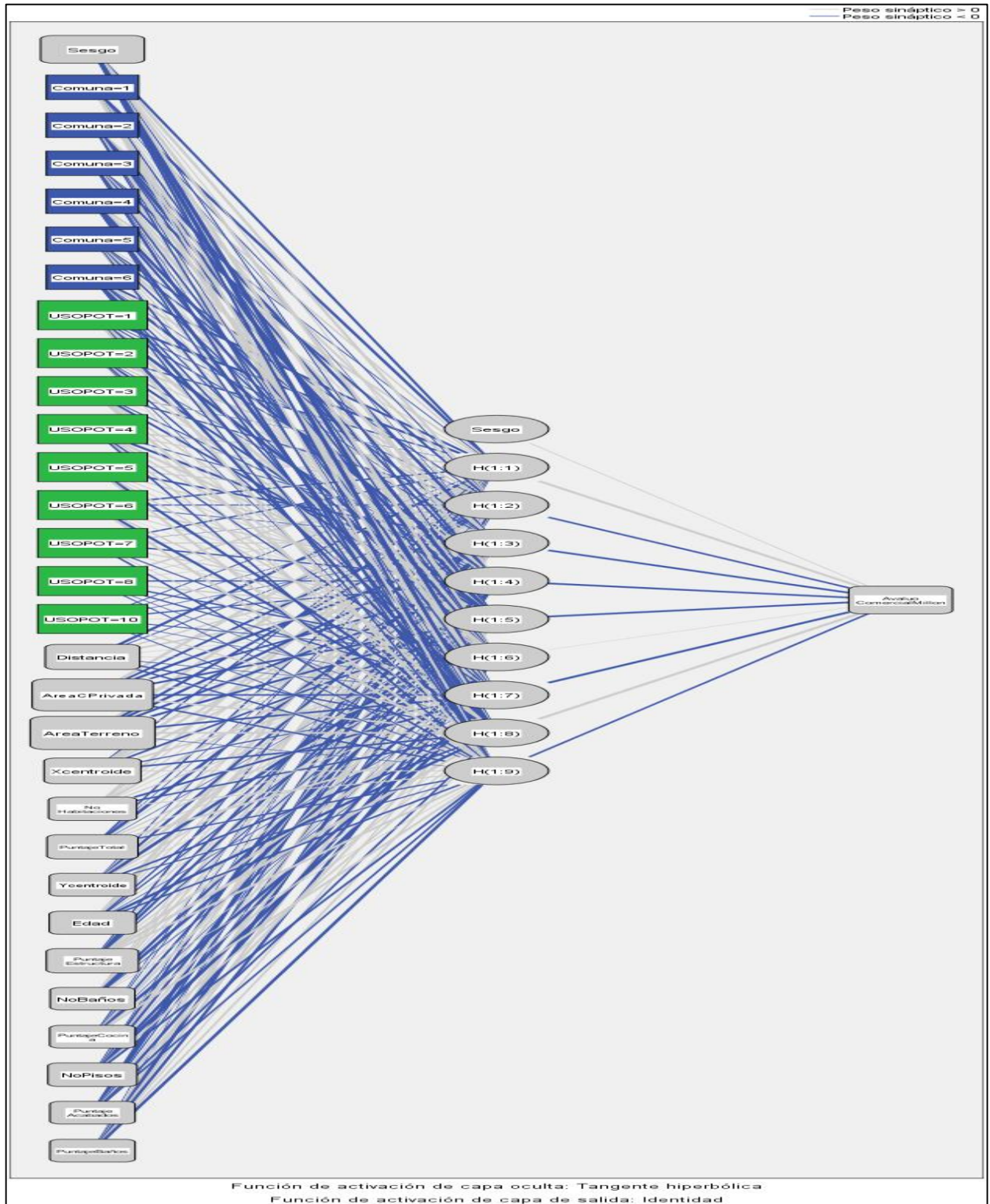


Figura 14. Esquema de la red neuronal (salida de SPSS) - Fuente: Elaboración propia en el SPSS

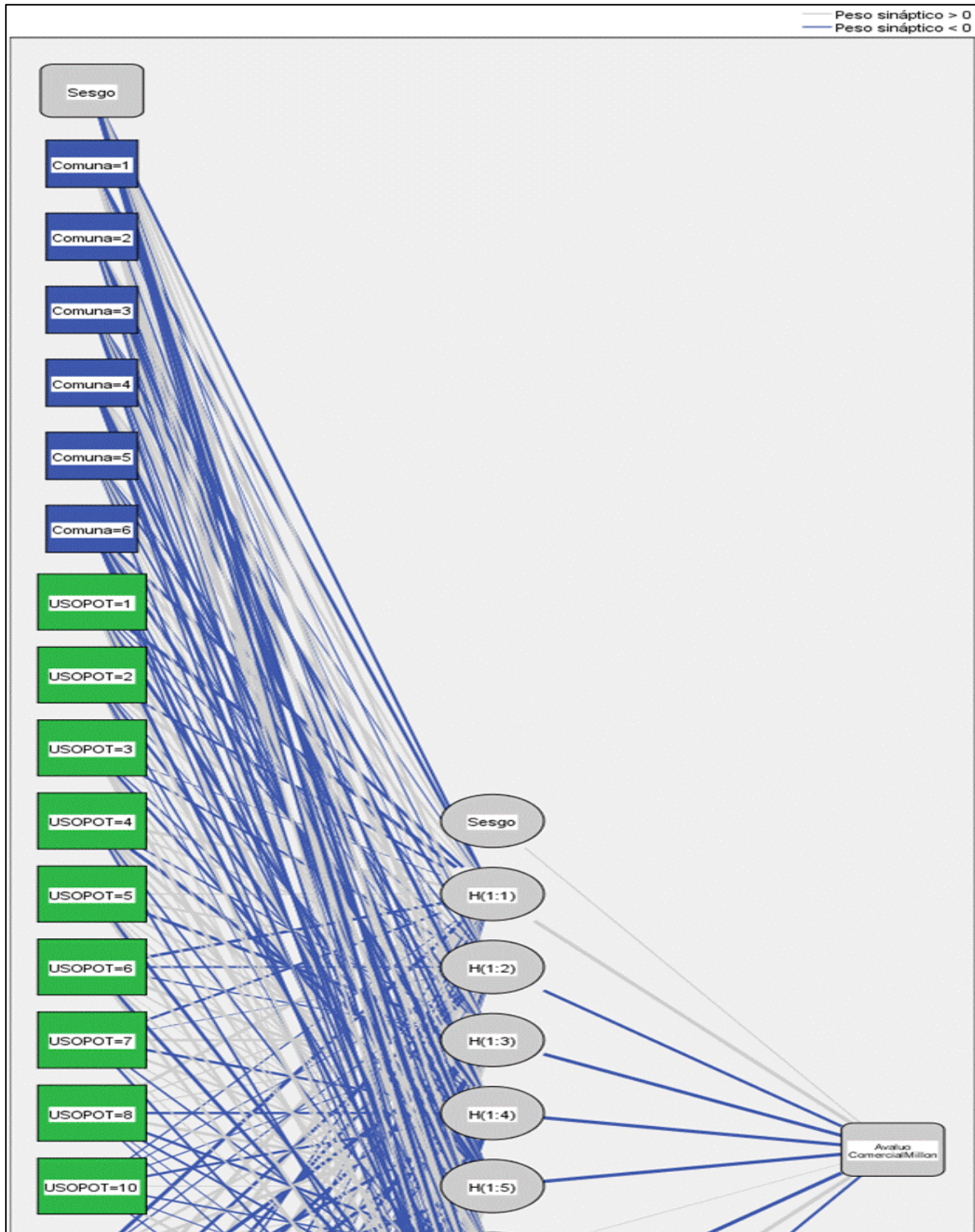


Figura 15. Red Neuronal Artificial. **Parte A**, donde se muestra variables cualitativas, COMUNA y USOPOT

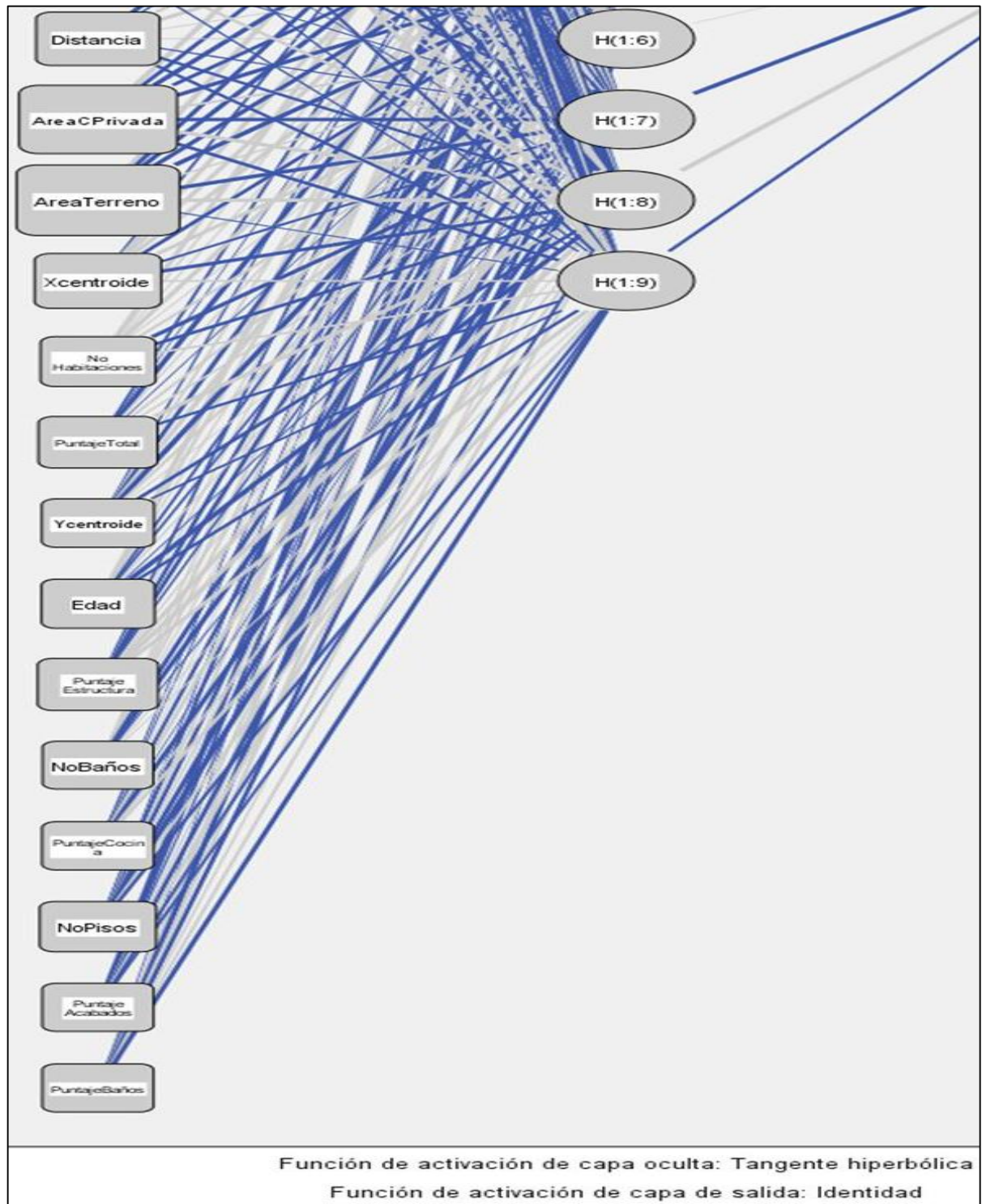


Figura 16. Red Neuronal Artificial. Parte B, donde se muestra variables cuantitativas

Otro resultado que se observa es la importancia relativa de cada variable independiente en la estimación del avalúo de la vivienda. La variable que más contribuye a explicar la totalidad del avalúo de la vivienda es el Área de Terreno, con una importancia relativa de 22,9%, seguida por el Área Cprivada, con una importancia relativa del 20,85%; en un tercer puesto se encuentra el USOPOT con 14,50% y en su orden, X Centriode con 7,6%; Distancia Centro con 6,30 %; Comuna con 5,0%; Puntaje Estructura con 4,9%; Puntaje Total con 4,4%; N.º Pisos con 3,2%; N.º Habitaciones con 2,7%; Edad con 1,9%; Y Centriode con 1,5%; Puntaje de Cocina 1,3 %; Puntaje Acabados 1,1%; Puntaje Baños 1,1% y N.º Baños con 0,8%. En la Figura 17 y Tabla 16 generados por SPSS, podemos ver la importancia relativa de cada una de las variables.

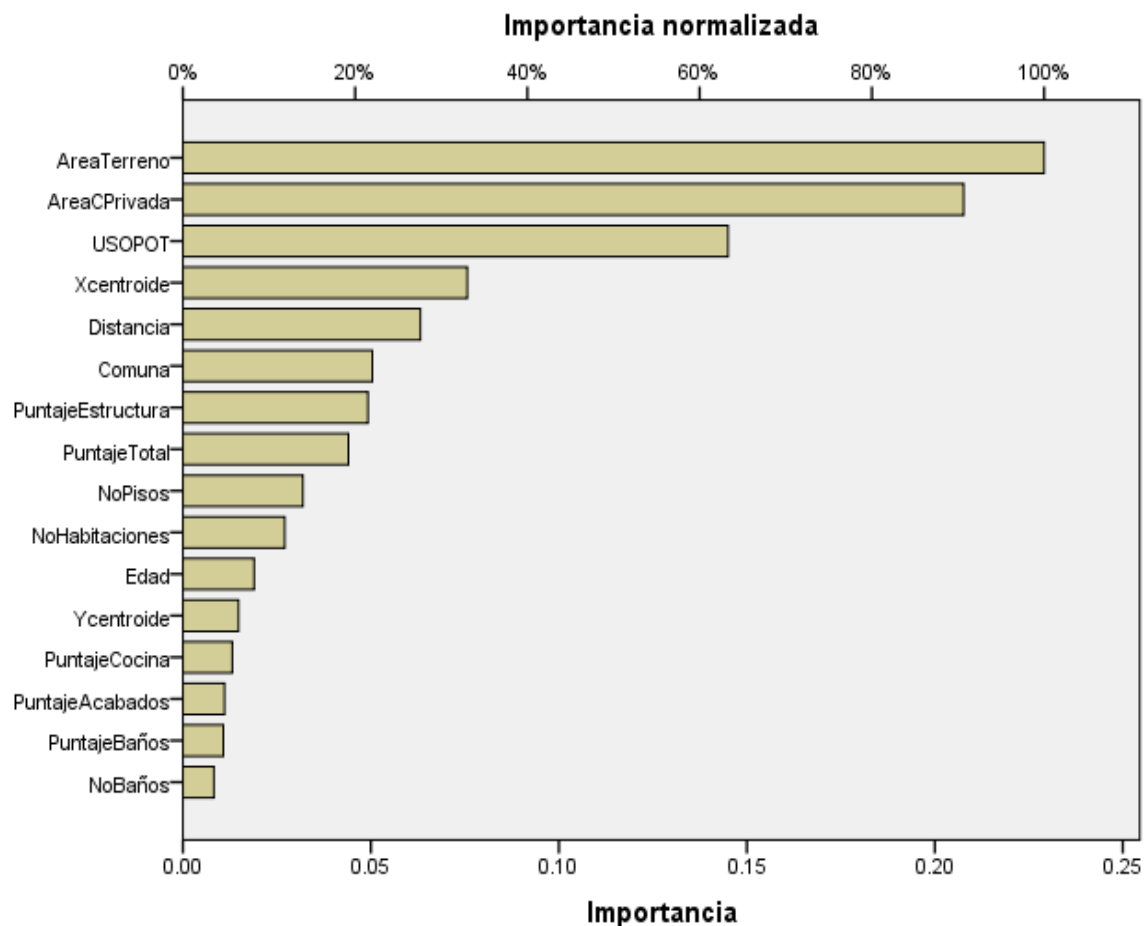


Figura 17. Importancia relativa de las variables utilizadas en la estimación de la red neuronal

Fuente: Elaboración propia en el SPSS

Tabla 16. Importancia relativa de las variables utilizadas en la estimación de la red neuronal

Variable	Importancia	Importancia Normalizada
Área Terreno	22,90%	100%
Área CPrivada	20,80%	90,70%
USOPOT	14,50%	63,30%
X Centroide	7,60%	33,00%
Distancia	6,30%	27,60%
Comuna	5,00%	22,00%
Puntaje Estructura	4,90%	21,50%
Puntaje Total	4,40%	19,20%
No Pisos	3,20%	13,90%
No Habitaciones	2,70%	11,80%
Edad	1,90%	8,30%
Y Centroide	1,50%	6,50%
Puntaje Cocina	1,30%	5,70%
Puntaje Acabados	1,10%	4,80%
Puntaje Baños	1,10%	4,70%
No Baños	0,80%	3,60%

Fuente: Elaboración propia en el SPSS

En la Figura 18 se muestra el avalúo observado Vs el avalúo estimado por la red, determinándose que el ajuste es bastante bueno, sobre todo para los predios cuyo valor comercial está situado alrededor de los \$200.000.000.

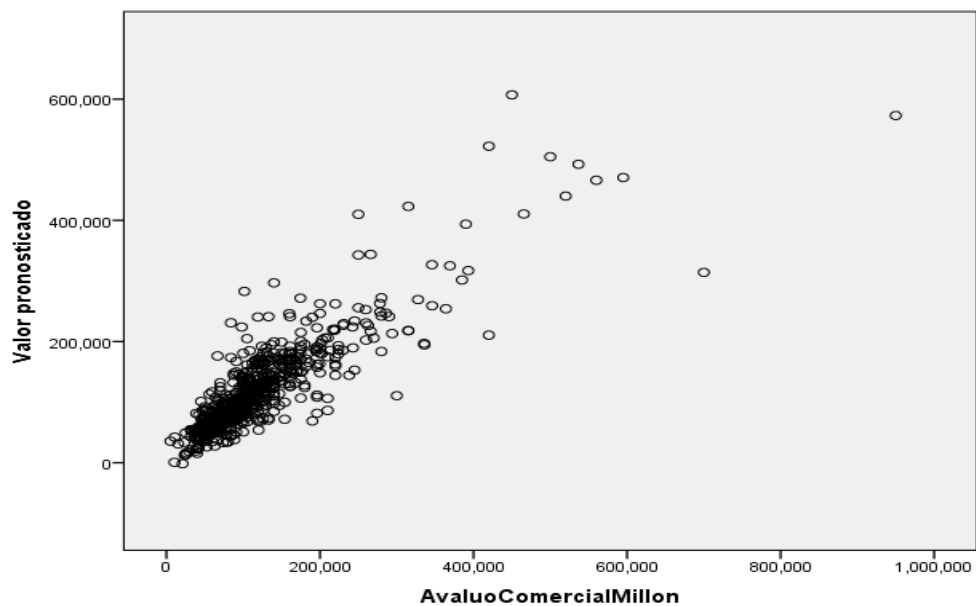


Figura 18. Precio observado vs el precio estimado por la red neuronal – Fuente: Elaboración propia en el SPSS

Observemos ahora como es el funcionamiento de una red neuronal, está en particular, siguiendo el camino de la información que ingresa en la capa de entrada hasta la capa de salida.

En primer lugar, el valor de las variables independientes que van a entrar en cada neurona de la capa de entrada se estandarizaron mediante la fórmula:

$$\frac{x - \bar{x}}{\bar{\sigma}}$$

Donde x representa el valor observado, \bar{x} es la media de los valores observados, y $\bar{\sigma}$ su desvío típico.

Estos valores entran en la red a través de la capa de neuronas de entrada. Se multiplican por los pesos de las conexiones entre las neuronas de esta capa y la capa oculta, y finalmente se suman. Por ejemplo, para el primer nodo de la capa oculta se tiene:

$$\alpha_1 = \dot{u}_{01} * x_0 + \dot{u}_{11} * x_1 + \dot{u}_{21} * x_2 + \dot{u}_{31} * x_3 + \dot{u}_{41} * x_4 + \dot{u}_{51} * x_5 + \dot{u}_{61} * x_6 + \dot{u}_{71} * x_7 + \dot{u}_{81} * x_8 + \dot{u}_{91} * x_9 + \dot{u}_{101} * x_{10} + \dot{u}_{111} * x_{11} + \dot{u}_{121} * x_{12} + \dot{u}_{131} * x_{13} + \dot{u}_{141} * x_{14} + \dot{u}_{151} * x_{15} + \dot{u}_{161} * x_{16} + \dot{u}_{171} * x_{17} + \dot{u}_{181} * x_{18} + \dot{u}_{191} * x_{19} + \dot{u}_{201} * x_{20} + \dot{u}_{211} * x_{21} + \dot{u}_{221} * x_{22} + \dot{u}_{231} * x_{23} + \dot{u}_{241} * x_{24} + \dot{u}_{251} * x_{25} + \dot{u}_{261} * x_{26} + \dot{u}_{271} * x_{27} + \dot{u}_{281} * x_{28} + \dot{u}_{291} * x_{29}$$

Dónde x_{ij} , son las variables que se relacionan en la Tabla 17

Tabla 17. Relación de Variables incorporadas al modelo

x_0	Sesgo	x_{10}	USOPOT 4	x_{20}	No Habitaciones
x_1	Comuna 1	x_{11}	USOPOT 5	x_{21}	Puntaje Total
x_2	Comuna 2	x_{12}	USOPOT 6	x_{22}	Y Centroide
x_3	Comuna 3	x_{13}	USOPOT 7	x_{23}	Edad
x_4	Comuna 4	x_{14}	USOPOT 8	x_{24}	Puntaje Estructura
x_5	Comuna 5	x_{15}	USOPOT 10	x_{25}	No Baños
x_6	Comuna 6	x_{16}	Distancia	x_{26}	Puntaje Cocina
x_7	USOPOT 1	x_{17}	Área Cprivada	x_{27}	No Pisos
x_8	USOPOT 2	x_{18}	Área Terreno	x_{28}	Puntaje Acabados
x_9	USOPOT 3	x_{19}	X Centroide	x_{29}	Puntaje Baños

Fuente: Elaboración propia

Los pesos de las conexiones ω_{ij} son calculados por el software SPSS y se pueden ver en la tabla siguiente (Tabla 18).

Tabla 18. Parámetros de la Red Neuronal

Predictor		Pronosticado									
		Capa oculta 1								Capa de salida	
		H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	H(1:8)	H(1:9)	AvaluoComercialMillon
Capa de entrada	(Sesgo)	-,326	,386	,074	,454	-,076	,542	-,794	-1,208	-,274	
	[Comuna=1]	-,415	-,068	-,008	,054	,119	-,138	,251	,392	,257	
	[Comuna=2]	,175	,311	-,481	-,314	-,720	-,211	,644	-,055	-,185	
	[Comuna=3]	-,056	-,315	,485	-,276	,227	,556	-,192	-,014	-,349	
	[Comuna=4]	,311	,019	,247	,215	-,521	-,342	,031	-,591	-,187	
	[Comuna=5]	,085	-,142	-,220	,422	-,189	,132	,094	-,235	,022	
	[Comuna=6]	-,428	,366	-,054	-,504	,127	-,141	-,203	-,507	-,147	
	[USOPOT=1]	-,670	-,066	-,138	,164	-,643	,597	,036	-,484	,405	
	[USOPOT=2]	-,344	,524	,136	-,063	-,377	,312	-,165	-,046	,060	
	[USOPOT=3]	-,037	-,294	,012	-,147	-,404	,247	,147	,600	,058	
	[USOPOT=4]	,702	,266	-,157	-,713	,762	,345	-,212	,261	-,241	
	[USOPOT=5]	,208	,372	-,717	,284	,096	,087	,301	-,498	,410	
	[USOPOT=6]	-,603	-,294	-,154	,586	,137	,011	,290	-,542	,472	
	[USOPOT=7]	-,385	-,009	,422	-,449	,409	,112	-,172	,098	-,276	
	[USOPOT=8]	,577	,132	,112	-,418	,599	-,417	-,294	-,053	-,119	
	[USOPOT=10]	-,120	-,057	-,080	-,397	,026	-,214	-,066	,381	-,114	
	Distancia	-,686	-,117	,407	,675	,056	,058	-,102	-,372	-,372	
	AreaCPrivada	1,128	-,462	-1,234	-,519	-,933	,308	-,825	,856	-,463	
	AreaTerreno	,873	-,538	-,695	-,656	-1,682	,148	-1,333	1,279	-,064	
	Xcentroide	-,158	,580	,560	,573	-,423	,540	-,253	-,686	,229	
	NoHabitaciones	,568	,113	,084	,561	,056	,269	-,612	-,449	,268	
	PuntajeTotal	-,208	-,022	-,636	,101	-,030	,239	,138	,647	-,341	
	Ycentroide	-,096	,175	-,409	,552	-,836	,269	,034	-,410	-,191	
Edad	,698	,282	,089	-,186	-,151	,087	-,049	-,837	-,351		
PuntajeEstructura	-,202	,070	-,328	-,365	-,123	,093	,560	1,064	,464		
NoBaños	,105	,270	,216	,057	-,416	-,535	-,050	,049	,055		
PuntajeCocina	-,042	-,046	,030	-,274	,366	-,011	-,523	,144	,354		
NoPisos	-,038	-,322	,254	-,312	,154	-,439	-,716	,234	-,307		
PuntajeAcabados	,010	-,663	-,231	-,087	,148	-,381	,351	-,252	-,222		
PuntajeBaños	-,015	,018	,449	-,580	,352	-,393	,129	,242	-,486		
Capa oculta 1	(Sesgo)									,074	
	H(1:1)									,816	
	H(1:2)									-,457	
	H(1:3)									-,849	
	H(1:4)									-1,022	
	H(1:5)									-1,600	
	H(1:6)									,025	
	H(1:7)									-1,350	
	H(1:8)									1,778	
H(1:9)									-,348		

Fuente: Elaboración propia en el SPSS

La función de activación utilizada en la capa oculta fue la tangente hiperbólica (selección personalizada en el SPSS), que se aplicó a los valores que se obtuvieron en la ecuación anterior.

$$\tanh(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$$

Los valores que resultan son los valores de las neuronas de la capa intermedia (o oculta).

El procedimiento para obtener el valor de la neurona de salida, a partir de las neuronas de la capa oculta, es lo mismo que se utiliza para obtener la capa intermedia a través de la capa de entrada. En cada neurona de la capa oculta, ha sido activada la función a partir de la que se procesará la respuesta de la red a través de la siguiente expresión:

$$y_k = \sum_{j=0}^m \hat{u}_{jk} * f(a_j)$$

Donde \hat{u}_{jk} son los pesos de las conexiones entre la capa oculta y la capa de salida, y f es la función de activación – en este caso, es la de **Identidad**.

Veamos ahora un ejemplo concreto, el predio N° 71 tenía un avalúo comercial de \$202.990.000, y el avalúo estimado por la red fue de \$202.806.000, es decir que se obtuvo una diferencia de sólo \$184 (la menor diferencia obtenida por la red) Las características de este predio se relacionan en la Tabla 19.

Tabla 19. Características del predio No 71 que procesa la red.

Variable	Valor	Variable	Valor
Área Terreno	158	No Pisos	2
Área CPrivada	128	No Habitaciones	5
USOPOT	2	Edad	52
X Centroide	968125	Y Centroide	970964
Distancia	1173	Puntaje Cocina	8
Comuna	6	Puntaje Acabados	5
Puntaje Estructura	19	Puntaje Baños	9
Puntaje Total	41	No Baños	2

Fuente: Elaboración propia

En la Tabla 19, podemos observar los valores de entrada que fueron dados a la red para cada una de las variables independientes. Estos valores de entrada, después de normalizados, se multiplican por los pesos respectivos de las conexiones entre las neuronas de la capa de entrada y la capa oculta, también es sumado el valor del *bias* (bla 18).

Después de calculado el valor de esta combinación lineal, es aplicada la respectiva función de transferencia que, en este caso, es la tangente hiperbólica. Este resultado es el valor de entrada en la primera neurona de la capa oculta. El valor de las otras neuronas en la capa intermedia se calcula del mismo modo, así como el valor final de la red. Sin embargo, nótese que el valor obtenido y_k es todavía estandarizado. Tenemos que convertirlo para obtener el valor deseado. El valor real, que corresponde al precio estimado de la vivienda por la red neuronal, viene dado por $y_k \times s + x$ (dada la fórmula de estandarización), dando como resultado el valor estimado por la red de \$202.806.000, como ya se mencionó, muy cerca del valor observado de \$202.990.000.

Es de precisar que el procedimiento anteriormente descrito de estimación del valor final es realizado de manera automática por el software SPSS.

De acuerdo con la revisión de literatura ya expuesta, cuando la Red Neuronal cuenta con muy pocas variables de entrada se corre el riesgo de reducir el espacio de búsqueda, dando lugar a un posible incremento del sesgo. Sin embargo, también demasiadas variables hacen que la red se vuelva muy compleja, y pueden no aportar información relevante para la reducción del error de generalización. Situación que se podría presentar para el presente estudio, pues la red finalmente seleccionada consta de diez y seis (16) variables de entrada, con un total de 280 parámetros estimados, no obstante para llegar a la obtención de este modelo se realizaron innumerables pruebas que fueron desde el cambio de orden de las variables, particiones de conjuntos, arquitectura y entrenamiento, hasta la conjugación de cada uno de estos con la eliminación progresiva de cada una de las variables, analizando el comportamiento de la red en cada caso. Fue así como se diseñaron redes sin las variables cualitativas COMUNA Y USOPOT, pues cada una de ellas generaba 6 y 9 clases, respectivamente, es decir que dos variables se convertían en 15, para un total final de 29 variables, situación que podría aumentar el sesgo de la red. Con el propósito de evaluar esta situación, se tomó la red seleccionada, dejándole exactamente los mismos parámetros y se le eliminó únicamente la variable COMUNA, obteniéndose como resultado el aumento del error relativo a 0,179. Luego, nuevamente se procedió de la misma manera, pero se eliminó solo la variable USOPOT y en esta oportunidad el error aumentó a 0,195. Por último, al mismo modelo se le eliminó simultáneamente las variables COMUNA y USOPOT, obteniéndose el más alto error relativo de 0,224. La situación antes descrita demuestra que las variables cualitativas COMUNA y USOPOT, tienen un aporte importante para la determinación del valor del inmueble. Lo anterior se ratifica, al analizar el gráfico de nivel de importancia de las variables obtenido del SPSS, ver Tabla 16, según el cual, del total de las

variables (16) involucradas en el modelo, la correspondiente a USOPOT ocupa el tercer puesto, con un aporte 14,5% de importancia en el modelo, así mismo la variable COMUNA ocupa el sexto puesto con una participación del 5%, es decir que las dos aportan información relevante para el aumento de eficiencia de la red y por lo tanto contribuyen a la disminución del sesgo.

Vale la pena precisar que para el caso de la variable USOPOT, se esperaba este nivel de respuesta del modelo, teniendo en cuenta que durante los últimos años la Oficina de Planeación del Municipio de Fusagasugá está limitando el desarrollo urbanístico al cumplimiento de lo normado en el Plan de Ordenamiento Territorial, por lo tanto la expedición de Licencias de Construcción se condicionan al uso normado que tenga el suelo donde se ubica el inmueble, aspecto que se ha convertido en un mecanismo de regulación efectiva del uso del suelo y ocupación del territorio. Como es lógico, la aplicación de esta política espacial de desarrollo, termina por reflejarse en el mercado de los inmuebles, pues el uso permitido finalmente es un condicionante de su valor, aspecto que evidentemente se está reflejando en el modelo de red neuronal propuesto.

Con relación a las demás variables de ubicación espacial, como lo son: Coordenada X (XCentroide), Distancia al Centro y Comuna, se determinó que estas también tienen un aporte importante al momento de definir el avalúo de un predio, teniendo en cuenta que dentro del nivel de importancia ocupan el cuarto, quinto y sexto puesto, con niveles de importancia del 7,6%, 6,3% y 5%, respectivamente. Por el contrario, la variable Coordenada Y (YCentroide) ocupa los últimos puestos (puesto 12) con un nivel de importancia de apenas el 1,5%.

Analizando las anteriores cifras, el primer cuestionamiento que surge es porque se da una diferencia tan significativa entre el nivel de importancia de la coordenada X (7,6%) y la Coordenada Y (1,5%), si sabemos que las dos en conjunto definen la ubicación de un punto en el plano cartesiano. Para encontrar la respuesta a este interrogante, se analizó la conformación del Plano de Conjunto Urbano del Municipio de Fusagasugá, observándose que éste tiene de manera predominante una forma alargada y que buena parte de su territorio es atravesado por una vía de importancia nacional como lo es la autopista que conduce a Melgar, la cual evidentemente ha jalando el desarrollo y crecimiento urbano de manera importante en este sentido (Eje X), ver Figura 11.

Con relación a la variable Distancia¹⁵, Inicialmente se consideró que de las variables de localización esta podría ser la que tendría mayor participación dentro del modelo, pues en Municipios como este, en proceso de desarrollo, los ciudadanos tienen mayor preferencia por vivir cerca al centro de la ciudad, pues allí es donde se concentran los centros de decisión y consumo de la ciudad. No obstante, al parecer esta preferencia está cambiando, porque el municipio de Fusagasugá hoy tiene una vocación más turística o de descanso, lo que hace que los ciudadanos prefieran tener sus viviendas ubicadas en zonas más alejadas, pero no distantes, que ofrecen mayor tranquilidad y calidad de vida.

De otra parte, con relación a la variable Comuna, se muestra que tiene una influencia importante en el modelo (5%), aunque inicialmente no se consideró que sería de esta forma, pues la misma obedece a una división político-administrativa que el municipio ha creado al interior de la zona urbana, la cual al parecer en este caso está creando unas estructuras espaciales que cobran una conciencia de clase, pues da un idea de las condiciones socioeconómicas de cada Comuna que refleja las características del entorno, pero sobre todo por el vecindario donde se encuentra ubicado el predio (comunales), aspecto de ubicación que indudablemente está ejerciendo influencia interesante al momento de determinar el valor de un inmueble.

Por último, es necesario analizar las variables de puntaje de la construcción correspondientes a cocina, acabados y baños, las cuales tienen los niveles más bajos de importancia, esto a pesar de que las mismas fueron diseñadas para reflejar la calidad de las características constructivas de estas zonas del predio, las cuales, como es lógico, afectan directamente el valor del mismo. Al respecto, vale la pena precisar que durante los últimos 25 años esta metodología de calificación del IGAC funcionó adecuadamente, sin embargo, ya se ha evidenciado que hoy se encuentra rezagada, pues durante estos años el mercado ha puesto a disposición de los constructores, una gran gama de materiales que no se encuentran previstos en el sistema de calificación desarrollado por el Instituto, como por ejemplo la madera, el porcelanato, piedras naturales, etc., por lo tanto, predios que han sido construidos con este tipo de materiales, la calificación dada por el IGAC seguramente no va a reflejar estas características. Esta puede ser una de las razones por las cuales los puntales (cocina, acabados y baños) tienen un nivel bajo de importancia dentro del modelo.

¹⁵ Es la medida en metros, desde el lugar de ubicación del predio al centro de la ciudad (plaza central).

5.4 Comparación de los Avalúos obtenidos con Metodología del IGAC y la Red Neuronal.

Teniendo en cuenta que uno de los objetivos del presente estudio es el de comparar los avalúos generados con el método tradicional de Zonas Homogéneas, con los obtenidos por la red neuronal seleccionada, se aplicó una prueba paramétrica T de *Student* para datos pareados o muestras relacionadas, para determinar si existen diferencias significativas entre las medias de las dos poblaciones de datos. Para lo cual en primera instancia se tomó la población de datos de los avalúos comerciales de los puntos de investigación y se comparó con los avalúos obtenidos con la metodología de Zonas Homogéneas del IGAC¹⁶, obteniéndose los resultados que se muestran en la Tabla 20.

Tabla 20. Prueba T, comparación avalúo Comercial Vs avalúo Calculado por el método de Zonas

a. Estadísticos descriptivos del procedimiento *prueba T* para muestras relacionadas

Estadísticos de muestras relacionadas					
		Media	N	Desviación típ.	Error típ. de la media
Par 1	Avalúo Comercial Millón	118518.10	685	86925.821	3321.265
	Avalúo Método IGAC	108227.32	685	91589.349	3499.449

b. Coeficiente de correlación de *Pearson*

Correlaciones de muestras relacionadas				
		N	Correlación	Sig.
Par 1	Avalúo Comercial Millón y Avalúo Método IGAC	685	,641	,000

¹⁶ En la metodología de Zonas Homogéneas del IGAC, el profesional que realiza el estudio, inicialmente expone al Comité de Avalúos los valores a nivel comercial obtenidos en el estudio. Posteriormente los miembros de dicho Comité deciden si adoptan el 60, 70, u otro porcentaje de dicho valor calculado para efectos catastrales. Finalmente, los avalúos que son puestos en vigencia corresponden al porcentaje del valor comercial que haya decidido el Comité. Estos avalúos son los que finalmente remite el IGAC a las Tesorerías municipales para efectos de la liquidación del impuesto predial. En este estudio para hacer comparables las dos metodologías, se tomaron los avalúos calculados a nivel comercial con la metodología de zonas y no los avalúos que hoy están en vigencia, que para el caso del municipio de Fusagasugá zona urbana corresponden al 60% del valor comercial investigado.

c. Tabla resumen del procedimiento *prueba T* para muestras relacionadas

	Diferencias relacionadas					Sig. (bilateral)
	Media	Desviación típ.	Error típ. de la media	99% Intervalo de confianza para la diferencia		
				Inferior	Superior	
Avalúo Comercial Millón – Avalúo Método IGAC	10290,788	75728,530	2893,438	2816,932	17764,644	,000

La Tabla 23 nos muestra el coeficiente de correlación de Pearson, donde se establece el nivel de asociación entre las dos variables, el cual fue altamente significativo ($P < 0.01$), pero bajo, es decir existe poca asociación entre el avalúo comercial y el avalúo calculado con la metodología del IGAC.

De otra parte, en la **Tabla 20**, puede observarse que la probabilidad asociada al estadístico t (prueba de t) es altamente significativa ($P < 0.01$) por lo que se rechaza la hipótesis nula de igualdad de las medias muestrales, es decir que el avalúo comercial es diferente del calculado con la metodología del IGAC.

Ahora, en la **Tabla 21** se muestra el análisis estadístico de la comparación entre los avalúos comerciales y los obtenidos con la red neuronal.

Tabla 21. Prueba T, comparación avalúo comercial Vs avalúo calculado por la red neuronal

a. Estadísticos descriptivos del procedimiento *prueba T* para muestras relacionadas

Estadísticos de muestras relacionadas					
		Media	N	Desviación típ.	Error típ. de la media
Par 1	Avalúo Comercial Millón	118518,10	685	86925,821	3321,265
	Avalúo Método RNA	118237,70	685	76386,642	2918,584

b. Coeficiente de correlación de *Pearson*

Correlaciones de muestras relacionadas				
		N	Correlación	Sig.
Par 1	Avalúo Comercial Millón y Avalúo Método RNA	685	,868	,000

C. Tabla resumen del procedimiento *prueba T* para muestras relacionadas

		Diferencias relacionadas				
		Media	Desviación típica.	Error típico. de la media	99% Intervalo de confianza para la diferencia	
					Inferior	Superior
Par 1	Avalúo Comercial Millón – Avalúo Método RNA	280,403	43155,130	1648,873	3978,695	4539,501

El coeficiente de correlación de Pearson fue cercano a 1 (Tabla 21) lo que indica que el nivel de asociación entre las dos variables es muy alto y es altamente significativo ($P < 0.01$) es decir existe una alta asociación entre el avalúo comercial y el avalúo calculado por la red neuronal.

En cuanto probabilidad, esta fue no significativa ($P > 0.01$) por lo que se acepta la hipótesis nula de igualdad de las medias muestrales, es decir que los avalúos comerciales son muy parecidos a los avalúos calculados con la red neuronal. Al determinarse que no hay diferencia entre las medias muestrales, se establece que el modelo de red obtenido, es muy preciso, y por consiguiente tiene una mejor capacidad de reproducir los avalúos frente a los obtenidos con el método que viene aplicando el IGAC.

El anterior resultado se ajusta a lo encontrado por otros autores que han puesto a prueba el potencial de las redes neuronales tipo *Perceptrón multi-capas* en el ámbito de la valoración de la propiedad. Por ejemplo, García, *et al.*, (2008), han realizado varias comparaciones entre diferentes redes y llegan a la conclusión, como muchos otros autores ya habían llegado, que estas redes presentan un buen desempeño con este tipo de datos (García, Gámez, & Alfaro, 2008). De igual forma, Según Borst, la exactitud de las RNA las torna rivales de los métodos de regresión lineal múltiple. El autor cree que estos pueden ser utilizados para evaluar en masa, así como para ejercer un control de calidad sobre los valores estimados por otros métodos (Borst R. , 1995).

Esta capacidad de las redes se le atribuye el hecho de que este tipo de sistemas somete las muestras a procesos matemáticos mucho más complejos que un modelo de regresión múltiple, el cual se limita a construir una simple ecuación polinómica, pues la red cuenta con una mayor potencia computacional por tener las neuronas conectadas entre sí en paralelo y de forma masiva formando capas. Esta estructura implica que cada neurona puede estar conectada con muchas otras, de manera que la información de entrada se distribuye entre una gran cantidad de neuronas que trabajan en paralelo y, en cada momento, hay una gran cantidad de estas procesando la información simultáneamente (Gallego, 2004).

Al revisar las pruebas estadísticas que aplica el IGAC para determinar los avalúos, se observó que estas son aún menos robustas que las regresiones múltiples, pues para estimar el avalúo de los terrenos la prueba que aplica es la estimación de Intervalos de Confianza y para la determinación del valor de las construcciones se elaboran modelos de Regresión Simples, los cuales en ambos casos son procedimientos estadísticos que revisten muy bajo poder, frente a la complejidad y heterogeneidad de la información que deben procesar, pues esta corresponde a la totalidad de los predios que integran un municipio.

De otra parte, se estableció que los resultados de la red también se vieron favorecidos al haberse incorporado al modelo variables de localización del predio¹⁷ que permitieron una mayor capacidad de predicción del mismo, pues estas desempeñan un papel determinante en la definición del avalúo de un inmueble. Tal como ya lo habían expresado otros autores según los cuales de todos los atributos del predio, la ubicación es la más importante, porque está relacionada con la configuración espacial del producto (inmovilidad). La importancia de la ubicación está relacionada con las accesibilidades (oferta y calidad de las carreteras y del transporte), con las características del barrio y con el uso del suelo en entorno de la propiedad. La medición de estos efectos es difícil, porque no son directamente cuantificables. Se usan variables *proxy* como la distancia al centro de comercio o urbano (González & Formoso, 2000)

Por último, también es de resaltar que en el ámbito de la valoración inmobiliaria, muchos de los estudios establecen que otra ventaja de los sistemas de IA¹⁸ es su capacidad para estimar el valor de las propiedades que presentan características significativamente diferentes de las de los bienes de su entorno (*outliers*¹⁹), lo cual también pudo beneficiar los resultados obtenidos, pues en el Municipio de Fusagasugá existe un alta heterogeneidad entre los predios que conforman la zona urbana, pues es fácil encontrar inmuebles de buenas características, seguidos de otros con menores condiciones constructivas, y es claro, que la metodología del IGAC no está en capacidad de incorporar información de este nivel de complejidad, por que utiliza estadística muy básica en sus análisis, es decir, que la definición de los valores que finalmente serán adoptados, en gran parte dependen del criterio del

¹⁷ Variables como: USOPOT, Comuna, Distancia al centro, Xcentroide y YCentroide,

¹⁸ Inteligencia Artificial: Es la ciencia e ingenio de hacer máquinas inteligentes, especialmente programas de computo inteligentes. (Gonzalo, 1987).

¹⁹ Es un valor atípico, es una observación que es numéricamente distante del resto de datos.

profesional que realiza el estudio y no de un método estadístico que integre todas las variables a evaluar y por consiguiente respalde el avalúo final asignado al inmueble.

Recientemente la Dirección General del Catastro Español ha puesto en marcha un proyecto en determinadas regiones (Madrid, La Coruña, Sevilla, Valencia y Castilla-La Mancha) (Fernández, Llorca, Valero, & Botti, 2012) para estimar el valor de la vivienda a precios de mercado precisamente utilizando una de las metodologías desarrolladas en este trabajo: Redes Neuronales Artificiales. Aunque, en principio, estos cálculos del valor de mercado de los inmuebles no tendrán repercusión en el Impuesto sobre Bienes Inmuebles, sí podrían verse afectados en un futuro otros impuestos como Transmisiones Patrimoniales o el impuesto sobre la Renta.

Conclusiones

La elección del tema tratado en este trabajo de grado queda justificada por la importancia que reviste en la economía del país la determinación de avalúos que se ajusten a la realidad física de los inmuebles y del mercado inmobiliario, pues estos son la base de la liquidación del Impuesto Predial Unificado que es la principal fuente de ingresos de la mayor parte de los municipios en Colombia. Sin embargo la más subjetiva y compleja de las características catastrales es el avalúo que se le asigna al inmueble, por lo tanto es, ha sido y probablemente será, la principal fuente de discusión en la implementación de los procesos Actualización Catastral. El motivo de esta controversia no es otro, sino el hecho de que el avalúo catastral es la base para la liquidación del Impuesto Predial Unificado que grava el inmueble.

El Instituto Geográfico Agustín Codazzi es la entidad encargada de adelantar en la mayor parte del territorio nacional los procesos de Actualización de la Formación Catastral, y para este fin, hace más de treinta años desarrolló una metodología para calcular los avalúos catastrales, a la cual no se le ha realizado ningún ajuste a pesar del tiempo transcurrido y de los cambios que se han dado en las características constructivas de los predios y en la complejidad del mercado inmobiliario, por lo tanto se viene fijando un valor a los predios, sin tener presente la totalidad de sus características físicas y de entorno que resultan determinantes a la hora de atribuir un valor al inmueble.

Motivo por el cual este trabajo tenía como objetivo principal desarrollar una metodología basada en Redes Neuronales Artificiales para calcular el avalúo catastral de un predio utilizando las variables que se encuentran disponibles en el Sistema de Información Catastral del IGAC, con la cual se espera mejorar los niveles de precisión de los mismos.

Como resultado de la investigación de obtuvieron las siguientes conclusiones:

1. Se desarrolló una metodología basada en RNA como herramienta para determinar el avalúo de un predio, la cual presenta importantes ventajas, dada la flexibilidad y capacidad de predicción de una red, permitiendo recoger relaciones no lineales entre las variables explicativas Áterreno, Áprivada, UsoPOT, Xcentroide, Ycentroide, Distancia, Comuna, Puntaje Estructura, Puntaje Total, No Pisos, No Habitaciones, Edad, Puntaje Cocina, Puntaje Acabados, Puntaje Baños y No Baños. Se considera que el uso de estas variables puede servir para explicar gran parte de la variación del precio de la vivienda en otros Municipios del país.

2. El presente estudio ha demostrado la importancia del uso de las variables de ubicación o localización del predio (USOPOT, COMUNA, DISTANCIA, XCENTROIDE y YCENTROIDE al momento de definir el avalúo de un inmueble. lo cual es lógico debido a que estas incorporan atributos que generan beneficios o satisfacción a los residentes. Es aquí donde la Geomática como ciencia de la información espacial tiene un papel fundamental en el cálculo de las variables de ubicación a través de los Sistemas de Información Geográficos, sin embargo, y a pesar que las mismas vienen siendo levantadas en campo por el IGAC y se encuentran disponibles para su uso y aplicación en la Geodatabase del plano digital del municipio de Fusagasugá, actualmente estas no son tenidas en cuenta por el Instituto dentro del procesamiento estadístico de la metodología de valoración masiva de Zonas Homogéneas para la determinación de los avalúos catastrales de los inmuebles.
3. De las variables que actualmente emplea el IGAC para la determinación del avalúo se destacan Area Terreno, Area CPrivada, Puntaje Estructura, Puntaje Total, N° pisos y N° Habitaciones, haciendo precisión que estas variables pueden cambiar su nivel de importancia dependiendo el municipio donde se realice el estudio.
4. La prueba paramétrica T de *Student* para datos pareados o muestras relacionadas, corroboró el mayor poder de la Red Neuronal Artificial para estimar los avalúos respecto de resultados obtenidos con la metodología del IGAC.
5. Los recientes e importantes avances en los Sistemas de Información Geográfica permiten, en la actualidad, el tratamiento estadístico de datos georreferenciados, constituyendo una línea de investigación fundamental en materia de valoración. Concretamente, en la aplicación que aquí se ha desarrollado, el plano digital de Fusagasugá ha constituido una herramienta fundamental para la localización de los predios muestrales en la ciudad. Las posibilidades de añadir información mediante la adición de otras capas de información son enormes, como por ejemplo, capa con zonas verdes, riesgos ambientales, capa con servicios, etc.
6. El proceso de selección de las variables explicativas ha sido largo, aunque en términos de cálculo, esta tarea sea muy simplificada con el uso de software estadístico. Tener una adecuada cantidad de información ha posibilitado, por una parte, la expectativa de encontrar un modelo más cercano de la realidad del mercado inmobiliario, pero también, aumentó considerablemente el trabajo de investigación y preparación de datos.
7. En Redes Neuronales Artificiales, como en cualquier otra aplicación de métodos estadísticos, es importante homogeneizar de alguna manera la información que se

va a suministrar al sistema. Nos referimos al rango o escala de medida de las variables. En el caso de RNA, normalizar previamente los inputs tiene efectos importantes como un aprendizaje más rápido.

8. Los modelos de precios de la vivienda, desde que sean actualizados, traerían beneficios para todos los intervinientes del mercado inmobiliario. Estos modelos, podrían servir de base tanto para el vendedor o el comprador con el fin de objetivar el valor de la transacción. Estos modelos, que serían relativamente sencillos de implementar por un ordenador, podría ser parte de un sistema de apoyo a la decisión, de las partes involucradas en el sector inmobiliario. De hecho, actualmente una misma propiedad puede tener múltiples valores, en función del objetivo previsto a su evaluación. La propiedad tiene un valor, que puede ser objetivamente calculado utilizando modelos matemáticos, y este valor de base, puede someterse a imposición de acuerdo a su finalidad prevista, la compra, el pago de impuestos municipales o para préstamos bancarios. En Colombia, pueden existir dos predios en la misma calle, uno de lujo y el otro normal, y pueden no tener ninguna, o prácticamente ninguna diferencia de valor para efectos fiscales, y creemos que esta situación no es justa. El modelo econométrico propuesto, proporciona la determinación del valor de la vivienda, para todas, las características concretas y mediante un cálculo sencillo y sin subjetividad.
9. Para el caso de las ciudades de tamaño mediano y grande, que tienen un número suficiente de información del mercado inmobiliario, las redes neuronales ofrecen una seria alternativa a los modelos tradicionales. En general, los errores obtenidos en modelos de redes neuronales varían entre el 5 y el 10%, mientras que en los modelos de regresión van de entre 10 a 15%. Esto se explica por la flexibilidad que las redes neuronales a reconocer las relaciones en la mayor parte de veces no lineales, entre los patrones de los datos.
10. Entre los usos posibles de la metodología analizada de Redes Neuronales Artificiales para valoración inmobiliaria cabe señalar los siguientes:
 - Posibles aplicaciones al campo de la teoría urbana, ya que nos permite analizar las diferencias de precio del bien vivienda en un área metropolitana por su distancia al centro ciudad –ciudades monocéntrico- o por la distancia a distintos centros de actividad económica –ciudades metacéntricas.
 - Utilización por los propios Agentes de la Propiedad Inmobiliaria como apoyo para ajustar de una forma relativamente fácil, rápida y barata el precio de la vivienda que desea ser vendida por un cliente.

- Los resultados también podrían ser tenidos en cuenta por los constructores de vivienda del ámbito en que se realiza el estudio, que habrán de prestar una mayor atención a aquellos atributos de mayor relevancia para la determinación del precio del inmueble.
- Finalmente, si el mismo Estado promovieran este tipo de investigaciones podrían realizarse con cierta facilidad comparaciones, tanto a nivel espacial como temporal, del precio de las transacciones inmobiliarias y de los principales determinantes del mismo.

Recomendaciones

1. Dependiendo las características de cada municipio sería conveniente incorporar más información de entrada para la estimación del precio de un inmueble, por ejemplo, sobre la existencia de zonas de verdes, facilidad de aparcamiento, afectación por riesgos ambientales o nivel de contaminación en la zona de ubicación, u otras variables que afecten directamente el valor del inmueble.
2. Construir modelos para viviendas regidos por reglamentos de propiedad horizontal.
3. Analizar la posibilidad de otros modelos con diferentes variables explicativas, sobre todo en términos de variables cualitativas.
4. Aplicación de la lógica difusa en las redes neuronales a fin de permitir un tratamiento diferente de la variable ubicación de la propiedad.

Anexo B. Ficha técnica de Ortofotomosaico



ORTOFOTOMOSAICO PLANCHA 227 1:10000

1. INFORMACIÓN GENERAL

El ortofotomosaico de Plancha 227 es entregado en formato .Jmg, con un área total de 194102.04 Ha, para un cubrimiento aproximado de 57 hojas (213750 Ha) de acuerdo a la grilla oficial para escala 1:10000 IGAC

Resolución Espacial: 0.50 m

Resolución Espectral: 3 Bandas

2. CARACTERÍSTICAS GENERALES

El ortofotomosaico de Plancha 227 es de Nivel 2 (Convencional) en la cual la precisión es similar para cartografía de la misma escala, siempre y cuando se realicen mediciones a nivel de suelo. En razón de que la ortorrectificación se realiza teniendo como base un modelo digital de terreno, la proyección de los elementos sobresalientes del suelo no se logra corregir completamente, debido a ello las partes altas de las construcciones pueden estar desplazadas de su localización verdadera, adicionalmente se presentan ocultamientos dificultando la superposición de información vectorial.

3. INFORMACIÓN DEL SISTEMA DE PROYECCIÓN

Coordenadas planas con proyección Gauss Krugger, referidas al datum Magna-Sirgas, elipsoide GRS1980, origen Bogota_Magna al cual se le han asignado los siguientes parámetros:

Latitud: 4°35'46.3215" N

Longitud: 74°04'39.0285" W

Falso Norte: 1'000.000 m

Falso Este: 1'000.000 m

El sistema de proyección se asignó mediante el software ArcGIS. Se recomienda reasignar el sistema de proyección de acuerdo al software en el que se desee visualizar el ortofotomosaico.

4. INFORMACIÓN DE LAS IMÁGENES FUENTE

SENSOR: Vexcel Ultracam D

VUELO: 127 0 30 04 12 2009

Fecha de Toma: 4 de Diciembre de 2009

Fotografías: 347 a 378 381 a 407 452 a 476

VUELO: 127 0 30 15 11 2009

Fecha de Toma: 15 de Noviembre de 2009

Fotografías: 1107 a 1186 1189 a 1226 1228 a 1267

VUELO: 127 0 30 25 11 2009

Fecha de Toma: 25 de Noviembre de 2009

Fotografías: 867 a 887 915 a 940 954 a 989 992 a 1027 1052 a 1087 1090 a 1124 1131 a 1165 1168 a 1238

VUELO: 127 0 30 27 11 2009

Fecha de Toma: 27 de Noviembre de 2009

Fotografías: 2013 a 2059 2070 a 2072 2076 a 2116 2119 a 2159

VUELO: 227 0 30 22 08 2009

Fecha de Toma: 22 de Agosto de 2009

Fotografías: 8 a 27 33 a 47 52 a 54 57 a 103 106 a 151 155 a 201 204 a 250 253 a 255 272 a 299

VUELO: 227 0 30 22 11 2009

Fecha de Toma: 22 de Noviembre de 2009

Fotografías: 346 a 377 777 a 791

Bibliografía

- Ley 44 de 1990, Impuesto Predial Unificado (1990).
- Alonso, W. (1964). *Location and Land Use*. Harvard University Press. Cambridge, 12-27.
- Anas, A. (1983). Residential localization markets and urban transportation economic theory, econometrics, and discrete choice models. *Academic Press*.
- Azoff, E. (1995). *Neural Network Time Series Forecasting of Financial Markets*. New York, USA.: John Wiley & Sons.
- Azqueta, D. (1994). La Función de Precios Hedónicos en la Práctica. 100 -116. McGraw-Hill.
- Bailly, A. S. (1978). *La organización Urbana. Teoría y Modelos*. Madrid: Instituto de Estudios de Administración Local. Col Nuevo Urbanismo No 28.
- Basher, I. A., & Hajmeer, M. (2000). *Journal Microbiological Methods*, 43:, 3.
- Boleat, M. (1973). Housing prices and the housing Market. Housing Market. *Housing Review*, 15-36.
- Bonilla, M., & Puertas, R. (1997). Análisis Redes Neuronales: aplicación a problemas de predicción y clasificación financiera. Quaderns de treball 43. Valencia: Departamento de Economía Financiera y Matemática. Universidad de Valencia.
- Borst, R. (1995). Artificial neural networks in mass appraisal. *Journal of Property Tax Assessment & Administración* 1(2), págs. 5-15.
- Borst, R. A., & Cluskey, M. (1996). The Role of Artificial Neural Networks in the Mass Appraisal of Real Estate. *Paper presented to the European Real Estate Society Conference, Belfast, June 26-28*.
- Boser, B., Guyon, I., & Vapnik, V. (1992). *A training algorithm for optimal margin classifiers*. San Mateo, C.A.:Morgan Kaufmann: In Fifth Annual Workshop on Computational Learning Theory.
- Bourassa, S. F. (1997). Defining residential Submarkets: Evidence from Sydney and Melbourne. Sydney: University of Auckland, Department of Property.
- Bourassa, S., Hamelink, F., M, H., & MacGregor. (1997). Defining residential Submarkets: Evidence from Sydney and Melbourne. Sydney: University of Auckland, Department of Property.

-
- Brío, B. M., & Molina, A. S. (2001). *Redes neuronales y sistemas barrosos*. Madrid: 2ª edición. Rama Editorial.
- Broomhead, D., & Lowe, D. (1988). *Multivariable functional interpolation and adaptive networks*. *Complex Systems*, 2:321-355.
- Cano Guervós, R. A. (1999). Adecuación y aplicación de técnicas cuantitativas análisis del valor de la vivienda. Granada: Tesis doctoral. Editorial Universidad de Granada.
- Caridad, J. M., & Ceular, N. (2001). Un análisis del mercado de la vivienda a través de sistemas de Redes Neuronales. *Revista de Estudios de Economía Aplicada*, 18, 67-81.
- Caridad, J. M., J., N. T., N, C. V., & G, M. V. (2009). 27 th International Conference. *Mathematical Methods in Economics*. Czech Republic.
- Caridad, J., & Brañas, P. (1998). El Precio de la Vivienda Urbana. La disyuntiva Superficie/Ubicación: una ampliación. *Actas del I Congreso de Ceincia Regional de Andalucía:Andalucía en el umbral del siglo XXI*, 792-812.
- Carjal, S. R. (1990). *New Ideas on the Structure of the Nervous System in Man and Vertebrates*. . MIT Press, Cambridge, MA, translation of the French edition of 1894.
- Carpenter, G., & Grossberg, S. (1988). *The ART of Adaptive Pattern Recognition by a Self-Organizing*. *Neural Network. Computer*, 21:77-88.
- Carvalho, J. M. (1999). Modelização e Caracterização do Mercado Imobiliário Português. *Tese de mestrado*. Portugal: Universidade Católica Portuguesa, Faculdade de Ciências Económicas e Empresariais.
- Chica Olmo, J. (1994). Teoría de las variables regionalizadas: aplicación en economía espacial y valoración inmobiliaria. Granada: Universidad de Granada. Servicio de Publicaciones .
- Cooke, W., & Hamilton, B. (1984). Evolution of Urban housing stocks; a model del applied to Baltimore and Houston. *Journal of Urban Economic*, 13, 3, 317 - 38.
- Cortez, P. A. (1997). Algoritmos Genéticos e Redes Neurais na Previsão de Séries Temporais. Minho: Tese de Mestrado em Informática, Universidade do Minho.
- Costa, P. T. (2003). Uma análise do consumo de energia em transportes nas cidades portuguesas utilizando Redes Neurais Artificiais. Tese mestrado em Engenharia Municipal. . Minho: Departamento de Engenharia Civil.Escola de Engenharia de Universidade do Minho.

- Couto, P. M. (2007). Avaliação Patrimonial de Imóveis para Habitação. *Tese de Doutorado*. Portugal: Laboratório Nacional de Engenharia Civil.
- Cybenko, G. (1989). *Approximation by superposition of a sigmoid function, Mathematics of Control, signals and Systems* (Vol. Vol 2).
- Derycke, P. (1983). *Economía y Planificación Urbana*. Madrid: Instituto de Estudios de Administración Local. Col. Nuevo Urbanismo No 38.
- Do, A. Q., & Grudnitski, G. (1993). A neural network analysis of the effect of age on housing values. *Real Estate Residential*. 8 (2), 253-264.
- Do, Q., & Grudnitski, G. (1992). A NeuralNetwork Approach to Residential Appraisal. *The Real Estate Appraiser*. 58, 38-45.
- Fernández, D. L., Llorca, P., Valero, C. S., & Botti, N. (2012). Incidencia de la localización en el precio de la vivienda a través de un modelo de red neuronal artificial. Una aplicación a la ciudad de Valencia. *CT Catastro*, 7-22.
- Follain, J., & Jiménez, E. (1985 a). Estimating the Demand for Housing Characteristics: A Survey and Critique. 15: 77-107. *Regionale Science and Urban Economics*.
- Freeman, J., & Skapura, M. D. (1993). Redes Neuronales Algoritmos, aplicaciones y técnicas de Programación. 26 - 241. Wilminton Addison - Wesley Díaz de Santos.
- Gallant, S. (1993). *Network Learning and Expert Systems*. Cambridge, USA.: MIT Press.
- Gallego, M. E. (2004). La inteligencia artificial aplicada a la valoración de inmuebles. Un ejemplo para valorar Madrid. *CT:Catastro no 50. Abril*.
- García, N., Gámez, M. G., & Alfaro, E. (2008). *ANN+GIS: An automated system for property valuation*. Elsevier, Neurocomputing 71.
- Goldstein, G., & Moses, L. (1973). A Survey of Urban Economics. *Journal of Economic Literature*, 11: 54-96.
- González, M. a., & Formoso, C. T. (2000). Análise conceitual das dificuldades na determinação de modelos de formação de preços através de regressão. *Revista Engenharia Civil da Universidade do Minho*, 8., 65-75.
- Gonzalo, L. M. (1987). *Inteligencia Humana e Inteligencia Artificial*. MADRID: ISBN 84-7118-490-7.
- Harrison, D., & Klein, J. F. (1969). Cumulative urban Growth and urban density functions. *Journal of Urban Economics*, 4(1), 113-117.

-
- Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. Prentice-Hall. New Jersey, USA: 2.^a Edição.
- Henderson, J. (1985). Economic Theory and the cities. *Academic Press*. Orlando.
- Henderson, J. (1985). Economic theory and zoning in urban Economy: Further Results. *American Economic Review*, 73, 98-113.
- Hilera, J. R., & Martínez, V. J. (1995). *Redes Neuronales Artificiales: Fundamentos, modelos y aplicaciones*. Madrid: Ra-Ma.
- Hochman, O., & Pines, D. (1982). Cost of Adjustyment and the spatial pattern of a growing open city. *Econometrica* 50, (6), 1371 - 1391.
- Hopfield, J. J. (1982). *Neural Networks and Physical Systems with Emergent Computational Abilities*. Washington, USA: Proceedings of the National Academy of Sciences, Vol. 79.
- Huawang, S. (2009). Determination of Real estate Price Based on Principal Component Analysis and Artificial Neural Networks. . Second International Conference on Intelligent Computación Technology and Automation.
- IGAC. (Diciembre de 2010). Manual para la Elaboración del Estudio de Zonas Homogéneas Físicas y Geoeconomicas y determinación del valor unitario por tipo de construcción y . Bogotá: IGAC.
- J., F., & M., S. D. (1993). Redes Neuronales Algoritmos, aplicaciones y técnicas de Programación. 26 - 241. Wilmington Addiso - Wesley Díaz de Santos.
- Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). *Artificial Neural Networks: A Tutorial*. *Computers JEEE*, 31.
- James, H., & Lam, E. (1996). The Realiability of artificial Neural Networks for Property data Analysis. *Paper presented to the third European Real Estate Society conference, june* , 26-28.
- Journel, A. (1975). *Guide Pratique de Geostatistique miniere*. París.
- Journel, A. (1977). *Géoestatisque Minere*. París: Thése Doctteur. CGMM.
- Kauko, T. (2003). On current neural network applications involving spatial modelling. *Journal of Housing and Built enviroment*, 18 (2), 159-181.
- Kauko, T. (2009). *Housing Studies*, Vol. 24, No. 5.

- Kershaw, P., & Rossini, P. (1999). Using Neural networks to Estimate Constant Quality House Price Indices. *Fifth Annual Pacific-Rim Real Estate Society Conference. Kuala Lumpur, Malaysia.*
- Kiel, K. A., & Zabel, J. E. (2008). Location, Location, location: The 3L Approach to house price determination. *Journal of Housing Economics* 17, 175-190.
- Kohonen, T. ((1998): The self-organizing map.). *The self-organizing map. Elsevier. Neurocomputing* 21: 1-6. Elsevier. *Neurocomputing* 21: 1-6.
- Kohonen, T. (1997). *Self-Organizing Maps*. Springer, Heidelberg: Series in Information Sciences. 2nd ed., vol. 30.
- Kohonen, T. (1997). *Self-Organizing Maps*. Springer, Heidelberg: Series in Information Sciences. 2nd ed vol. 30.
- Kontrimas, V., & Verikas, A. (2009). *The mass appraisal of the real estate by computational intelligence*. Elsevier. *Journal Applied Soft Computing*.
- Kusan, H., Aytekin, & Ozdemir, I. (2010). *The use of fuzzy logic in predicting house selling price.*, 1808 - 1813. Elsevier. *Expert Systems with Applications* 37.
- Lara Cabeza, J. (2005). Aplicación de las Redes Neuronales Artificiales al Campo de la Valoración Inmobiliaria. *Mapping, 104.*, 64-71.
- Lee, J. S., & Li, M. (1989). *The impact of detention basin on residential property value: Case studies using*. Elsevier. *Landscape and Urban Planning* 89.
- Ley 14, 1. (1983). Normas sobre Catastro, Impuesto Predial e Impuesto de Renta y Complementarios. Bogotá: Congreso de la República de Colombia.
- Ley 1450, 2. (2011). Plan de Desarrollo del Gobierno Nacional "Prosperidad para todos, 2.010 - 2.014".
- Ley 44, 1. (1990). Impuesto Predial Unificado. Bogotá: Congreso de la República de Colombia.
- Ley 489, d. 1. (29 de Diciembre de 1998). *Organización y funcionamiento de las entidades del orden nacional*. Bogotá: Congreso de Colombia.
- Matheron, G. (1969). *Le Krigeage Universal*. París: CGMM Fasc. ENSMP.
- McCluskey, W., & Borst, R. (1997). An evaluation of MRA, comparable sale analysis, and ANNs for the mass appraisal of residential properties in north Ireland. *Assess. J.4(1)*, 47-55.

-
- McCulloch, W., & Pitts, W. (1943). *A logical calculus of the ideas immanent in nervous activity*. Bulletin of Mathematical Biophysics, 5:.
- Mellado, C. (2002). *El Mercado Inmobiliario Urbano en España*. Madrid: Pirámide.
- Minsky, M., & Papert, S. (1969). *Perceptrons: An Introduction to Computational Geometry*. Cambridge: MIT Press.
- Montero Lorenzo, J. M. (2004). El precio medio del metro cuadrado de la vivienda libre: Una aproximación metodológica desde la perspectiva de la geoestadística. *Revista de estudios de economía aplicada, Vol 22-3 (2004)*, 1-18.
- Neves, J., & Cortez, P. (2000). *Redes Neuronais Artificiais*. Braga: Departamento de Informática da Escola de Engenharia da Universidade do Minho.
- Ngugen, N., & Cripps, A. (2001). Predicting housing value: a comparison of multiple regression analysis and artificial neural networks. 314 - 336. *Journal Networks of Real Estate Research*, 22(3).
- Pérez, D. M., & Q., M. M. (2003). *Aplicación de la Redes Neuronales a la Estadística*. Madrid: La Muralla.
- Pérez, D., & Martín, M. (2003). *Aplicación de las Redes Neuronales a la Estadística*. Madrid: La Meralla.
- Pulido, C., Montesinos, P., Roldan, J., & Ruiz - Navarro, F. (2007). Linear regressions and neural approaches to water demand forecasting in irrigation districts telemetry systems. 283 - 293. *Biosystems Engineering* 97.
- Refenes, P. (1995). *Neural Networks in the Capital Markets*. ED Wiley.
- Resolución 620, d. 2. (23 de Septiembre de 2008). Por la cual se establecen los procedimientos para los avalúos ordenados dentro del marco de la Ley 388 de 1997. Bogotá, Colombia: Instituto Geográfico Agustín Codazzi.
- Resolución 70, d. 2. (4 de Febrero de 2011). Reglamenta técnicamente la Formación, la actualización de la Formación y la Conservación Catastral. Bogotá, Colombia: Instituto Geográfico Agustín Codazzi.
- Richardson, H. W. (1973). *Economía Regional. Teoría de la localización, estructuras urbanas y crecimiento regional*. Vicens Vives.
- Riedmiller, M. (1995). Advanced Supervised Learning in Multi-Layer Perceptrons - From. *International Journal of Computer Standards*.

- Rojas, R. (1996). *Neural Networks- A systematic Introducción* Springer-Verlag. Germany.
- Rosenblatt, F. (1959). *Principales of neurodynamics..* New York: Spartan Books.
- Rossini, P. A. (1997). Artificial Neural Networks versus Multiple Regression in the Valuation Residential Property. *Australian Land Economics Review*. V:3No1.
- Rumelhart, D. E., & Hinton, G. E. (1986). *Learning representations by backpropagation. Nature*, N 323.
- Rumelhart, D., Hinton, G., & Clelland, J. M. (1986). *Learning representations by backpropagation. Nature*.
- Saura, P. (1995). Demanda de Características de la Vivienda en Murcia. *Secretariado de Publicaciones de la Universidad de Murcia*, 153-157.
- Shaaf, Mohamad, Erfani, G., & Rod. (1996). Air pollution and the housing market. *Aneural Network Approach*.
- Sharda, R., & Rampal, R. (1996). *Neural Networks and Management Science/Operations Research: A. Encyclopedia of Library and Information Science*, 62.
- Solow, R. (1972). Congestion, density and the use of land in transportation. *Journal of Economics*, 74 , 161 - 173.
- Soria, E., & Blanco, A. (2001). Redes Neuronales Artificiales. *ACTA (Autores científico-técnicos y académicos)*. Id. nº 19.
- Tay, D. P., & Ho, D. K. (1992). Artificial Intelligence and the Mass appraisal of Residential Apartments . *Journal of Property Valuation and investment*. 10:2, 525-40.
- Thurston, J. (2002). *GIS & Artificial Neural Networks: does your GIS Think?* . GIS Vision Magazines.
- Tiebout, C. (1956). A pure theory of local expenditures I. *Journal of Political Economy*, 64, 416 - 535.
- Turnbull, G. K. (1990). The pure theory of Household, localtion: An Axiomatic Approach. *Journal of regional Science*, 30, (4), 549-562.
- Wheaton, W. C. (1982). Urban Spatial Development wiht durable but replaceable capital. *Journalof Urban Economics*, 12 ,63 - 67., 12, 63 - 67 .
- Wilkinson, R. K. (1973). *Measuring the deterninants of relative houses prices*. Environment and Planning, 5.

- Windrow, B. (1962). *Generalization and Information Storage in Networks of ADELIN*
Neurons. In: Self-Organization Systems. Washington: Spartan Books.
- Worzola, E., Lenk, M., & Silva, A. (1995). An exploration of neural networks and its
application to real estate valuation . *Journal of Real Estate Research*, 10(2), 185-
202.
- Zhang, G. P. (1998). Forecasting with artificial networks: The state of the art. *International
Journal of Forecasting*, 14. 35-62. Elsevier.