

UNIVERSIDAD AUTÓNOMA DE CHIHUAHUA

FACULTAD DE INGENIERÍA

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO



**COMPARACIÓN DE LOS MÉTODOS HEDÓNICO Y
LENGUAJE MÁQUINA EN VIVIENDAS DE CIUDAD
CHIHUAHUA**

POR:

ING. MARÍA DE LOURDES FLORES PORTILLO

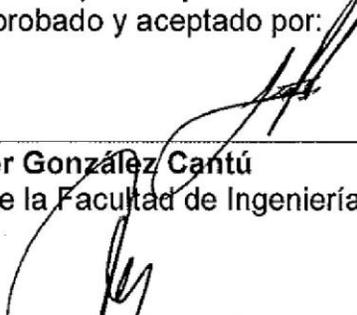
**TESIS PRESENTADA COMO REQUISITO PARA OBTENER EL GRADO DE
ESPECIALIDAD EN VALUACIÓN**

CHIHUAHUA, CHIH., MÉXICO

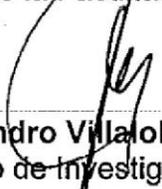
DICIEMBRE 2020



Comparación de los métodos Hedónico y Aprendizaje Máquina en viviendas en Ciudad Chihuahua. Tesis presentado por María de Lourdes Flores Portillo como requisito parcial para obtener el grado de Especialista en Valuación, ha sido aprobado y aceptado por:



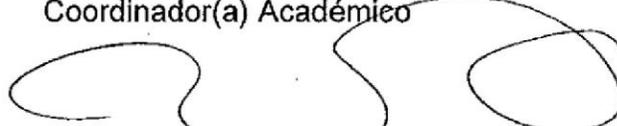
M.I. Javier González Cantú
Director de la Facultad de Ingeniería



Dr. Alejandro Villalobos Aragón
Secretario de Investigación y Posgrado



M.I. Miguel Ángel González Núñez
Coordinador(a) Académico



Dr. Abelardo Casas Comadurán
Director de Tesis

Fecha diciembre 2020

Comité:

Dr. Abelardo Casas Comadurán
M.V. Alonso Villa Ortiz
M.I. Roberto López Santillán
M.I. David Maloof Flores

© Derechos Reservados

**Portillo Ing. María de Lourdes
Flores**

Dirección personal o de la
institución

Chihuahua, Chih. México

Octubre 2020



ING. MARÍA DE LOURDES FLORES PORTILLO

Presente

En atención a su solicitud relativa al trabajo de tesis para obtener el grado de Especialista en Valuación, nos es grato transcribirle el tema aprobado por esta Dirección, propuesto y dirigido por el director **Dr. Abelardo Casas Comadurán** para que lo desarrolle como tesis, con el título: **"COMPARACIÓN DE LOS MÉTODOS HEDÓNICO Y APRENDIZAJE MÁQUINA EN VIVIENDAS DE CIUDAD CHIHUAHUA"**.

Índice de Contenido

RESUMEN

ABSTRACT

ÍNDICE GENERAL

ÍNDICE DE CUADROS

ÍNDICE DE FIGURAS

1 INTRODUCCIÓN

- 1.1 Antecedentes
- 1.2 Problema de investigación
- 1.3 Objetivos de estudio
- 1.4 Justificación
- 1.5 Viabilidad
- 1.6 Delimitación del estudio
- 1.7 Formulación de la hipótesis

2 FUNDAMENTACIÓN TEÓRICA

- 2.1 Marco conceptual
- 2.2 Marco teórico

3 CRITERIOS METODOLÓGICOS

- 3.1 Tipo de investigación
- 3.2 Hipótesis de investigación
- 3.3 Enfoque de la investigación
- 3.4 Tipo de investigación
- 3.5 Diseño de la investigación



UNIVERSIDAD AUTÓNOMA DE CHIHUAHUA

3.6 El marco muestral
3.7 La unidad de análisis

- 3.8 Análisis de regresión múltiple (MPH)
- 3.9 Algoritmo de vectores de soporte de máquina (Support vector machine)
- 3.10 Desarrollo de algoritmo de soporte de vectores de regresión (support vector regression)
- 3.11 Desarrollo de algoritmo de Bosques aleatorios (Random forest clasification)
- 3.12 Desarrollo de algoritmo de Perceptron Multicapa (MPL Multi Layer Perceptron)

4 RESULTADOS Y DISCUSIÓN

Resultados

- 4.1 Selección de la ecuación de MPH
- 4.2 Resultado de vectores de soporte de regresión (Support vector regression)
- 4.3 Resultado de Bosques aleatorios (Random Forest)
- 4.4 Resultado de Multicapa Perceptron
- 4.5 Discusión

5 CONCLUSIONES Y RECOMENDACIONES

- 5.1 Conclusiones
- 5.2 Recomendaciones

6 BIBLIOGRAFÍA O REFERENCIAS BIBLIOGRÁFICAS

7 ANEXOS

ATENTAMENTE
"naturam subiecit aliis"

EL DIRECTOR

M.I. JAVIER GONZÁLEZ CANTÚ

EL SECRETARIO DE INVESTIGACIÓN
Y POSGRADO

DR. ALEJANDRO VILLALOBOS ARAGÓN

DEDICATORIA

Este trabajo de investigación lo dedico a mi familia, quienes han sido parte importante en el desarrollo de este proceso.

A mi esposo Mario, mis hijos Eunice, Mario Rubén y Rubí, a mi madre Eva.

Gracias por todo el soporte durante el estudio y la elaboración de esta investigación, los amo y son mi motor para seguir adelante.

Al Dr. Abelardo Casas Comadurán, quien me ha apoyado, poniendo la inquietud de esta investigación y por todo su apoyo.

Al único y sabio Dios a quien me ha dado la vida y los medios para seguir adelante, a El sea toda la Gloria.

AGRADECIMIENTOS

Todo mi reconocimiento a los Maestros y personal de la Universidad Autónoma de Chihuahua, a mis maestros de la Especialidad en Valuación a quienes nombro a continuación:

Dr. Abelardo Casas Comadurán, un incansable investigador y científico que se ha destacado por su anhelo de nuevas técnicas y conocimiento en el área de la valuación.

M.V. Alonso Villa Ortiz, que por el amor a su profesión se entrega como un guía para los estudiantes que hemos sido parte de sus clases.

M.I. Jesús Roberto López Santillán, quien me apoyo de una manera extraordinaria para continuar con esta tesis.

M.I. David Maloof Flores compañero y amigo gracias por tu apoyo.

Dr. Alejandro Villalobos Aragón compañero y amigo de la Facultad gran ejemplo de responsabilidad y dedicación.

Giovanny Alfonso Chávez Cenicerros por el apoyo en el proceso de los algoritmos de lenguaje de máquina.

RESUMEN

El objetivo del trabajo fue analizar cuatro métodos de valuación automática asistida por computadora, para esta investigación se abordaron los modelos de valuación por Precios Hedónicos, Support Vector Regression, Random Forest y el de Valuación por Redes Neuronales, la meta que se alcanzó es un análisis de la validez estadística de estos modelos con el objeto de conocer cuál de los cuatro algoritmos estadísticos alcanzó el resultado óptimo. En la actualidad los modelos automáticos que integran la inteligencia artificial para la solución de problemas han sido muy socorridos en varias áreas de la ciencia y la tecnología, de esta forma el manejo de métodos novedosos en valuación se suma como un potencial a descubrir y agilizar procesos de valuación masiva, de tal forma que los profesionales en valuación podrán alcanzar sus metas de precisión y objetividad en los avalúos que se realicen en corto tiempo. El resultado encontrado consiste en evidencia de la validez estadística de los cuatro algoritmos estudiados, identificada por su capacidad de predicción del valor de mercado, medida esta en función de la menor desviación estándar asociada a un coeficiente de determinación R^2 (ajustada) mayor del 95 %. Se concluyó que en orden de mayor a menor validez se encuentran los algoritmos: Regression, Random Forest (98.86 %), Support Vector Regression (98.70 %) y el Método de Precios Hedónicos (96.90 %) Un algoritmo que no superó el objetivo del 95 %, fue el de las Redes Neuronales Artificiales (83.06 %).

Palabras clave: (Método de precios hedónico, redes neuronales artificiales, aprendizaje de máquina, valuación automática de inmuebles).

ABSTRACT

The objective of the work was to analyze four methods of computer-assisted automatic valuation, for this research the valuation models by Hedonic Prices, Support Vector Regression, Random Forest and Valuation by Neural Networks were approached, the goal that was reached is an analysis of the statistical validity of these models in order to know which of the four statistical algorithms achieved the optimal result. At present, automatic models that integrate artificial intelligence for problem solving have been very helpful in various areas of science and technology, in this way the management of novel valuation methods is added as a potential to discover and streamline processes of massive valuation, in such a way that valuation professionals will be able to achieve their goals of precision and objectivity in the appraisals that will be carried out in a short time. The result found consists of evidence of the statistical validity of the four algorithms studied, identified by their ability to predict the market value, measured as a function of the lowest standard deviation associated with a coefficient of determination R^2 (adjusted) greater than 95%. . It was concluded that in order from highest to lowest validity are the algorithms: Regression, Random Forest (98.86%), Support Vector Regression (98.70%) and the Hedonic Price Method (96.90%). An algorithm that did not exceed the target of 95 %, was that of Artificial Neural Networks (83.06%).

Keywords: (Hedonic pricing method, artificial neural networks, machine learning, automatic property valuation).



ÍNDICE GENERAL

RESUMEN	VII
ABSTRACT	VIII
ÍNDICE GENERAL.....	IX
1 Índice de Tablas.....	XII
2 Índice de figuras	XIII
1 INTRODUCCIÓN	1
1.1 ANTECEDENTES	4
1.2 PROBLEMA DE INVESTIGACIÓN	9
1.3 OBJETIVOS DEL ESTUDIO.....	9
1.3.1 Objetivo general.....	9
1.4 JUSTIFICACIÓN.....	10
Justificación	10
1.5 Viabilidad.....	11
1.6 DELIMITACIÓN DEL ESTUDIO	11
Delimitación	11
1.7 FORMULACIÓN DE LA HIPÓTESIS.....	11
2 FUNDAMENTACIÓN TEÓRICA	11
2.1 Marco conceptual	11
2.1.1 Enfoque de mercado (o comparativo)	12
2.1.2 Enfoque físico o de costos.....	13
2.1.3 Enfoque residual	13
2.1.4 Homologación	13
2.1.5 Propósitos en un trabajo de valuación.....	13
2.1.6 Metodología de mercado.....	14
2.1.7 Valuación.....	14
2.1.8 Valor comercial.....	14
2.1.9 Regresión múltiple	14
2.1.10 Método Hedónico	14
2.1.11 Redes neuronales artificiales	15
2.1.12 Algoritmo.....	15
2.1.13 Inteligencia artificial	15



2.1.14	Deep learning o Aprendizaje profundo	15
2.1.15	Aprendizaje de máquina o aprendizaje de automático	16
2.2	Marco teórico	17
2.2.1	Métodos tradicionales	17
2.2.2	Métodos avanzados	17
2.2.3	Metodología de Precios Hedónicos.....	18
2.2.4	Redes Neuronales Artificiales.....	19
2.2.5	Aprendizaje de máquina	20
2.2.6	Arquitectura básica de una red neuronal	21
	Estado del arte	25
2.2.7	Estado del arte de las redes neuronales artificiales.....	25
3	CRITERIOS METODOLÓGICOS.....	27
3.1	Tipo de investigación.....	27
3.2	Hipótesis de investigación.....	27
3.3	Enfoque de la investigación	27
3.4	Tipo de investigación.....	27
3.5	Diseño de la investigación	27
3.5.1	Transeccional explicativo	28
3.6	El marco muestral	28
3.7	La unidad de análisis	28
3.7.1	El tipo de muestreo	28
3.7.2	El tamaño de la muestra	29
3.7.3	La selección de la muestra	29
3.7.4	Variable predictoria del estudio.....	29
3.7.5	Codificación de las variables independientes	29
3.7.6	Medidas estadísticas básica	33
3.7.7	Prueba de hipótesis t de student	33
3.8	Análisis de regresión múltiple (MPH).....	33
3.9	Algoritmo de vectores de soporte de máquina (Support vector machine)	34
3.10	Desarrollo de algoritmo de soporte de vectores de regresión (support vector regression) 34	
3.11	Desarrollo de algoritmo de Arboles de decisión (Random forest).....	35
3.12	Desarrollo de algoritmo de Redes Neuronales artificiales, Perceptrón Multicapa (MPL Multi Layer Perceptron)	37
4	RESULTADOS Y DISCUSIÓN.....	37



Resultados	37
4.1 Selección de la ecuación de MPH.....	37
4.2 Resultado de vectores de soporte de regresión (Support vector regression)	38
4.3 Resultado de Bosques aleatorios (Random Forest)	39
4.4 Resultado de Redes neuronales artificiales (Multicapa Perceptron).....	39
4.5 Discusión	39
5 CONCLUSIONES Y RECOMENDACIONES.....	41
5.1 CONCLUSIONES:	41
5.2 RECOMENDACIONES	42
6 BIBLIOGRAFÍA O REFERENCIAS	43
7 ANEXOS:	46
7.1 ANEXO A CEDULA DE DATOS DEL INMUMUEBLE.....	46
7.2 ANEXO B. CEDULA PARA CALCULAR LOS RESULTADOS DE LOS EJEMPLOS DE APLICACIÓN DE LOS CINCO CONJUNTOS TESTIGOS	¡Error! Marcador no definido. 6
7.3 ANEXO C. ECUACIONES MPH	46
7.4 ANEXO D PRUEBA T DE STUDENT DE LAS MPH.....	¡Error! Marcador no definido. 50
7.5 ANEXO E LENGUAJE DE MÁQUINA.....	51
1 RF vs SVM vs MLP	51
1 Cargando el dataset	52
1.1 Métodos para preparar los datos según las necesidades.....	55
1.2 Visualizar el pipeline (opcional)	56
(A) Support Vector Machine	58
(B) Random Forest	59
(C) Multi-Layer Perceptron	59



Índice de Tablas

Tabla 1.	Nivel de Infraestructura	30
Tabla 2.	Nivel de equipamiento urbano	30
Tabla 3.	Clasificación de la zona.....	30
Tabla 4.	Referencia de proximidad	30
Tabla 5.	Calidad predominante en las calles en las calles circundantes.....	31
Tabla 6.	Nivel socioeconómico	31
Tabla 7.	Referencia de proximidad urbana.....	31
Tabla 8.	Calidad predominante.....	31
Tabla 9.	Clase general del inmueble	32
Tabla 10.	Estado de conservación.....	32
Tabla 11.	Estadística.....	33
Tabla 12.	Resultados obtenidos en los valores de desviación estandar y coeficiente de determinación R2 (ajustada).....	40



Índice de figuras

Figura 1.	Perceptrón sin bias.....	22
Figura 2.	Perceptrón con bias.....	23
Figura 3.	Red neuronal simple.....	24
Figura 4.	Support vector machine.....	35
Figura 5.	Bosque de decisión.....	36
Figura 6.	Perceptrón multicapa.....	37

1 INTRODUCCIÓN

En la nueva economía global la Inteligencia artificial se ha convertido en un tema central, aunque la inteligencia artificial existe desde hace más de 50 años, actualmente los avances tecnológicos, big data, y nuevos algoritmos han hecho que la inteligencia artificial se considere un elemento fundamental para la transformación digital de la sociedad.

Los avances recientes en la inteligencia artificial han aumentado la necesidad de transformar a la sociedad digitalmente. Las actividades diarias como la publicidad en línea, búsquedas en la web, asistentes personales digitales, ciudades, infraestructura, automóviles inteligentes, ciberseguridad, transporte, fabricación, alimentación y agricultura. La administración y los servicios públicos, un sin número de actividades económicas en donde interviene la inteligencia artificial, para hacer las tareas cotidianas de manera más eficiente y en un menor tiempo.

Una importante cuestión técnica es implementar en la disciplina de la valuación la aplicación de la inteligencia artificial, a través de nuevos algoritmos en los cuales se apliquen nuevas técnicas o modelos automáticos para la valuación.

Sin embargo, hasta ahora a nivel nacional ha habido poca información en el área de la valuación y su aplicación.

La aplicación de valores a la tierra, o a los bienes inmuebles, es una actividad que se realizó desde tiempos antiguos. La valuación es un procedimiento técnico basado en metodologías específicas por las cuales se realiza una investigación del inmueble, del entorno social y económico, para de esa manera establecer variables cuantitativas y cualitativas y dar un valor a los bienes. (Hernández, 2017)

La valuación en todos los activos es necesaria. En la sociedad es imprescindible saber el valor de los activos para realizar trámites o transacciones económicas.

Algunos ejemplos son expropiaciones gubernamentales, impuestos prediales, compra venta, créditos hipotecarios entre otros. (Preciado, 2015)

El propósito de esta investigación fue mostrar los modelos automáticos de la valuación a través de cuatro métodos, El Modelo de Precios Hedónicos, Support Vector Regression, Random Forest y las Redes Neuronales Artificiales. Con el fin de analizar cuál de los cuatro métodos es el óptimo y tiene mejores resultados, en viviendas en la ciudad de Chihuahua.

Las cuestiones principales tratadas en esta investigación se plasmaron de la siguiente manera.

En el capítulo uno se expresa los antecedentes de la investigación sobre los métodos automáticos de la valuación. En Sudamérica, México y países europeos. Las redes neuronales y la aplicación de la inteligencia artificial ya se han trabajado en la disciplina de la valuación, las tesis de investigación muestran la inquietud por ingresar esta disciplina en el área de la inteligencia artificial.

El problema de investigación fue la propuesta de mostrar los métodos automáticos como una alternativa viable para los profesionales de la valuación. Mostrando que dichas aplicaciones automáticas apoyan al desarrollo de la profesión de manera más eficiente y eficaz, sin sustituir al valuador, sino apoyándolo en las nuevas tecnologías inteligentes.

Objetivo de la investigación: Comparar los precios de predicción del valor del mercado obtenidos por cuatro algoritmos: Método de Precios Hedónicos, Support Vector Regression, Random Forest, y Redes Neuronales Artificiales en las viviendas de calidad, interés social, media, semi lujo y residencial de la Ciudad de Chihuahua.

Justificación de la investigación: La investigación se realizó con el propósito de conocer otros métodos de valuación utilizando la estadística y aprendizaje de máquina, para optimizar la valuación inmobiliaria.

El proyecto contó con factibilidad, ya que se tiene la información necesaria, es decir, se tomaron muestras representativas de cada colonia, siendo como máximo seis, para realizar la investigación, el Método de Precios Hedónicos, Support Vector Regression, Random Forest y las redes neuronales.

El estudio se realizó en la ciudad de Chihuahua contando con una base de datos elaborada por estudiantes de la Especialidad en Valuación en el año 2015 y consta de 1,904 avalúos de viviendas de clase: interés social, media, semilujo, residencial.

La formulación de la hipótesis: El método de valuación que aplica redes neuronales artificiales produce resultados con mayor precisión que los métodos de precios hedónicos, vectores de soporte de regresión (Support Vector Regression) y bosques aleatorios (Random Forest).

El capítulo dos la fundamentación teórica se mostró las investigaciones más recientes en la aplicación de las redes neuronales artificiales y el método hedónico en algunos estudios de valuación. Así también se describieron los conceptos de los métodos tradicionalmente aceptados en el gremio de los valuadores. Como el enfoque físico, de mercado, residual.

Se describió los conceptos de inteligencia artificial, aprendizaje de máquina y algoritmo. Así como la arquitectura básica de una red neuronal, support vector regression, random forest y la metodología de precios hedónicos.

En el capítulo tres se muestra el método que se realizó para producir las ecuaciones que predecirán el valor de mercado neto.

Se contó con la base de datos de los avalúos de la Ciudad de Chihuahua. Dicha base se configuró en partes, en cada una de ellas se seleccionaba el 20% de la base para esa selección tomarla como testigo de prueba después de realizar los cálculos estadísticos con el fin de determinar la ecuación de regresión.

Dicho procedimiento se repitió en cinco ocasiones para aplicar el Método de Precios Hedónicos, obteniendo cinco polinomios para predecir el valor de mercado. Acto seguido se procedió a seleccionar la ecuación que produjo la menor desviación estándar de la diferencia de su predicción del valor de mercado respecto de los correspondientes valores de mercado contenidos en su testigo de prueba.

La ecuación así obtenida se comparó con los valores de la desviación estándar que resultaron de las predicciones del valor de mercado derivadas de los algoritmos correspondientes a las Redes Neuronales Artificiales, Support Vector Regression y Random Forest. La prueba de hipótesis se realizó por comparación de los valores de la desviación estándar, seleccionando como óptima a la que produjo la menor desviación estándar.

1.1 ANTECEDENTES

“La valuación de inmuebles es un procedimiento técnico basado en metodologías específicas que, a través de la indagación física, económica y social, permite establecer las variables cuantitativas que inciden en el valor que se le asigna a los bienes”. (Hernández, 2017)

“El objetivo del valuador es reflejar, con la mayor objetividad y certeza posibles, el valor que el mercado otorga a los bienes en un determinado del tiempo. Para lo cual, deberá obtener la información disponible sobre los precios o los bienes comparables y utilizará la metodología adecuada para dar el valor que tenga el bien cuya valoración se le ha encargado”. (AEV Asociación Española de Análisis de Valor, 2020).

El uso de un avalúo inmobiliario generalmente es para un préstamo bancario, es por ello por lo que las entidades bancarias solicitan un avalúo del bien a los profesionales capacitados en el área. Así el valor del inmueble sirve como garantía de esto se deriva el valor del préstamo.

Tradicionalmente existen metodologías aceptadas entre el gremio de los valuadores con las que se realizan los avalúos de los bienes inmuebles, sin embargo, existen nuevas técnicas matemáticas y estadísticas que pueden obtener los valores del mercado de una manera distinta a las tradicionales.

Se han publicado una cantidad considerable de tesis y artículos científicos sobre las redes neuronales artificiales y el modelo de precios hedónicos. Las redes neuronales han tomado la delantera en distintas disciplinas como medicina, química, ingeniería entre otras. La valuación no es la excepción y los investigadores se adentraron en el ámbito de la inteligencia artificial y la regresión lineal para explicar los resultados obtenidos en los valores de viviendas o bienes inmuebles.

En México como en el mundo, se hace más difícil ignorar los métodos automáticos para la elaboración de un avalúo. Los avances más recientes en inteligencia artificial han aumentado la necesidad de utilizar las nuevas tecnologías en cualquier ámbito de la ciencia y en este estudio en la valuación.

Sin embargo, este tema es poco utilizado entre el gremio de los valuadores, trabajando con los métodos tradicionales.

La innovación en los métodos de valuación es prioritaria, las nuevas tecnologías, como la inteligencia artificial, usada en diferentes disciplinas, han apoyado en mejoras significativas.

Las investigaciones recientes en el área de la valuación han mostrado un creciente interés por implementar alternativas analíticas objetivas apoyadas en herramientas estadísticas.

En México cada año se celebra un congreso nacional de la federación de colegios de valuadores (FECOVAL) en donde se aborda problemáticas relacionadas al ámbito de la valuación en el país.

En la ponencia titulada: “UN MODELO AUTOMÁTICO DE VALORACIÓN (AVM) APLICADO EN LIMA – PERÚ, presentada en el congreso nacional de valuación 2018, el autor presentó los objetivos de su investigación fueron: Desarrollar un modelo automático de valoración (AVM) implementado en Lima Perú, el cual se construyó con los estándares europeos; desde la construcción de la base de datos, el cálculo de un modelo matemático y su implementación mediante una interfaz web (BERNALES GHACHAM, 2018).

La metodología utilizada en la investigación fue la creación de una base de datos, el análisis exploratorio de los datos, eligiendo las variables: Precio (en US\$) - Área total del apartamento (m²) - Edad (años) - Número de habitaciones - Número de baños - Número de estacionamientos - Vista (al mar, interior o calle) – Piso. Elaboración del modelo matemático, interfaz avm – usuario, análisis e interpretación de los resultados, modelo de regresión clásico, conclusiones y recomendaciones fueron: Los modelos automáticos de valoración son sumamente útiles y aplicables en diversos campos tanto públicos como privados.

En la Tesis titulada “Valoración de Inmuebles urbanos: Comparativa Modelo Regresión Multivariable versus Redes Neuronales Artificiales”, para la ciudad de Morelia, Michoacán, México. En donde el objetivo es la comparación entre el modelo econométrico (regresión lineal múltiple) y una red neuronal del tipo multicapa back propagation para la estimación y predicción de precios de inmuebles sobre los diferentes segmentos del mercado inmobiliario. (Preciado, 2015)

La hipótesis es que las redes neuronales artificiales, tienen la ventaja de ser capaces de encontrar relaciones más complejas entre las variables y por lo tanto mayor capacidad explicativa de los precios de los inmuebles frente a los modelos clásicos econométricos.

Otro objetivo planteado se relaciona con la elección de un modelo econométrico para la modelización del precio, por medio del modelo de regresión lineal múltiple, el modelo de regresión no lineal y la red multicapa back propagation, con salida lineal y no lineal.

Se expone un estudio para la comparación entre un modelo econométrico y la red neuronal multicapa back propagation para determinar los resultados.

En los resultados se observa que el modelo de red neuronal con 16 unidades en la capa oculta, supera en cada una de sus mediciones estadísticas a la regresión múltiple, logarítmica y al otro modelo de RNA propuesto.

Con base en los resultados, en línea con los resultados de otros investigadores, se concluye que para el caso de México también las redes neuronales son capaces de explicar mejor la variabilidad del precio de los inmuebles que los modelos de regresión múltiple. (Preciado, 2015)

Además (Lara, 2019) en su tesis Validez de una ecuación del método de precios hedónicos para el mercado de vivienda en la ciudad de Chihuahua, Chihuahua. Menciona que el objetivo de su investigación fue: comprobar la validez de la ecuación matemática que se obtuvo mediante un procedimiento de regresión múltiple (MPH).

La metodología utilizada en el estudio proporcionó una comprobación de los modelos soportados por la regresión múltiple en el método de precios hedónicos.

Los resultados obtenidos: se aplicó el método de regresión múltiple para diez conjuntos para calcular la ecuación MPH y se obtuvieron diez ecuaciones que se probaron con los diez conjuntos testigo. De esa manera se obtuvieron diez pruebas de hipótesis mediante la aplicación de la t Student.

Como ya se dijo en las cinco ocasiones que se aplicó la t Student se encontró que la media de la diferencia entre los valores obtenidos con las ecuaciones MPH respecto de los valores de los conjuntos testigo, siempre fue considerablemente menor al 20%. La ecuación seleccionada como la de menor error fue la MPH 4 que se muestra a continuación:

Ecuación de regresión

$$\begin{aligned} Vmn = & 33018 Id - 109185 Cz - 40374 Nr - 167501 Nb + 182155 Sb \\ & + 26180 See - 10971 Tfr \\ & - 5349 Tfo + 124295 Cg + 16506 Ec - 255.1 Ed \\ & + 291.2 Eu_1 + 2703 Ssh + 4668 Sv \\ & - 438651 Cus \end{aligned}$$

Resumen del modelo

	R-	R-cuad.	R-
S	cuad. (ajustado)	(pred)	cuad.
142421	96.94%	96.91%	96.81%

Esta investigación presenta un trabajo sobre los precios de bienes raíces, en donde se usaron modelos de precios hedónicos. Los modelos tienen diversos propósitos, como el cálculo de impuestos de bienes raíces, también los análisis financiero de los bienes inmuebles. Para los casos de impuestos, los modelos deben ser extensos, dirigido a la valorización de un conjunto de inmuebles, lo que hace que la investigación parezca más compleja. Esta investigación aborda dos ciudades brasileñas de tamaño mediano, Lajeado y Montenegro, que demuestran

el proceso de estudio y el desarrollo de modelos de precios de bienes inmuebles. Los datos fueron recabados de agencias inmobiliarias y anuncios comerciales de evaluación. La investigación estadística, dio paso para obtener las ecuaciones que mejor se ajustaban a los datos recolectados. Se realizaron pruebas significativas para las variables y el modelo, de esta forma mostraron que los modelos tienen un buen poder explicativo con respecto a los precios. Se puede determinar que los modelos desarrollados son factibles en valuaciones fiscales.

1.2 PROBLEMA DE INVESTIGACIÓN

El mundo actual evoluciona de manera vertiginosa, la tecnología ha llegado a los hogares de tal forma que ahora contamos con edificios inteligentes, por tanto sería ingenuo dejar de lado la tecnología en los métodos de valuación, así pues evolucionar en conocimientos novedosos para realizar avalúos es una tarea a desarrollar en la investigación de los mejores aprendizajes, utilizando método hedónico, programación genética y otros métodos computacionales, y otros métodos de aprendizaje de máquina (Machine Learning) , a través del método científico para mejorar y optimizar los resultados de los avalúos.

¿Cómo se diseña una muestra de viviendas representativas de los valores de mercado de vivienda de la ciudad de Chihuahua?

¿Cuáles son las características que son relevantes para el precio de mercado de las viviendas de clase: interés social, media, semilujo, ¿residencial?

1.3 OBJETIVOS DEL ESTUDIO

1.3.1 Objetivo general

Comparar los precios de predicción del valor de mercado obtenidos por el Método de Precios Hedónico, con los algoritmos de aprendizaje de máquina conocidos como: Redes Neuronales Artificiales, Vectores de Soporte de Regresión (Support Vector Regression) y Bosques Aleatorios (Random Forest) en las

viviendas de calidad, interés social, media, semi lujo y residencial de la Ciudad de Chihuahua.

1.4 JUSTIFICACIÓN

Justificación

La investigación se realizó con el propósito de conocer otros métodos de valuación utilizando la estadística y aprendizaje de máquina, para optimizar la valuación inmobiliaria. Para determinar la justificación del trabajo se usan los cinco criterios que expresa (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2014) resultando como sigue:

Criterio conveniencia. La investigación sirve para difundir conocimientos de valuación asistida por computadora fundamentados en métodos estadísticos asistidos por computadora

Criterio de relevancia social. La proyección del trabajo está vinculada con problemas genuinos del país. Ello, debido a que puede apoyar la aplicación de valuación masiva en las instituciones que administran los impuestos sobre la propiedad inmobiliaria. Además, son herramientas de avanzada que pueden incorporarse a la praxis de la valuación en el país proporcionando objetividad y certeza en los reportes de valuación que emiten los profesionales del medio.

Criterio de implicaciones prácticas. El uso de los Métodos estadísticos asistidos por computadoras en la valuación de bienes raíces optimizará la práctica profesional, aportando rapidez, objetividad y certeza de los resultados de los profesionales y las entidades del medio.

Criterio de valor teórico. Se aporta información novedosa sobre la capacidad de predicción de los cuatro métodos estudiados, resultando en un análisis comparativo de las ventajas y limitaciones de sus resultados.

1.5 Viabilidad

El proyecto tuvo factibilidad, ya que se contó con la información necesaria para realizar la investigación, el método hedónico y algoritmos de aprendizaje de máquina

1.6 DELIMITACIÓN DEL ESTUDIO

Delimitación

El estudio se realizó en la ciudad de Chihuahua contando con una base de datos elaborada por estudiantes de la Especialidad en Valuación en el año 2015 y consta de 1,904 avalúos de viviendas de clase: interés social, media, semilujo, residencial. La antigüedad de las viviendas que se seleccionaron como sujetos de investigación se ubica en un rango de 0 a 30 años.

1.7 FORMULACIÓN DE LA HIPÓTESIS

Hipótesis General

El método de valuación que aplica redes neuronales artificiales produce resultados con mayor precisión que los algoritmos de: Método de Precios Hedónicos, Vectores de Soporte de Regresión (Support Vector Regression) y Bosques Aleatorios (Random Forest).

2 FUNDAMENTACIÓN TEÓRICA

2.1 Marco conceptual

El objetivo de la valuación inmobiliaria es otorgar el valor comercial de un bien inmueble.

En casos específicos la valuación se puede realizar para otros propósitos (valores referidos, de reproducción o reposición, de rescate, residual, etc.). (Almirall, 2007).

Enfoque de capitalización (o de ingresos): método que considera las rentas las rentas producidas o beneficios generados por el objeto de la valuación. La capitalización de rentas es un proceso por el cual se obtiene el valor de un bien o derecho, con base en la cantidad de las rentas o beneficios netos futuros que se obtienen durante la vida económica del bien, y con una tasa que involucra la productividad y todos los riesgos asociados con el bien o derecho que se trate. (Hernández, 2017)

Modelos de Valuación Automática (MVA). Son programas de cómputo para valuación automática que usan variables de los bienes raíces como la demografía, las características de la propiedad y las de su entorno, así como los precios de ofertas y ventas del mercado para calcular el valor de una propiedad. Los métodos que se utilizan para los MVA se distinguen en dos tipos genéricos: 1) los procesos estadísticos de regresión múltiple que requieren apoyo teórico formal previo que apoye sus conclusiones, conocidos como modelos hedónicos de regresión y; 2) los procesos de regresión múltiple flexible, que no precisan de soporte teórico previo, entre los que destacan las redes neurales. (Casas Comadurán, 2007)

2.1.1 Enfoque de mercado (o comparativo)

El enfoque de mercado está basado en el principio de sustitución, es decir, se basa en el análisis y comparación de los precios de bienes similares al sujeto de estudio en el ámbito de inmuebles, que se encuentran en venta en zonas similares. Homologando los valores obtenidos si existen diferencias entre los inmuebles comparables al sujeto de estudio. Entre las que se encuentran:

- Ubicación.
- Edad.
- Superficie.
- Calidad del proyecto.
- Estado de conservación.

Método que se fundamenta en el empleo de datos estadísticos de transacciones de mercado y razonamientos que reflejan el pensamiento de los participantes de este, utiliza procesos que incluyen la comparación y considera también el uso de listas de precios y ofertas publicadas, denominándose al resultado obtenido “valor de mercado”. (Hernández, 2017)

2.1.2 Enfoque físico o de costos

Método aplicable en la valuación de bienes que se fundamentan en el costo actual de su reproducción o reposición, reduciéndolo por los deméritos propios específicos de los mismos. (Hernández, 2017)

2.1.3 Enfoque residual

Método cuya premisa fundamental establece que es posible determinar el valor de cambio individual de un elemento integrante de una unidad económica, en un momento dado, a partir del valor de uso o de cambio de toda ella, reduciéndolo con el costo neto de reposición propio del complemento de dicha unidad. (Hernández, 2017).

2.1.4 Homologación

Es un proceso que tiene como finalidad, por una parte, la identificación de aquellos elementos físicos que conforman las pequeñas diferencias entre los bienes comercializados en el mercado y aquel que se requiere valorar. Y por la otra cuantificar monetariamente dichas diferencias, con el propósito de conocer el bien en cuestión. (Hernández, 2017).

2.1.5 Propósitos en un trabajo de valuación

Un avalúo se elabora con uno o más propósitos, para determinar el valor de un bien; para estimar el costo de producir, poseer, cambiar o terminar un bien; para estimar el importe de los daños a un bien: y para pronosticar las rentas potenciales que puede generar un bien.

2.1.6 Metodología de mercado

El método de mercado consiste en la estimación de un valor, a partir de un conocimiento general y consensado del mercado y de análisis de la oferta de propiedades parecidas al objeto de tasación. (Hernández, 2017).

2.1.7 Valuación

Es un sustantivo femenino, cuyo significado es conocido como acción y efecto de valorar, por lo que valorar es un verbo transitivo que consiste en señalar la cuantía económica de un intercambio. (Hernández, 2017)

2.1.8 Valor comercial

Es el valor operante en el ámbito del comercio para la realización de un acto jurídico donde hay que establecer el valor de los bienes, servicios, derechos u obligaciones que se involucren en este o para valorar las consecuencias de un hecho jurídico. (Hernández, 2017)

2.1.9 Regresión múltiple

La regresión múltiple es la técnica o herramienta estadística empleada para identificar cuantitativa y cualitativamente los atributos de los inmuebles que producen satisfacción -en el sentido de la doctrina hedónica- a los individuos.

2.1.10 Método Hedónico

La metodología de los precios hedónicos constituye herramienta importante para estudiar las correlaciones que existen entre los niveles de calidad, que puede atribuirse a una unidad habitable, en términos de un conjunto de características y el valor de la misma (ya sea el valor de compra y venta o el valor de la renta que se paga por el derecho de habitarla). Los precios hedónicos están basados en la hipótesis de que los bienes son valorados por el bienestar relacionada con sus características. Los precios hedónicos se definen como los precios implícitos de las características que proporcionan satisfacción y son revelados por los sujetos al ejercer transacciones de mercado, ya que en ellas puede observarse los precios de productos diferenciados, así como las diferentes cantidades de atributos de

bienestar que poseen para resolver las necesidades y deseos individuales del bien (Rosen, 1974). La técnica de los precios implícitos en los atributos de los inmuebles es ya bastante conocida en la práctica económica y pueden distinguirse dos etapas en la utilización de esta, la segunda de ellas a partir de 1974 se caracterizó por un incremento en el refinamiento de la teoría.

2.1.11 Redes neuronales artificiales

Son objetos de programación que imitan el funcionamiento de las neuronas biológicas (Esteban, 2009). Formadas a través de una interconexión de redes, de forma paralela, estas redes poseen una organización jerárquica que permiten interactuar con el mundo, además se conoce como un sistema de computación constituida por un gran número de elementos simples de procesamiento muy interconectados, que procesan información en respuesta para algún estímulo externo. (Wilmer & Olivo, 2018)

2.1.12 Algoritmo

Conjunto ordenado y finito de operaciones que permiten hallar la solución de un problema. Método y notación en las distintas formas del cálculo. (Real academia de la lengua española, 2019). Es el núcleo de donde parte la inteligencia artificial. Son las fórmulas matemáticas o comandos de programación que indican a la computadora como resolver situaciones o problemas.

2.1.13 Inteligencia artificial

Es una ciencia multidisciplinaria que busca desarrollar una inteligencia similar a la humana en algoritmos resueltos mediante mecanismos y/o softwares programados para tomar decisiones o realizar tareas que típicamente requieren de inteligencia humana y que se alimentan por grandes bancos de datos.

2.1.14 Deep learning o Aprendizaje profundo

Son redes neuronales artificiales con varias capas, de modo que su aprendizaje es más completo y menos superficial, pueden aprender de forma automática sin entrenamiento humano y a partir tan solo de reglas.

2.1.15 Aprendizaje de máquina o aprendizaje de automático

Es una rama de la inteligencia artificial basada en la idea de que los sistemas pueden aprender de datos, identificar patrones y tomar decisiones con mínima intervención humana.

2.2 Marco teórico

2.2.1 Métodos tradicionales

Los métodos de valuación tradicionalmente aceptados en el gremio de los valuadores son los técnicos conocidos como: Costos, Comparación del mercado, Capitalización de rentas, Residual - Estático – Dinámico. Cada país establece las normas para las finalidades de los avalúos con los métodos de valuación de inmuebles en función del tipo de inmueble. Las asociaciones internacionales que elaboran normas de valoración entre las cuales se encuentran: Normas Internacionales de Valoración (NIV) y Royal Institution of Chartered Surveyors (RICS) (Guadalajara , 2014). Cada nación adopta de manera individual los métodos según el tipo de valuación del inmueble, las características del mercado, del régimen de la propiedad y de las posibilidades en el mercado inmobiliario. Las normas deben de cumplirse en los avalúos de garantía. Sin embargo, se utilizan por los valuadores de inmuebles para los avalúos de inmuebles urbanos con otro tipo de finalidades, por lo que se aplica para distintas finalidades. Se desarrolla los llamados métodos técnicos de valuación, que coinciden con las Normas Europeas de Valuación. Los métodos son los siguientes: 1) Método del coste. 2) Método de comparación. 3) Método de actualización de rentas. 4) Método residual.

2.2.2 Métodos avanzados

Los métodos avanzados como Precios Hedónicos, Inteligencia Artificial (Redes Neuronales Artificiales, Lógica difusa, Sistemas Expertos y Algoritmos Genéticos) Análisis Espacial, K-Vecinos, Basados en la Teoría de decisión multicriterio, Técnicas Box-Jenkins (Modelos ARIMA). Llamados así por diversos autores (Pagourtzi, V. , T. , & French , 2003), (Selim, 2009), (Kusan, Aytakin, & Özdemir, 2010). Notablemente distintas a los métodos tradicionales o técnicos, se basan en las técnicas matemáticas para estimar el valor, a través de procedimientos científicos, por lo que poseen más objetividad (Gallego , 2008). Dichos métodos tienen una aplicación práctica, se destacan por la valuación masiva en el parque inmobiliario de cualquier nación. Su uso se compara a ser como una “calculadora

de valores”, ya que permite la introducción de los datos de un inmueble dando como resultado un valor.

2.2.3 Metodología de Precios Hedónicos

Metodología de Precios Hedónicos (MPH). Este método se aplica a muchos activos con características heterogéneas. En el campo de la valuación de bienes inmuebles y particularmente en inmuebles urbanos residenciales donde existen un número mayor y puede aplicarse el método. Se han llevado a cabo varios estudios que investiga los factores determinantes en el precio de un inmueble urbano, entre los cuales se encuentra el análisis de 125 estudios de casas, realizado en la década 1994-2004, llegando a las siguientes conclusiones: Cada inmueble posee características específicas las cuales determinan su valor, ya que son valoradas de acuerdo con la zona geográfica de manera distinta. Un ejemplo es una cochera tiene más valor en zonas con clima muy frío, mientras que una alberca tiene mayor valor en climas cálidos. Por lo que los resultados obtenidos para una determinada zona geográfica son complicados de generalizar en otros lugares, por las variantes del clima y los aspectos sociales y culturales. Así también un comprador da el valor al inmueble de manera distinta por lo que el proceso de valoración se complica. (Rey-Carmona & Núñez-Tabal, 2017)

Las variables más utilizadas son las siguientes: la edad del inmueble, (con signo negativo), la superficie, con signo positivo. A continuación, el garaje, el tamaño del terreno, con signo positivo, la chimenea tiene efectos negativos en determinados trabajos. También el número de dormitorios (con signo negativo en algunos estudios), el número de baños (rara vez con signo negativo), y la piscina y el sótano (estos no tienen efecto negativo, aunque en muchas ocasiones fueron catalogados como “no importantes”). El ambiente natural, viene catalogado en muchos casos por la visualización de un lago, mar u océano o simplemente “buena vistas” (siempre con signo positivo sobre el precio). Entre las variables ambientales por proximidad o ubicación destaca la ubicación propiamente dicha medida por barrio, vecindario o Código Postal que tiene un efecto positivo, el índice de criminalidad en la zona

(efecto en contra), cercanía a campo de golf (efecto a favor) y presencia de árboles (efecto positivo). También hay variables ambientales resultantes de los servicios públicos: distrito escolar, presencia de población minoritaria creciente en escuela (signo negativo) y acceso a alcantarillado. Otras características relacionadas con la ocupación y factores de venta son las condiciones de habitabilidad, estado de conservación, si se encuentra o no vacía en el momento de la compra o tiempo en el mercado, así como aspectos relacionados con la financiación, como impuestos sobre la propiedad o financiación favorable. (Rey-Carmona & Núñez-Tabal, 2017).

2.2.4 Redes Neuronales Artificiales

El discernimiento del cerebro humano, han dado lugar a un avance de las RNA. Básicamente el cerebro humano almacena la información en forma de patrones, estos patrones tienen alta complejidad, por lo que se realizan tareas como el reconocimiento facial desde diferentes ángulos (Corchado, Díaz, Borrajo, & Fernández, 2000). Es por ello por lo que, al almacenar la información en patrones, para después usarlos y de esa forma solucionar problemas con ellos, esta conducta es la que imitan las redes neuronales artificiales.

Una Red Neuronal es un procesador paralelo distribuido y masivamente interconectado que almacena conocimiento experimental (Haykin, 1999). Los primeros estudios sobre valoración de inmuebles que utilizan la IA datan de principios de los años 90, por consiguiente, la aplicación de estas técnicas en el ámbito de determinación de precios inmobiliarios tiene más de dos décadas de vida. La mayor parte de los sistemas de IA aplicados a este campo son RNA, destacando el análisis pormenorizado efectuado por Gallego (2004) a la ciudad de Madrid. Especialmente interesante resulta la comparativa entre las técnicas tradicionales de precios hedónicos y la inteligencia artificial. En la mayoría de los estudios realizados con ese objetivo, las RNA han demostrado una mayor robustez y capacidad para detectar relaciones no lineales entre variables (Tay y Ho, 1992; Borst, 1991; Do y Grudnitski, 1993; Kauko, 2009; Peterson y Flanagan, 2009; Cannavaro,

2011; Núñez, 2007 y, en el ámbito de los locales comerciales, Rey, 2014). s. (Rey-Carmona & Núñez-Tabal, 2017)

2.2.5 Aprendizaje de máquina

El aprendizaje automático es un subcampo de la inteligencia artificial y la ciencia cognitiva. La inteligencia artificial se divide en tres ramas principales: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje reforzado. El aprendizaje profundo es un enfoque especial en el aprendizaje automático que cubre las tres ramas y busca también extenderlas para abordar otros problemas de inteligencia artificial que no suelen estar incluidos en el aprendizaje automático como la representación del conocimiento, el razonamiento, la planificación, etc. (Skansi, Introduction to Deep Learning from logical calculus to artificial intelligence, 2018)

El aprendizaje supervisado es una clasificación. El truco que supera la mayoría de los problemas puede verse como problemas de clasificación, por ejemplo, el problema de reconocer un vehículo en una imagen puede considerarse como clasificar la imagen en una de las dos clases: "tiene vehículo" o "no tiene vehículo". (Skansi, Introduction to Deep Learning from logical calculus to artificial intelligence, 2018). El aprendizaje supervisado se basa en descubrir la relación entre una variable de entrada y una de salida, es decir enseñarle al algoritmo cual es el resultado que se desea obtener tras mostrarle determinado valor. Después de enseñarle muchos ejemplos si se dan las condiciones el algoritmo será capaz de dar un resultado correcto, aún y cuando se le dé un valor que no haya visto antes. Mostrarle a un algoritmo suficientes datos de entrada y de salida y si existe una relación, será capaz de aprenderla, por lo anterior el aprendizaje supervisado es el que más aplicación práctica ha tenido.

El aprendizaje no supervisado en este caso los datos de entrenamiento son sin etiquetar. El sistema trata de aprender sin un profesor.

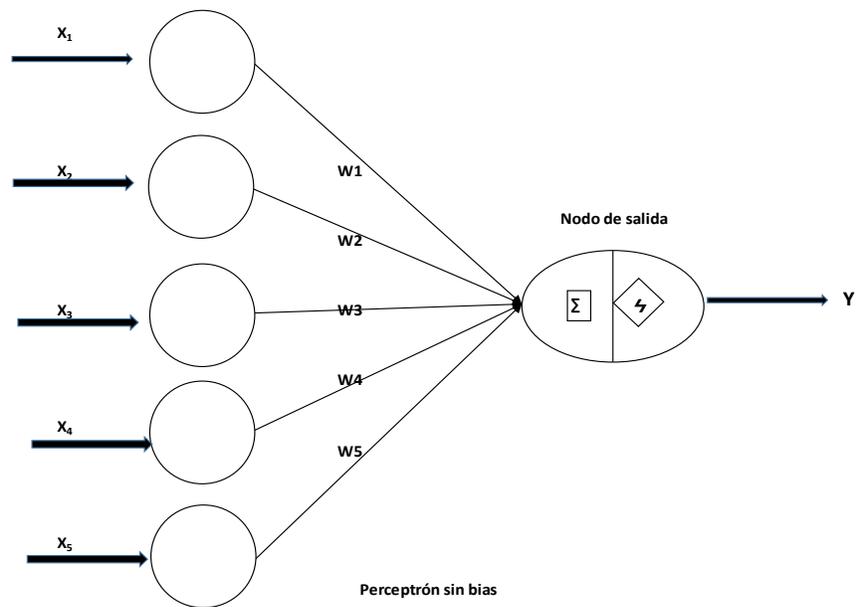
Los algoritmos de visualización son también buenos ejemplos de aprendizaje sin supervisión, les das de comer mucho complejo y sin etiquetar datos, y generan un 2D o Representación 3D de sus datos que se pueden trazar fácilmente (Skansi, Introducción al aprendizaje profundo, 2018). El aprendizaje no supervisado es un paradigma que consigue adquirir conocimiento solamente con los datos que se proporcionan como entrada sin necesidad, en ningún momento de explicar al sistema cual es el resultado que se desea obtener. Es adentrarse en un mundo desconocido y del que no se tiene información.

2.2.6 Arquitectura básica de una red neuronal

En la red neuronal de una sola capa, un conjunto de entradas se asigna directamente a una salida, mediante el uso de una variación generalizada de una función lineal. Esta simple instanciación de una red neuronal también se conoce como perceptrón. En las redes neuronales multicapa, las neuronas están dispuestas en capas, en las que las capas de entrada y salida están separadas por un grupo de capas ocultas. Esta arquitectura de capas de la red neuronal también se conoce como red de alimentación directa. (Charu C. , 2018)

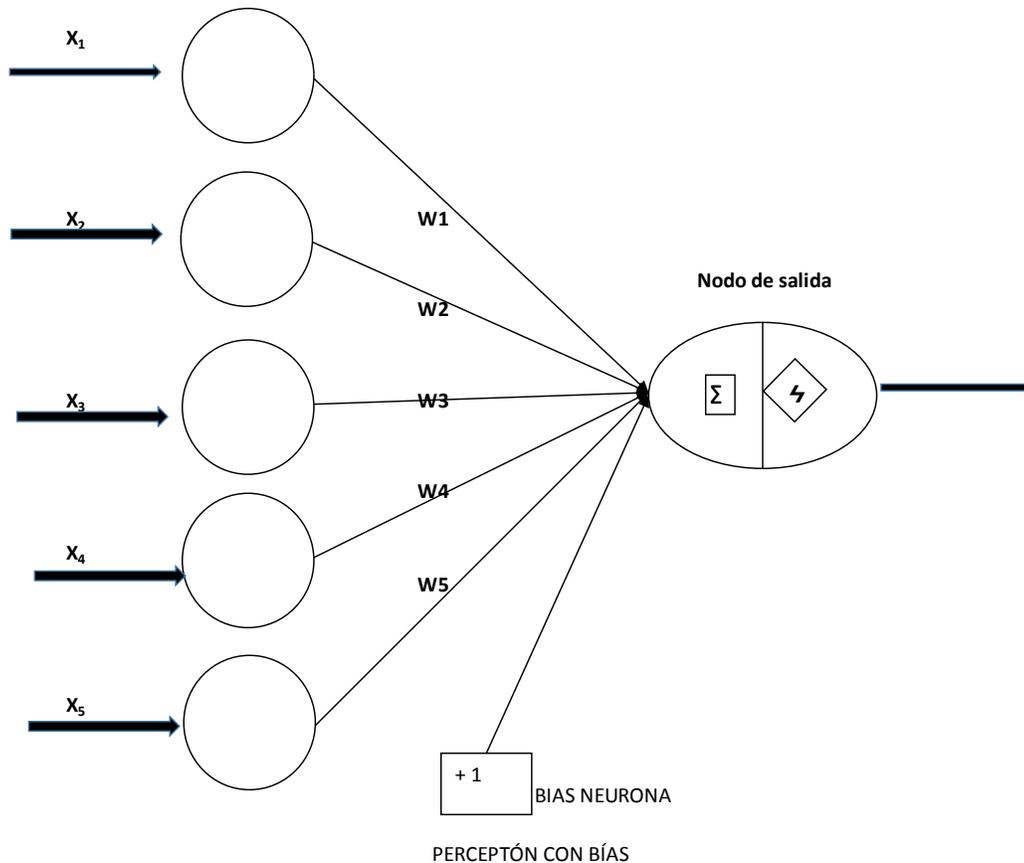
Capa computacional única: el perceptrón Esta red neuronal contiene una sola capa de entrada y un nodo de salida.

La red neuronal más simple se conoce como perceptrón. Esta red neuronal contiene una sola capa de entrada y un nodo de salida.



Fuente: Redes neuronales y aprendizaje profundo

Figura 1. Perceptrón sin bias



Fuente: Redes neuronales y aprendizaje profundo

Figura 2. Percepción con bias

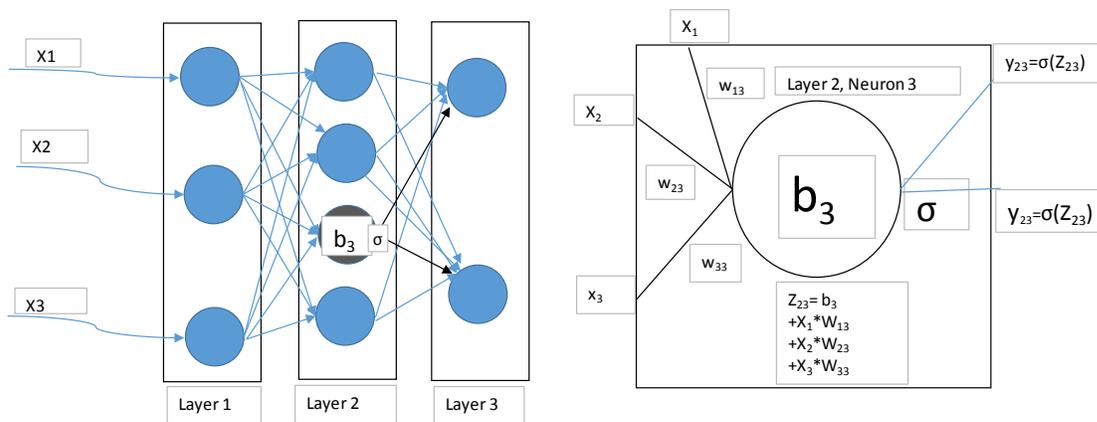
Perceptrón es la red neuronal más simple. Esta red neuronal contiene una sola capa de entrada y un nodo de salida. La arquitectura básica del perceptrón se muestra en la figura 1, se considera donde cada entrenamiento es desde (X, y) donde:

$X = [x_1 \dots x_d]$ contiene distintas variables, $y \in \{-1, +1\}$ contiene el valor observado de la clase binaria. Por “valor observado” nos referimos al hecho de que se da como parte de los datos de entrenamiento, y el objetivo es predecir la variable de clase para los casos en los que no se observa. La capa de entrada contiene d nodos que transmiten las d características $X = [x_1 \dots x_d]$ con bordes de peso $W = [w_1 \dots w_d]$ a un nodo de salida. La capa de entrada contiene d nodos que transmiten

las d características $X = [x_1 \dots x_d]$ con bordes de peso $W = [w_1 \dots w_d]$ a un nodo de salida. La capa de entrada no realiza ningún cálculo por derecho propio. La función lineal $W \cdot X = \sum_{i=1}^d w_i x_i$ se calcula en el nodo de salida. Posteriormente, se utiliza el signo de este valor real para predecir la variable dependiente de X . Por lo tanto, la predicción \hat{y} se calcula de la siguiente manera:

$$\hat{y} = \text{sign} \left\{ \sum_{j=1}^d \omega_j X_j \right\}$$

La función de signo asigna un valor real a $+1$ o -1 , lo cual es apropiado para la clasificación binaria. (Charu C. , 2018)



Fuente: Redes neuronales y aprendizaje profundo

Figura 3. Red neuronal simple

La estructura de una red neuronal simple de tres capas, misma que se ilustra en la figura 3, se muestra que la neurona de una capa está conectada a todas las neuronas de la siguiente capa, pero se multiplica por un llamado peso que determina qué cantidad de la cantidad de la capa anterior se transmite a una neurona de la siguiente capa. Por supuesto, el peso no depende de la neurona inicial, sino del par inicial de destino. Esto significa que el vínculo entre, la neurona N5 y la neurona M7 tiene un peso w_k , mientras que el vínculo entre las neuronas N5 y M3 tiene un peso

diferente, estos pesos pueden llegar a tener el mismo valor por accidente, pero en la mayoría de los casos, tendrán valores diferentes. El flujo de información a través de la red neuronal va desde las neuronas de la primera capa (capa de entrada), a través de las neuronas de la segunda capa (capa oculta) hasta las neuronas de la tercera capa (neuronas de salida). Ahora volvemos a la Figura 3. puede aceptar un valor de entrada, y están representados por las variables x_1 , x_2 , x_3 (los valores de entrada reales serán los valores para estas variables). Aceptar la entrada es lo único que hace la primera capa.

La capa toma una sola salida. Es posible tener menos valores de entrada que las neuronas de entrada (entonces puede pasar a las neuronas no utilizadas), pero la red no puede tomar más valores de entrada que las neuronas de entrada. Las entradas se pueden representar como una secuencia x_1, x_2, \dots, x_n (que en realidad es lo mismo que un vector de fila) o como un vector de columna $x = (x_1, x_2, \dots, x_n)$. Estas son representaciones diferentes de los mismos datos, y siempre elegiremos la representación que haga más fácil y rápido el cálculo de las operaciones que podemos necesitar. Nuestra elección de representación de datos no está restringida por nada más que la eficiencia computacional. (Skansi, Introducción al aprendizaje profundo, 2018)

Estado del arte

2.2.7 Estado del arte de las redes neuronales artificiales

La situación actual que atraviesa el mundo entero, hace cada vez más imprescindible el uso de las redes neuronales para analizar el impacto de las múltiples variables a las que se enfrenta la sociedad mundial. Como el acceso a la salud, educación y los servicios básicos de la vivienda. El Uso de las redes neuronales para analizar el impacto de dichos servicios contra la incidencia de contagio y muerte por COVID19 usando la metodología de las redes neuronales artificiales. (Sosa, Ortiz, & Cabello, 2020)

Los métodos automáticos de valuación tienen sus bases científicas y con algoritmos matemáticos de los cuales se desprenden objetividad al momento de

valuar, ya que se sustentan en metodologías estudiadas hasta llegar a los resultados deseados.

Dentro de los métodos que se expresan en esta investigación, se resalta que todos ellos manejan metodologías científicas en las cuales se presenta más precisión y objetividad en los resultados presentados.

Varios autores muestran la efectividad de las metodologías de lenguaje máquina.

En la ponencia presentada en Chihuahua (Casas & Lara, 2019) expresa” que los métodos de valuación automática forman parte de la vanguardia en los conocimientos y técnicas de valuación: Se apoyan en los avances del análisis estadístico y la tecnología de computación”. Desde el año 1970 se inicia con el desarrollo de análisis estadísticos y computacionales en la valuación. Lo anterior con el fin de elaborar con mejor predicción los valores de mercado en menor tiempo. Dichos métodos son apoyados por algoritmos computacionales y bases de datos con la descripción de las propiedades.

Los métodos automáticos de valuación que alcanzan más precisión en las predicciones son aquellos que manejan la inteligencia artificial, ya que utilizan algoritmos matemáticos, dejando la subjetividad a un lado, no se puede negar la confiabilidad que la tecnología ofrece a los usuarios, la inteligencia artificial ofrece una multitud de posibilidades en varias áreas de la ciencia y en esta investigación se refleja en los diferentes resultados ofrecidos.

El modelo de los precios hedónicos trabaja de distinta forma que los de lenguaje de máquina ofrecen, no por ello signifique que no presente resultados confiables. El MPH trabaja como una variable comparativa utilizando los métodos de comparación de mercado apoyada con la regresión múltiple.

3 CRITERIOS METODOLÓGICOS

3.1 Tipo de investigación

Tipo de investigación explicativa, debido a que se obtuvo la comparación de los dos métodos de valuación automática, con lo que se obtuvo evidencia que soporte la explicación de cual método es el más exacto en los términos y alcances de la investigación.

3.2 Hipótesis de investigación

El método de valuación que aplica redes neuronales artificiales produce resultados con mayor precisión que los métodos de precios hedónicos, vectores de soporte de regresión (Support Vector Regression) y bosques aleatorios (Random Forest).

3.3 Enfoque de la investigación

El enfoque de la investigación fue cuantitativo, ya que se evaluaron los cuatro algoritmos ya citados, respecto de la validez estadística de sus predicciones del valor de mercado de las viviendas en Cd. Chihuahua.

3.4 Tipo de investigación

La investigación fue de tipo aplicada y explicativa, ya que principalmente el resultado obtenido en ambos métodos se comparó entre sí, con el fin de conocer cuál fue el mejor método para tomar una decisión objetiva, utilizando un método matemático, y posteriormente generar alternativas de solución.

3.5 Diseño de la investigación

El diseño de la investigación fue de carácter no experimental transeccional, ya que no se manipuló ninguna de las variables de estudio, y se trabajó sobre los hechos ya existentes, solo se observaron y midieron los resultados de cada método.

3.5.1 Transeccional explicativo

Transeccional porque la evaluación del trabajo sólo se desarrolló en un tiempo específico y descriptivo debido a que la propuesta solo describe los resultados obtenidos, ya que los datos fueron los avalúos de bienes inmuebles de la Ciudad de Chihuahua. Chihuahua. La base de datos se capturó con información de avalúos realizados los años 2012 al 2015, período en que hubo relativa estabilidad en los precios de venta de viviendas

El trabajo se llevó a cabo en la ciudad de Chihuahua, entre los meses de junio a noviembre del 2020.

La población de interés con la que se trabajó fueron avalúos de vivienda de la ciudad de Chihuahua.

3.6 El marco muestral

El universo muestral está constituido por avalúos de vivienda de la ciudad de Chihuahua, cuyo propósito sea otorgar el inmueble como garantía hipotecaria a una institución crediticia y mismos que se encuentren inscritos en el Registro Público de la Propiedad, del municipio de Chihuahua, durante el periodo de estudio.

3.7 La unidad de análisis

Se considera el sujeto de estudio al avalúo realizado por unidades de valuación para créditos garantizados para la vivienda en el municipio de Chihuahua durante el 2015. Estos avalúos proporcionan los datos de las variables que se requieren para realizar la investigación.

3.7.1 El tipo de muestreo

El tipo de muestreo fue no probabilístico, con base a los 1928 avalúos identificados con las características del trabajo y a sus objetivos, avalúos para créditos garantizados para la vivienda en el municipio de Chihuahua.

3.7.2 El tamaño de la muestra

El tamaño de la muestra fue de 1928 datos obtenidos del Registro Público de la Propiedad de operaciones de compraventa como garantías hipotecarias. Dicha base de datos se normalizó, sacando de ella los datos poco comunes o fuera de rango quedando 1,845 datos. Con la base de datos se formaron conjuntos de avalúos, uno que contiene el 20% de la información de la base de datos, o sea 341 sujetos de investigación que fueron usados como testigo en la aplicación de la ecuación MPH que se obtuvo con el segundo conjunto compuesto con 1354 sujetos, o sea el 80% restante de la base de datos. Esta operación se repitió cinco veces para formar cinco conjuntos testigo y cinco conjuntos para calcular la ecuación MPH.

3.7.3 La selección de la muestra

No probabilístico por cuotas, ya que se capturaron un máximo de seis avalúos por colonia que cumplieran con los requisitos de ser de viviendas en garantía hipotecaria y que hubieran sido realizados entre los años 2012 al 2015. Algunas colonias no tuvieron operaciones de compraventa en el periodo seleccionado y en otras colonias varía de una a seis.

3.7.4 Variable predictora del estudio

La variable predictora del estudio o variable dependiente que se evaluó fue el valor de mercado neto de los avalúos de viviendas en la ciudad de Chihuahua.

3.7.5 Codificación de las variables independientes

Para poder realizar la regresión múltiple, se codificaron a valor numérico las variables alfabéticas como se indica a continuación:

Tabla 1. Nivel de Infraestructura

Id Nivel de infraestructura (SHF 2.1)	
Valor	Código
Nivel 1: No tiene alguno de los tres servicios básicos del nivel 2	1
Nivel 2: Cuenta con agua potable, drenaje y luz en la zona	2
Nivel 3: Cuenta con alumbrado público y vialidades terminadas (con banquetas) además de los servicios del nivel 2	3
Nivel 4: Cuenta con gas natural y vigilancia privada además de los servicios del nivel 3	4

Tabla 2. Nivel de equipamiento urbano

Eu : Nivel de equipamiento urbano (SHF 2.11)	
Valor	Código
Nivel 1: Cuando en la zona no exista algún elemento del nivel 2	1
Nivel 2: Cuando la zona cuente con iglesia, mercado o comercios, escuelas y parques y jardines.	2
Nivel 3: cuando la zona tenga los elementos del nivel 2 mas acceso o estación de transporte público.	3
Nivel 4: Cuando en la zona se hallen los elementos del nivel 3 mas hospitales	4

Tabla 3. Clasificación de la zona

Clasificación de la zona (Cz):	
Valor	Código
Primer Orden 1	1
Segundo Orden 2	2

Tabla 4. Referencia de proximidad

Referencia de proximidad urbana (Rp):	
Valor	Código
Céntrica	1
Intermedia	2
Periférica	3
De expansión	4
Rural	5

Tabla 5. Calidad predominante en las calles en las calles circundantes

Calidad predominante en las calles circundantes (Ca):	
Valor	Código
Económica	1
Medio	3
Residencial	5

Tabla 6. Nivel socioeconómico

Nivel socioeconómico (Catastro Chih.)	
Valor	Código
Económico	1
Interés social	2
Medio	3
Semilujo	4
Residencial	5

Tabla 7. Referencia de proximidad urbana

Rp.- Referencia de proximidad urbana (SHF)	
Valor	Código
Céntrica	1
Intermedia	2
Periférica	3
De expansión	4

Tabla 8. Calidad predominante

Ca.- Calidad predominante (SHF 2.14)	
Valor	Código
Económica	1
Interés social	2
Media	3
Semilujo	4
Residencial	5
Residencial Plus	6

Tabla 9. Clase general del inmueble

Cg.- Clase general del inmueble (SHF 4.3.1.1)	
Valor	Código
Económica	1
Interés social	2
Medio	3
Semilujo	4
Residencial	5

Tabla 10. Estado de conservación

Ec.- Estado de conservación (SHF)	
Valor	Código
Regular	1
Remodelado	2
Bueno	3
Muy bueno	4
Nuevo	5

(Nr) Número de recamaras.

(Nb) Número de baños completos.

(Sb) Suma de baños.

(Ee) Estacionamiento.

(See) Suma de estacionamientos.

Tamaño del Terreno.

(Tfr) Frente.

(Tfo) Tamaño de fondo.

(St) Superficie del terreno.

(Ed) Edad aproximada de las construcciones en meses.

(Eu) Vida útil remanente en meses.

(Sh) Superficie construida habitacional.

(Ssh) Suma de superficie construida habitacional.

(Sv) Superficie vendible.

(Cus) Coeficiente de utilización de suelo.

Tamaño del frente.

Tamaño del fondo.

La recolección de datos se realizó a través de la captura de los avalúos del registro público de la propiedad de la Ciudad de Chihuahua, por 15 estudiantes de la especialidad en valuación de la Facultad de Ingeniería

La codificación de la información se realizó dando un valor numérico a cada variable ponderando así el valor.

3.7.6 Medidas estadísticas básica

Las medidas de estadística básica como: máximos, mínimos, media, desviación estándar, varianza, varianza de medias y límite de error de la estimación.

3.7.7 Prueba de hipótesis t de student

La media de la diferencia entre los valores de mercado del conjunto testigo respecto de las predicciones del mercado.

Tabla 11. Estadística

Estadística	
Tamaño de la muestra	167
Media	0.010132
IC de 95 %	0.038475
Desviación estandar	0.22144
Objetivo	0.05

Prueba: se puede concluir que la media es menor que 0.05 en el nivel de significancia de 0.05

- IC: Cuantifica la incertidumbre asociada a la estimación de la media a partir de los datos de las muestras. Se puede estar 95% seguro de que la media verdadera es menor que 0.038475.

3.8 Análisis de regresión múltiple (MPH)

Se realizó el cálculo de cinco ecuaciones de regresión con el fin de definir la relación entre las variables dependientes y las variables independientes que son notables.

3.9 Algoritmo de vectores de soporte de máquina (Support vector machine)

El algoritmo de vectores de soporte es un clasificador discriminatorio, se define por un hiperplano de separación. Es decir, con los datos de entrenamiento etiquetados, el algoritmo origina un hiperplano ideal, el cual clasifica los nuevos ejemplos en dos espacios dimensionales, este hiperplano es una línea que divide un plano en dos partes donde en cada clase se encuentra en cada lado.

Este algoritmo se puede utilizar para dos acciones, de clasificación y de regresión.

3.10 Desarrollo de algoritmo de soporte de vectores de regresión (support vector regression)

Este algoritmo está basado en encontrar la curva o hiperplano que modele la tendencia de los datos de entrenamiento y según ella predecir cualquier dato en el futuro. Esta curva siempre viene acompañada con un rango (máximo margen), tanto del lado positivo como en el negativo, el cual tiene el mismo comportamiento o forma de la curva.

El total de los datos que esten fuera del rango se consideran como errores, ya que se necesita hacer un calculo de la distancia entre el mismo y los rangos. Esta distancia lleva por nombre epsilon y afecta directamente a la ecuación final del modelo.

El modelo se adecua de acuerdo como se comporten los datos, lo relevante es que se cumpla a cabalidad con el fin de obtener resultados fidedignos.

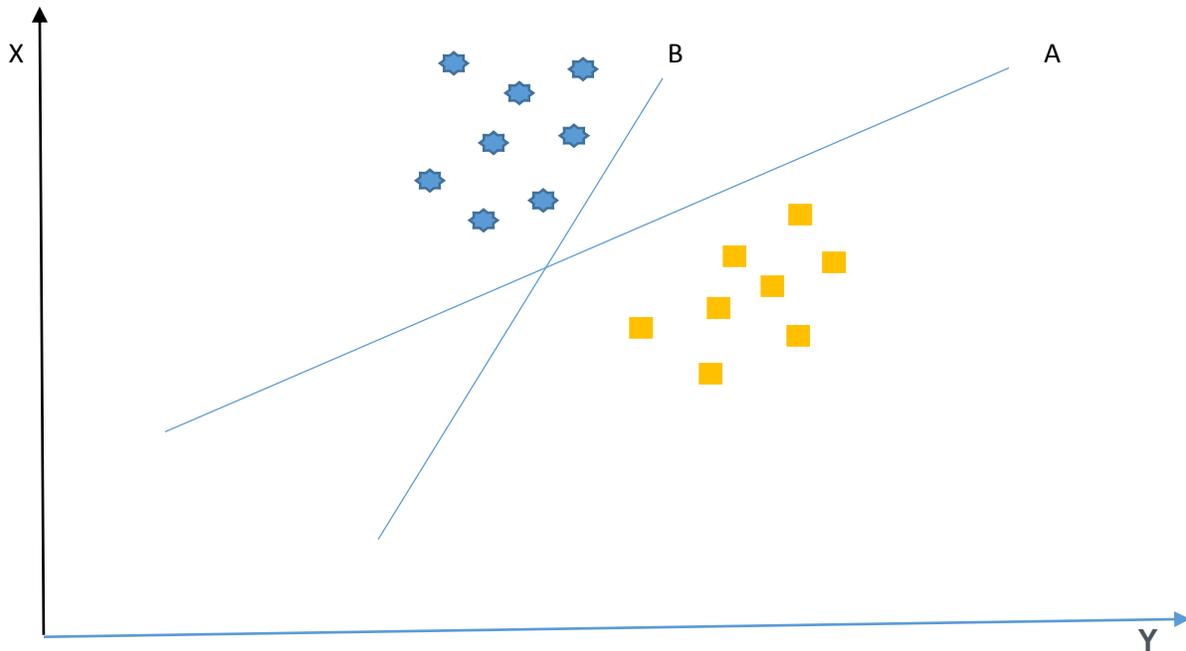


Figura 4. Support vector machine

3.11 Desarrollo de algoritmo de Bosques Aleatorios (Random forest)

Bosques Aleatorios es una combinación de árboles predictivos (clasificadores débiles) el cual trabaja serie de árboles intercorrelacionados y elabora un promedio. (Hastie, Friedman, & Tibshirani, 2001), en particularmente cada árbol depende de los valores de un vector aleatorio de la muestra, de manera independiente y con la misma distribución de todos los árboles en el bosque. (Medina Merino & Ñique Chacón, 2017)

Los bosques aleatorios son una combinación de predictores de árboles tal que cada árbol depende de los valores de un vector aleatorio muestreado de forma independiente y con el misma distribución para todos los árboles del bosque. Los errores de generalización para bosques converge en un límite a medida que el número de árboles en el bosque se vuelve grande. El error de generalización de un bosque de árboles clasificadores depende de la fuerza de cada árbol en el bosque y la correlación entre ellos. Usando una selección aleatoria de características para dividir cada una nodo produce tasas de error que se comparan favorablemente con Adaboost (Freund y Schapire [1996]), pero son más robusto con respecto al ruido. Estimaciones internas monitorean el error, la fuerza y la correlación y estos son

utilizado para mostrar la respuesta al aumento del número de características utilizadas en la división. Estimaciones internas también se utilizan para medir la importancia de las variables. Estas las ideas también son aplicables a la regresión. (Freund & Schapire, 1996)

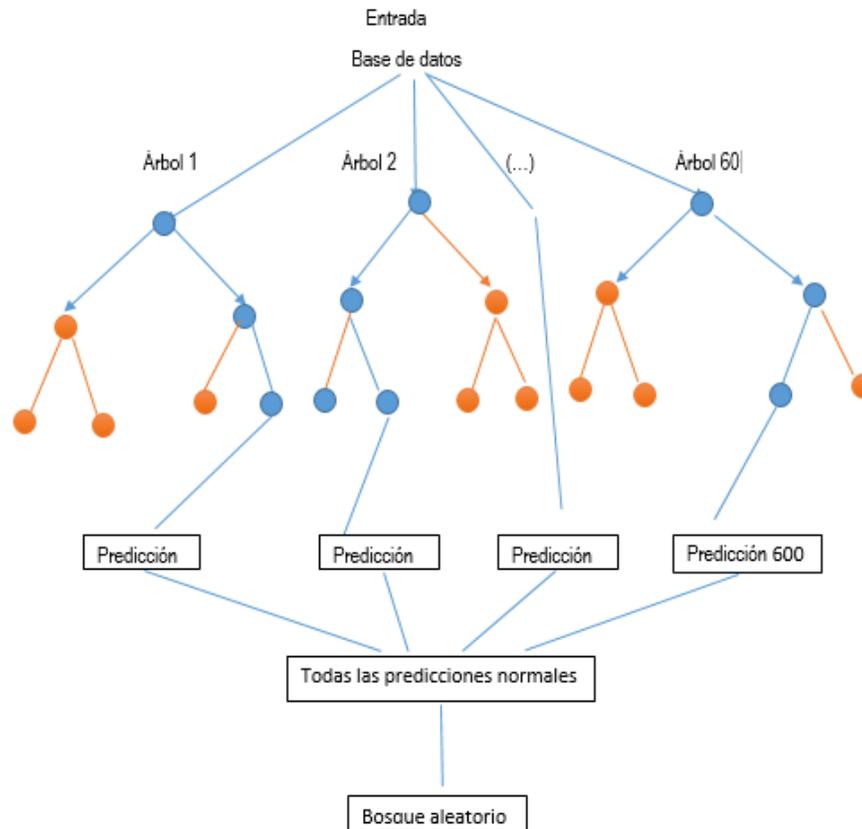


Figura 5. Bosque de decisión

3.12 Desarrollo de algoritmo de Redes Neuronales artificiales, Perceptrón Multicapa (MPL Multi Layer Perceptron)

El algoritmo usado para el proceso RNA fue el perceptrón multicapa o MLP (Multi-Layer Perceptron) se suele entrenar por medio de un algoritmo de retropropagación de errores o BP (Back Propagation) de ahí que dicha arquitectura se conozca también bajo el nombre de red de retropropagación.

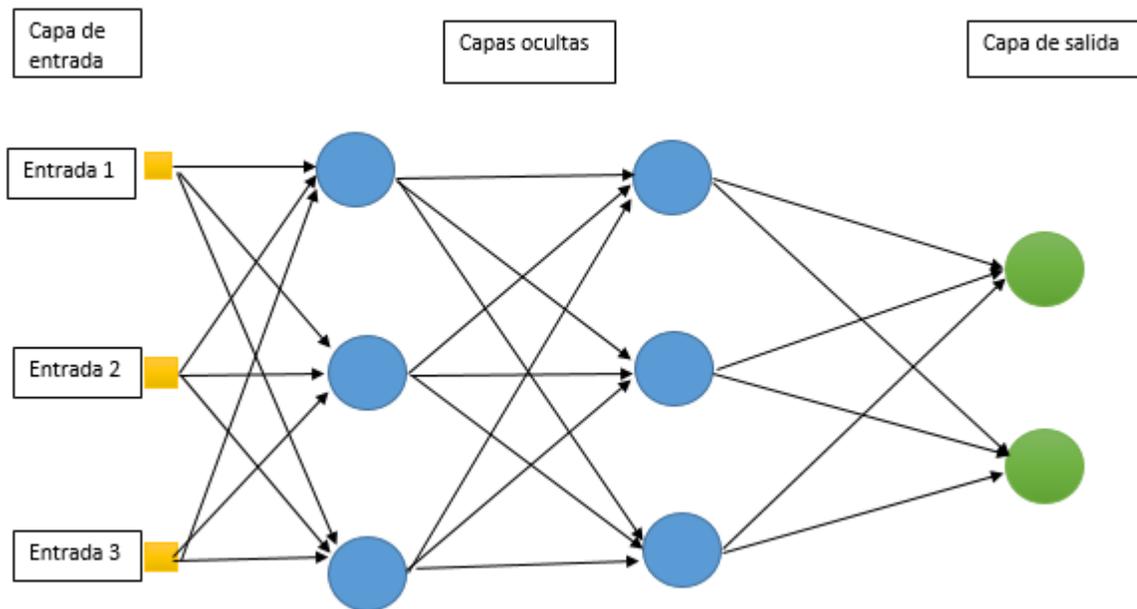


Figura 6. Perceptrón multicapa

4 RESULTADOS Y DISCUSIÓN

Resultados

4.1 Selección de la ecuación de MPH

La primera parte del método después de aplicar la regresión múltiple a los cinco subconjuntos de la base de datos se obtuvieron cinco ecuaciones, dichas

ecuaciones fueron probadas con los datos testigos correspondientes. Al final Se realizó el cálculo de la desviación estándar y la prueba t de Student al 95% de confiabilidad. Después de hacer la prueba se obtuvo que la ecuación que menor error tiene y será la que represente al Método MPH contra las redes neuronales. En la tabla 12 se muestran los resultados obtenidos.

Tabla 12. Resultados para la selección del Polinomio MPH óptimo

PRUEBA BASE MPH	MEDIA VERDADERA AL 95 %	COEFICIENTE DE DETERMINACIÓN R2 AJUSTADA	VALOR DE p	DESVIACIÓN ESTANDAR
MPH1	0.32347%	97.16	P<0.001	\$ 163,167.32
MPH2	0.04500%	97.08	P<0.001	\$ 138,756.18
MPH3	0.11765%	96.96	P<0.001	\$ 135,020.58
MPH4	0.00312%	96.91	P<0.001	\$ 126,827.61
MPH5	0.01374%	97.00	P<0.001	\$ 141,294.79

La ecuación con la menor desviación estandar resultó ser la ecuación **MPH4**

Ecuación de regresión

$$V_{mn} = 33018 I_d - 109185 C_z - 40374 N_r - 167501 N_b + 182155 S_b + 26180 S_{ee} - 10971 T_{fr} - 5349 T_{fo} + 124295 C_g + 16506 E_c - 255.1 E_d + 291.2 E_{u_1} + 2703 S_{sh} + 4668 S_v - 438651 C_{us}$$

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
142421	96.94%	96.91%	96.81%

4.2 Resultado de vectores de soporte de regresión (Support vector regression)

Después de aplicar el algoritmo de SVR a la base de datos y realizar el entrenamiento con el 80% de los datos, se procedió a realizar la prueba con el 20% restante. Se obtuvieron los siguientes resultados:

Desviación estándar \$ 24,943.17

R2 (ajustada) 98.70%

4.3 Resultado de Bosques aleatorios (Random Forest)

Después de aplicar el algoritmo de RF a la base de datos y realizar el entrenamiento con el 80% de los datos, se procedió a realizar la prueba con el 20% restante. Se obtuvieron los siguientes resultados:

Desviación estándar \$ 13,381.17

R2 (ajustada) 98.86 %

4.4 Resultado de Redes neuronales artificiales (Multicapa Perceptron)

Después de aplicar el algoritmo de MPL a la base de datos y realizar el entrenamiento con el 80% de los datos, se procedió a realizar la prueba con el 20% restante. Se obtuvieron los siguientes resultados.

Desviación estándar \$ 211,062.58

R2 (ajustada) 83.06

4.5 Discusión

En estudios anteriores acerca de los Métodos de precios Hedónicos y las Redes Neuronales Artificiales, apuntaron hacia las Redes Neuronales Artificiales como la opción que mejor describe el valor del mercado.

Los resultados obtenidos en este estudio demostraron que los algoritmos de inteligencia artificial alcanzan resultados de alta eficiencia, sin embargo en este caso en particular, el algoritmo de Redes Neuronales Artificiales, sorprendentemente tuvo un desempeño no esperado, dando como resultado el valor mas alejado del valor de mercado esperado.

Después del análisis de los resultados obtenidos en cada algoritmo, el mejor resultado fue el de bosques aleatorios, (Random Forest) debido a la naturaleza de su algoritmo, en donde se contruyen varios arboles de decisión y con la base de datos, llegó a ser la opción que mejor describe los valores concluídos. A continuación se muestra una tabla que ilustra los resultados de los cuatro métodos aplicados en esta investigación.

Tabla 13. Resultados obtenidos en los valores de desviación estándar y coeficiente de determinación R2 (ajustada)

Método	Desviación estándar	R2 (ajustada)
Método de Precios Hedónicos (MPH4)	\$ 135,020.58	96.91%
Support Vector Regression (SVR)	\$ 24,943.17	98.70%
Random Forest (RF)	\$ 13,381.17	98.86%
Redes Neuronales Artificiales RNA (Multicapa Perceptron)	\$ 211,062.58	83.06%

5 CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES:

Después de analizar los resultados obtenidos en los métodos automáticos de valuación presentados en la presente investigación, se destaca que en dos de los métodos de aprendizaje de máquina y en el Método de Precios Hedónicos se obtienen los resultados más certeros con valores del coeficiente de determinación superiores al 97%. En dos de los métodos de aprendizaje de máquina y en el Método de Precios Hedónicos se obtienen los resultados más certeros con valores del coeficiente de determinación superiores al 96%, dejando la subjetividad de lado, y pasando a un método más novedoso y nuevas formas de aplicar la valuación.

No dejando de lado el trabajo que los expertos en el área de la valuación aplican en cada uno de sus avalúos, pero abriendo nuevos horizontes del conocimiento para predecir con certeza y de manera objetiva el comportamiento del mercado actual.

El objetivo de la investigación fue la comparación de los métodos antes mencionados, dando como resultado que la hipótesis planteada no se comprobó, ya que el resultado final concluyó que el método de árboles de decisión (Random Forest) como la opción que mejor describe el valor de mercado de las viviendas en la Ciudad de Chihuahua.

La evidencia que se encontró en el estudio consiste en información de la validez estadística de los cuatro algoritmos estudiados, identificada por su capacidad de predicción del valor de mercado, medida está en función de la menor desviación estándar asociada a un coeficiente de determinación R^2 (ajustada) mayor del 95 %. Por lo tanto, se concluye que en orden de mayor a menor validez se encuentran los algoritmos: Regression, Random Forest (98.86 %), Support Vector Regression (98.70 %) y el Método de Precios Hedónicos (96.96 %). Un algoritmo no superó la meta de R^2 (ajustada) > 95 %, fue el de las Redes Neuronales Artificiales (83.06 %).

5.2 RECOMENDACIONES

Esta investigación fue realizada con el propósito de representar el fenómeno de valor de mercado de las viviendas de la ciudad de Chihuahua, puede ser replicado en a lo largo del territorio nacional, debido a que la información necesaria se encuentra en las dependencias del Registro Público de la Propiedad de los 32 estados de la República Mexicana. Las características hedónicas de las viviendas varían según la zona geográfica .

6 BIBLIOGRAFÍA O REFERENCIAS

- AEV Asociación Española de Análisis de Valor. (12 de octubre de 2020). *AEV Asociación Española de Análisis de Valor*. Obtenido de Asociación Española de Análisis de Valor 2020: <https://n9.cl/ghee>
- Almirall, M. P. (2007). *Introducción a la Valoración Inmobiliaria*. Barcelona: UPC.
- BERNALES GHACHAM, L. A. (2018). UN MODELO AUTOMÁTICO DE VALORACIÓN (AVM) APLICADO EN LIMA – PERÚ. *Retos de la valuación ante la globalización de los mercados* (pág. 18). Tijuana: LIV CONGRESO NACIONAL DE VALUACIÓN .
- Casas Comadurán, A. (2007). Valuación automática de vivienda. Chihuahua, Chihuahua, México.
- Casas, A., & Lara, B. (10 de octubre de 2019). Once años de experiencia profesional y académica en investigación aplicada para valuación automática de viviendas. Chihuahua, Chihuahua, México: LV Congreso Nacional de Valuación Chihuahua, Chihuahua.
- Charu C. , A. (2018). Redes Neuronales y aprendizaje profundo. En A. Charu C., *Redes Neuronales y aprendizaje profundo* (pág. 493). New York: Springer.
- Collado, R. H., & Lucio, P. B. (2014). *METODOLOGÍA DE LA INVESTIGACIÓN*. México: McGraw Hill Education.
- Corchado, J., Díaz, F., Borrajo, L., & Fernández, F. (2000). *Redes neuronales artificiales: Un enfoque práctico*. Vigo: Servicio de Publicaciones de la Universidad de Vigo.
- Fierro, O. (15 de noviembre de 2010). *Derecho* . Recuperado el 15 de abril de 2013, de <http://www.google.com/cse?cx=002866702082184079149%3Ahvjb6-5ekau&q=requisitos+para+publicar+en+Synthesis&sa=Buscar#gsc.tab=0&gsc.q=requisitos%20para%20publicar%20en%20Synthesis&gsc.page=1>
- Flores, C. (2011). *Tecnología de Información*. Madrid: Pearson.
- Flores, C. (2013). *Metodología de la Investigación*. México: Pearson.
- Freund, Y., & Schapire, R. (1996). Experiments with a new boosting algorithm, *Machine Learning: Proceedings. Experiments with a new boosting algorithm, Machine Learning: Proceedings I Conference*, (págs. 148 - 156).
- Gallego , M.-E. (2008). Modelos de valoración automatizada. *CT: Catastro*, 7-26.
- González, R. (2013). *Trata de Personas*. Madrid: Mc Graw Hill.
- Guadalajara , O. (2014). *Métodos de valoración inmobiliaria*. Madrid: Ediciones Mundi-Prensa.
- Hastie, T., Friedman, J., & Tibshirani, R. (2001). *The Elements of Statistical Learning*. Nueva York, Estados: Springer New York.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). *METODOLOGÍA DE LA INVESTIGACIÓN*. México: McGraw Hill Education.
- Hernández, E. (2017). *Valuación Inmobiliaria*. México: Trillas.
- Kauko, T. (2009). The Housing Market Dynamics of Two Budapest Neighbourhoods. *Housing Studies*, 587-610.
- Kusan, H., Aytakin, O., & Özdemir, I. (2010). The use of fuzzy logic in predicting house selling Price. *Expert Systems with Applications*, 1808-1813.
- Lara, B. (septiembre de 2019). VALIDEZ DE UNA ECUACIÓN DEL MÉTODO DE PRECIOS . Chihuahua, Chihuahua, México: Universidad Autónoma de Chihuahua.
- Mantilla, R. (2008). *Derecho Mercantil*. Mexico: Mc Graw Hill.
- Medina Merino, R. F., & Ñique Chacón, C. I. (2017). Bosques aleatorios como extensión de los árboles de. *Interfases*, 165 - 189.

- Navia, C.E., M. Ossa. (Enero de 2001). *Revista de Estudios Sociales*. Recuperado el 24 de mayo de 2013, de Universidad de los Andes. Facultad de Ciencias Sociales: <http://res.uniandes.edu.co/view.php/203/pdf/view.php>
- Pagourtzi, E., V. , A., T. , H., & French , N. (2003). Real estate appraisal: A review of valuation methods. *Journal of Property Investment & Finance* , 383-401.
- Pérez, R. (2011). *Metodología*. México: Pearson.
- Preciado, J. (octubre de 2015). VALORACIÓN DE INMUEBLES URBANOS: COMPARATIVA MODELO REGRESIÓN MULTIVARIABLE VERSUS REDES NEURONALES ARTIFICIALES PARA LA CIUDAD DE MORELIA, MICHOACÁN, MÉXICO. Valencia, España: Universidad Politecnica de valencia.
- Real academia de la lengua española. (1 de Noviembre de 2019). *Real Academia de la Lengua Española*. Obtenido de Real Academia de la Lengua Española: <https://dle.rae.es/algorithmo>
- Rey-Carmona , F., & Núñez-Tabal, J. (2017). Alternativas para la valoración de inmuebles urbanos . *Revista Publicando*, 3-19.
- Rodriguez, C. (2012). *Derecho Penal*. México: Pearson.
- Rosen, S. (1974). “Precios hedónicos y mercados implícitos: producto”Precios hedónicos y mercados implícitos diferenciación en pura competencia. *Revista de Economía Política*, 35 - 55 .
- S., P. (2019). Análisis de precios hedónicos de viviendas . *Revista Ingeniería de Construcción RIC* , PAG 215-220 .
- Selim, H. (2009). Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. *Expert Systems with Applications*, 2843–2852.
- Skansi, S. (2018). *Introducción al aprendizaje profundo*. Croatia: Springer.
- Skansi, S. (2018). *Introduction to Deep Learning from logical calculus to artificial intelligence*. Croacia: Springer International Publishing AG.
- Sosa, M., Ortiz, E., & Cabello, a. (agosto de 2020). Impacto del rezago social en el número de muertes y contagios. 1-17.
- Wilmer, R. A., & Olivo, B. M. (2018). *Redes Neuronales Artificiales aplicadas*. Machala: UTMACH.

7 ANEXOS:

7.1 ANEXO A CEDULA DE DATOS DEL INMUEBLE

2.1 Nivel de infraestructura urbana (Id)	2.11 Nivel de equipamiento urbano (Eu)	2.12 Clasificación de la zona (Cz) (capturar el No. Del orden)	Nivel Socioeconómico (Catastro Chih.) (Ns)	2.13 Referencia de proximidad urbana (Rp)	2.14 Calidad Predominante en las calles circundantes (Ca). Capturar el tipo de	3.5.1 Número de recámaras (Nr)	3.5.2 Número de baños completos (Nb)	suma de baños	estacionamiento		Tamaño del terreno		4.2 Superficie del terreno en estudio (St)	4.3.1.1 Clase general del inmueble (Cg)	4.3.1.2 Estado de conservación (Ec)	4.3.1.3 Calidad del proyecto (Cy)	4.3.1.4 Edad aproximada de las construcciones en meses (Ed)	4.3.1.5 Vida útil remanente en meses (Eu)	4.3.2.1 Superficie construida habitacional (Sh)		5.3.1 Superficie vendible (Sv)	CUS	Valor de mercado - Áreas y elementos adicionales - Vmn
									Descubiertos	suma	Frente	Fondo							T1	Suma			
Id	Eu	Cz	Ns	Rp	Ca	Nr	Nb	Sb	Ee	See	Tfr	Tfo	St	Cg	Ec	Cy	Ed	Eu	Sh	Ssh	Sv	Cus	Vmn
3	2	2	4	4	4	3	1	1.5	2	2	7	17.15	120.05	3	5	1	0	720	83.19	83.19	90.58	0.69296127	\$697,346.08

7.2 ANEXO C. ECUACIONES MPH

HOJA DE TRABAJO 1

MPH4 :Análisis de regresión: Vmn vs. Id, Cz, Nr, Nb, Sb, See, Tfr, Tfo, Cg, Ec, Ed, Eu_1, Ssh, Sv, Cus

Método

Filas no utilizadas 2

Ecuación de regresión

$$\begin{aligned} \text{Vmn} = & 33018 \text{ Id} - 109185 \text{ Cz} - 40374 \text{ Nr} - 167501 \text{ Nb} + 182155 \text{ Sb} + 26180 \text{ See} \\ & - 10971 \text{ Tfr} - 5349 \text{ Tfo} + 124295 \text{ Cg} + 16506 \text{ Ec} - 255.1 \text{ Ed} + 291.2 \text{ Eu}_1 \\ & + 2703 \text{ Ssh} + 4668 \text{ Sv} - 438651 \text{ Cus} \end{aligned}$$

Resumen del modelo

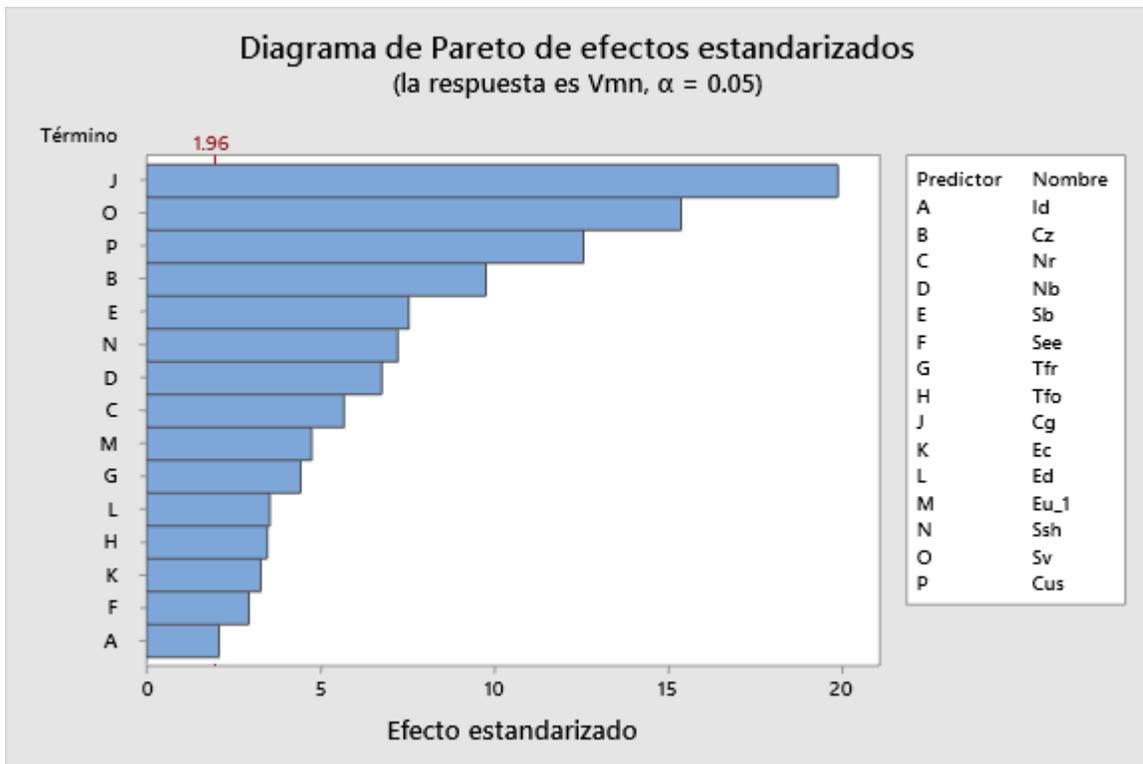
S	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
142421	96.94%	96.91%	96.81%

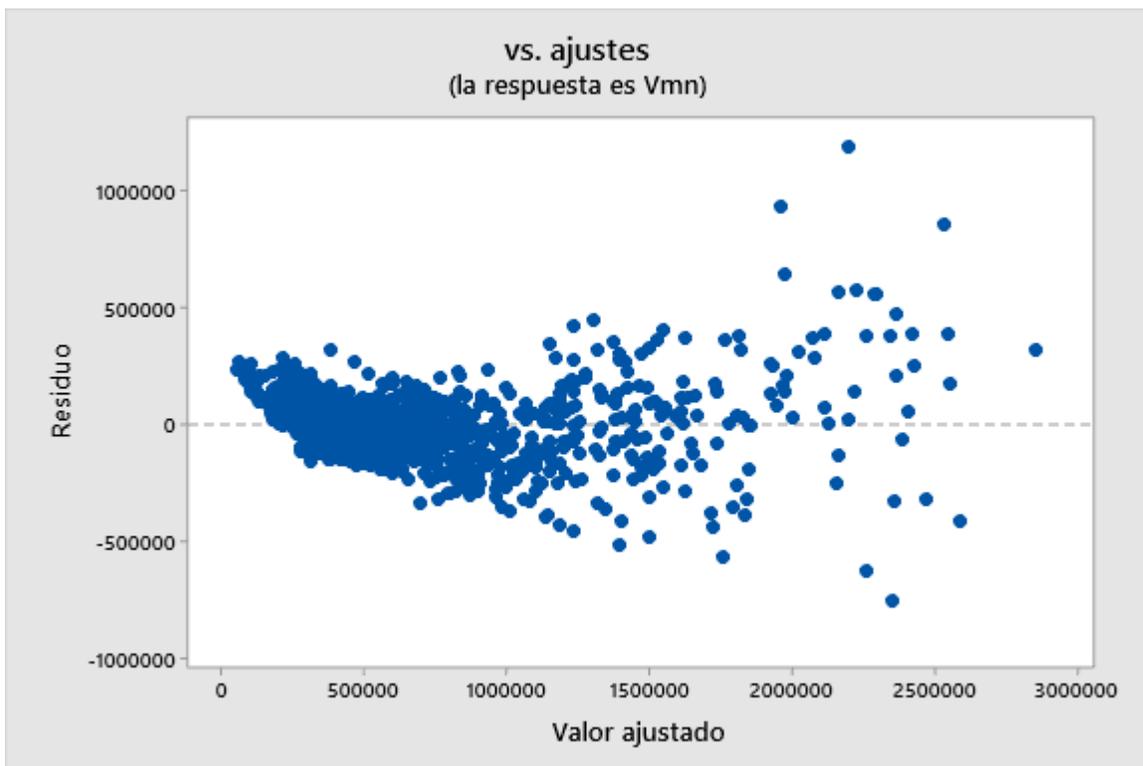
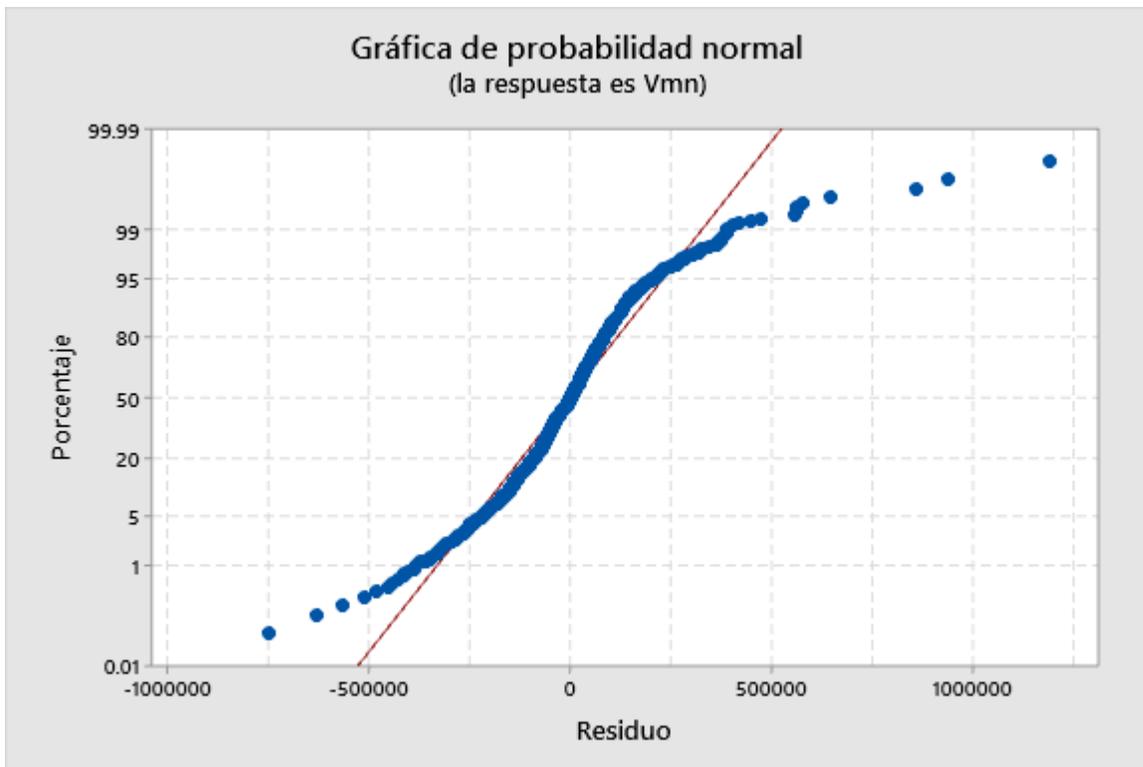
Análisis de Varianza

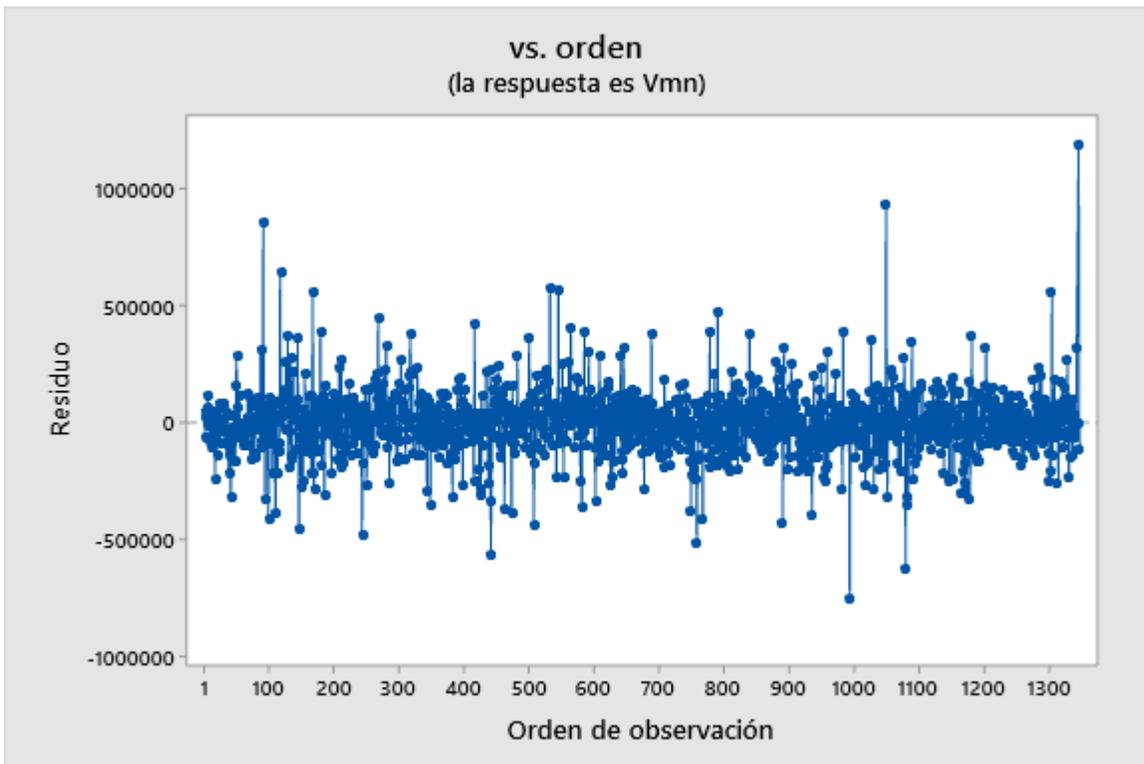
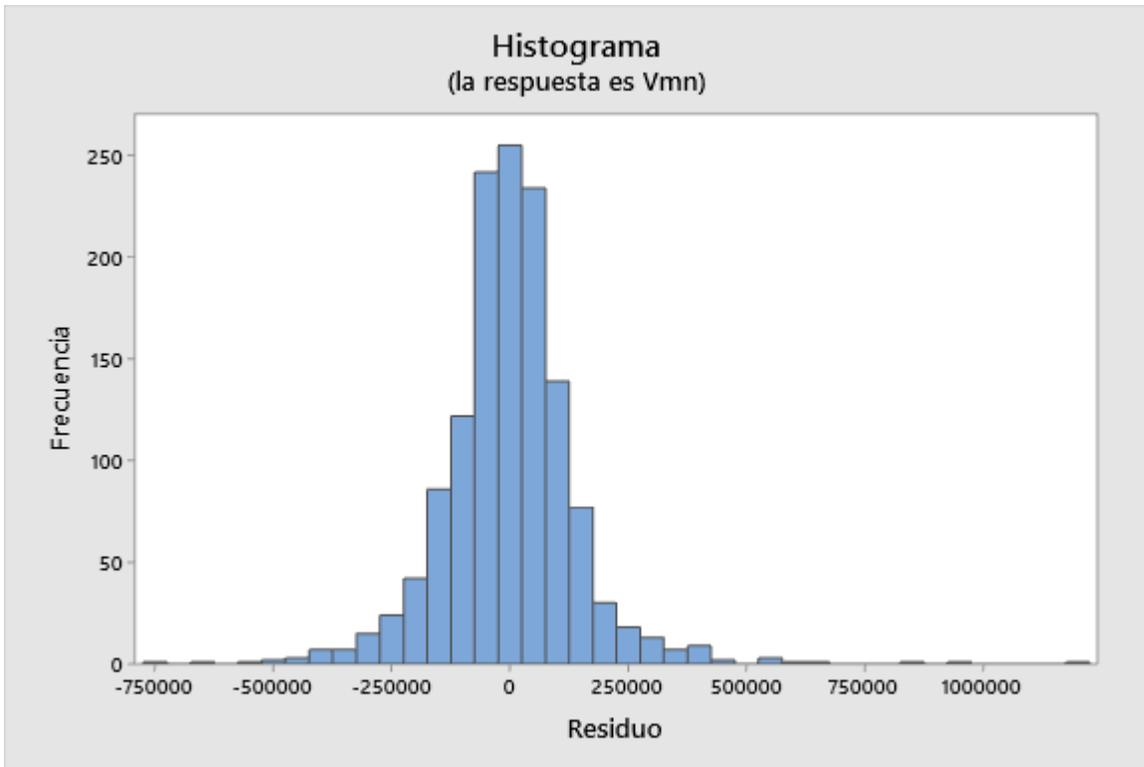
Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	15	8.55609E+14	5.70406E+13	2812.13	0.000
Id	1	88412894554	88412894554	4.36	0.037
Cz	1	1.93713E+12	1.93713E+12	95.50	0.000
Nr	1	6.54332E+11	6.54332E+11	32.26	0.000
Nb	1	9.29280E+11	9.29280E+11	45.81	0.000
Sb	1	1.15608E+12	1.15608E+12	57.00	0.000
See	1	1.75779E+11	1.75779E+11	8.67	0.003

Tfr	1	4.00012E+11	4.00012E+11	19.72	0.000
Tfo	1	2.45932E+11	2.45932E+11	12.12	0.001
Cg	1	8.02223E+12	8.02223E+12	395.50	0.000
Ec	1	2.19488E+11	2.19488E+11	10.82	0.001
Ed	1	2.54935E+11	2.54935E+11	12.57	0.000
Eu_1	1	4.58566E+11	4.58566E+11	22.61	0.000
Ssh	1	1.06170E+12	1.06170E+12	52.34	0.000
Sv	1	4.79628E+12	4.79628E+12	236.46	0.000
Cus	1	3.20598E+12	3.20598E+12	158.06	0.000
Error	1330	2.69774E+13	20283778016		
Falta de ajuste	1298	2.69672E+13	20775966820	65.05	0.000
Error puro	32	10219828398	319369637		
Total	1345	8.82587E+14			

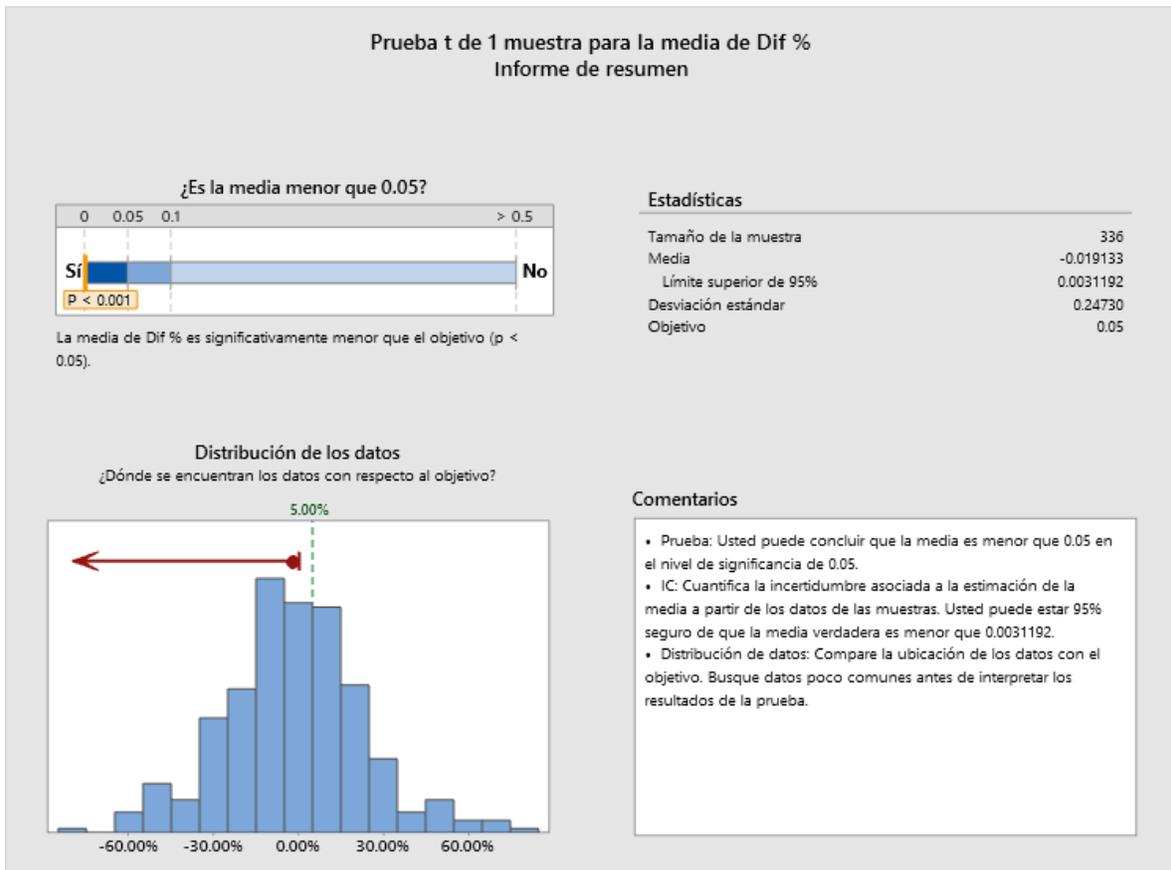
Residuo grande R
X poco común X







Prueba t de 1 muestra para la media de Dif% correspondiente a la MPH4



7.3 ANEXO E LENGUAJE DE MÁQUINA ELABORADO POR GIOVANI ALFONSO CHÁVEZ CENICEROS

1 RF vs SVM vs MLP

Giovanny Alfonso Chávez Cenicerros

Cargué el dataset en una carpeta de mi drive y accedí a través de ella.

Actualizando la librería scikit-learn (opcional)

```
! pip install --upgrade scikit-learn
```

```
Requirement already up-to-date: scikit-learn in /usr/local/lib/python3.6/dist-pa
Requirement already satisfied, skipping upgrade: numpy>=1.13.3 in /usr/local/lib
Requirement already satisfied, skipping upgrade: joblib>=0.11 in /usr/local/lib/
Requirement already satisfied, skipping upgrade: scipy>=0.19.1 in /usr/local/lib
Requirement already satisfied, skipping upgrade: threadpoolctl>=2.0.0 in /usr/lo
```

Montando nuestro directorio de Drive

```
from google.colab import drive
```

```
drive.mount("/content/gdrive")
!pwd
%cd "/content/gdrive/My Drive/DATA SCIENCE/datasets"
%ls
from google.colab import files
uploaded = files.upload()
```

```
Drive already mounted at /content/gdrive; to attempt to forcibly remount, call d
/content/gdrive/My Drive/DATA SCIENCE/datasets
/content/gdrive/My Drive/DATA SCIENCE/datasets
'luly (1).csv'          luly.csv
                       X_test.csv
                       y_train.csv 'luly
(1).gsheet'           luly.gsheat
                       X_train.csv
'luly (2).csv'        my_pipeline.html      y.csv
```

'luly (3).csv' X.csv y_test.csv
Choose Files luly.csv

- **luly.csv**(text/csv) - 451826 bytes, last modified: 11/9/2020 - 100% done
Saving luly.csv to luly (4).csv



1 Cargando el dataset

```
import pandas as pd
```

```
import io  
df = pd.read_csv(io.BytesIO(uploaded['luly.csv']))
```

```

import pandas as pd
import io
df = pd.read_csv(io.BytesIO(uploaded['luly.csv']))
#df = pd.read_csv("luly.csv") df.info()

```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 1737 entries, 0 to 1736 Data
```

```
columns (total 43 columns):
```

#	Column	Non-Null Count	Dtype
0	REG1	1737 non-null	int64
1	REG2	1737 non-null	int64
2	REG3	1737 non-null	int64
3	REG4	1737 non-null	int64
4	REG5	1737 non-null	object
5	REG6	1737 non-null	object
6	REG7	1737 non-null	object
7	REG8	1737 non-null	object
8	REG9	1737 non-null	object
9	Id	1737 non-null	int64
10	Eu	1737 non-null	int64
11	Cz	1737 non-null	int64
12	Ns	1737 non-null	int64
13	Rp	1737 non-null	int64
14	Ca	1737 non-null	int64
15	Nr	1737 non-null	int64
16	Nb	1737 non-null	int64
17	Sb	1737 non-null	float64
18	Ee	1737 non-null	float64
19	See	1737 non-null	int64
20	Tfr	1737 non-null	float64
21	Tfo	1737 non-null	float64
22	St	1737 non-null	float64
23	Cg	1737 non-null	int64
24	Ec	1737 non-null	int64
25	Cy	1737 non-null	int64
26	Ed	1737 non-null	int64
27	Eu.1	1737 non-null	int64
28	Sh	1737 non-null	float64
29	Ssh	1737 non-null	float64
30	Sv	1737 non-null	float64
31	Cus	1737 non-null	float64
32	\$T	1737 non-null	float64
33	\$C	1737 non-null	float64
34	Vt	1737 non-null	float64
35	Vo	1737 non-null	float64
36	Va	1737 non-null	float64
37	Vf	1737 non-null	float64
38	\$M	1737 non-null	float64
39	\$Mn	1737 non-null	float64
40	Vm	1737 non-null	float64
41	Vmn	1737 non-null	float64
42	Vco	1737 non-null	float64

dtypes: float64(20), int64(18), object(5) memory
usage: 583.6+ KB

1.1 Métodos para preparar los datos según las necesidades

```
class Preprocessor: def _init
    (self):
        # etiquetas de la información del registro inmobiliario
        self.R = ['REG1', 'REG2', 'REG3', 'REG4', 'REG5', 'REG6', 'REG7', 'REG8', 'REG9']
        # etiquetas de los datos categóricos
        self.C = ['Id', 'Eu', 'Cz', 'Ns', 'Rp', 'Ca', 'Cg', 'Ec']
        # etiquetas de los datos numéricos
        self.N = ['Nr', 'Nb', 'Sb', 'Ee', 'See', 'Tfr', 'Tfo', 'St', 'Cy', 'Ed', 'Eu.1',
                 'Sh', 'Ssh', 'Sv', 'Cus', '$T', '$C', 'Vt', 'Vo', 'Va', 'Vf', '$M', '$Mn', 'Vm', 'Vmn']

    def retrieve(self, X):
        X.drop(self.R, axis=1, inplace=True)
        y = pd.DataFrame(X['Vco'], columns=['Vco'])
        X.drop('Vco', axis=1, inplace=True)
        return X, y

    def retrieve2(self, X):
        y = pd.DataFrame(X['Vco'], columns=['Vco']) return X,
        y

    def drop_categories(self, X):
        X.drop(self.R + self.C, axis=1, inplace=True) y =
        pd.DataFrame(X['Vco'], columns=['Vco']) X.drop('Vco',
        axis=1, inplace=True)
        return X, y

    def drop_except(self, X, features):
        feats = [feat for feat in self.N if feat not in features] X.drop(self.R + self.C,
        axis=1, inplace=True) X.drop(feats, axis=1, inplace=True)
        y = pd.DataFrame(X['Vco'], columns=['Vco'])
        X.drop('Vco', axis=1, inplace=True)
        return X, y
```

```
preprocessor = Preprocessor()
```

▼ Preprocesando los datos implementando un *pipeline*

```
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import OneHotEncoder, StandardScaler from
cat_pipe = Pipeline([ ('onehot',
    OneHotEncoder())
])
```

```

from sklearn.pipeline import Pipeline
from sklearn.preprocessing import OneHotEncoder, StandardScaler
from sklearn.compose import ColumnTransformer

cat_pipe = Pipeline([ ('onehot',
    OneHotEncoder())
])

num_pipe = Pipeline([ ('scaler',
    StandardScaler())
])

prep = ColumnTransformer(transformers=[ ('cat',
    cat_pipe, processor.C), ('num', num_pipe,
    processor.N)
])

prep_pipe = Pipeline([ ('prep', prep)])

```

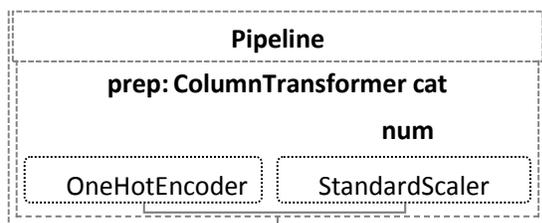
1.2 Visualizar el pipeline (opcional)

```

# actualizar la libreria sklearn antes de usarse! from sklearn
import set_config

set_config(display='diagram')
prep_pipe

```



```

# actualizar la libreria sklearn antes de usarse! from sklearn.
utils import estimator_html_repr

with open('my_pipeline.html', 'w') as f:
    f.write(estimator_html_repr(prepare_pipe))

```

Aplicando el *pipeline* sobre *trainy test*

```
import numpy as np
from sklearn.model_selection import train_test_split
```

```
X, y = preprocessor.retrieve(df.copy())
_X_train, _X_test, _y_train, _y_test = train_test_split(X, y, test_size=0.2, random_s
```

```
X2, y2 = preprocessor.retrieve2(df.copy())
_X2_train, _X2_test, _y2_train, _y2_test = train_test_split(X2, y2, test_size=0.2, ra
```

```
X2.to_csv('X.csv', index = False, header=True) y2.to_csv
('y.csv', index = False, header=True)
_X2_train.to_csv('X_train.csv', index = False, header=True)
```

```

_y2_train.to_csv('y_train.csv', index = False, header=True)
_y2_test.to_csv('y_test.csv', index = False, header=True)

# aplicamos el pipeline para ambas particiones del dataset X_train =
prep_pipe.fit_transform(_X_train)
X_test = prep_pipe.transform(_X_test) y_train =
np.array(_y_train).ravel() y_test =
np.array(_y_test).ravel()
print(_X_train.shape, _X_test.shape)
print(X_train.shape, X_test.shape)

(1389, 33) (348, 33)
(1389, 60) (348, 60)

```

(A) Support Vector Machine

```

from sklearn.svm import SVR
from sklearn.model_selection import GridSearchCV

%%time
svr = GridSearchCV(SVR(), [
    {'C': [1, 10, 100, 1000], 'kernel': ['linear']},
    {'C': [1, 10, 100, 1000], 'gamma': [0.001, 0.0001], 'kernel': ['rbf']},
], cv=10, verbose=0)
svr.fit(X_train, y_train)

print('Best C:', svr.best_estimator_.C) print('Best Kernel:',
svr.best_estimator_.kernel) print('Best Gamma:',
svr.best_estimator_.gamma) print("R^2:", svr.score(X_test,
y_test))
preds = svr.predict(X_test)

from numpy import savetxt savetxt('preds_svr.csv', preds,
delimiter=',') savetxt('truths_svr.csv', y_test, delimiter=',')

from sklearn.metrics import mean_squared_error
rms = mean_squared_error(y_test, preds, squared=False)
print("RMSE: ", rms)

```

```

☞ Best C: 1000
Best Kernel: linear Best
Gamma: scale
R^2: 0.9977789338602067

```

```

RMSE: 24943.167442794194

```

CPU times: user 28.7 s, sys: 14.2 ms, total: 28.8 s Wall time: 28.8 s

(B) Random Forest

```
from sklearn.ensemble import RandomForestRegressor from
sklearn.model_selection import GridSearchCV

%%time
forest = GridSearchCV(RandomForestRegressor(), [
    {'n_estimators': [1, 10, 100, 200], 'max_depth': [5, 10, 15]}
], cv=10, verbose=0)
forest.fit(X_train, y_train)

print('Best number of trees:', forest.best_estimator_.n_estimators) print('Best max
depth:', forest.best_estimator_.max_depth) print("R^2:", forest.score(X_test, y_test))
preds = forest.predict(X_test)

from numpy import savetxt savetxt('preds_rf.csv', preds,
delimiter=',') savetxt('truths_rf.csv', y_test, delimiter=',')

from sklearn.metrics import mean_squared_error
rms = mean_squared_error(y_test, preds, squared=False)
print("RMSE: ", rms)
```

```
Best number of trees: 200 Best max
depth: 10
R^2: 0.9993607858337313
RMSE: 13381.168887572372
CPU times: user 2min 11s, sys: 123 ms, total: 2min 11s Wall time:
2min 12s
```

(C) Multi-Layer Perceptron

```
from sklearn.neural_network import MLPRegressor from
sklearn.model_selection import GridSearchCV

%%time
mlp = GridSearchCV(MLPRegressor(), { 'solver':
    ['sgd'],
    'activation': ['tanh'], 'max_iter':
    [10000]
}, cv=5, verbose=0)
mlp.fit(X_train, y_train)
```

```
print('Best solver:', mlp.best_estimator_.solver) print('Best activation:',
mlp.best_estimator_.activation) print('Best max_iter:',
mlp.best_estimator_.max_iter) print("R^2:", mlp.score(X_test, y_test))
preds = mlp.predict(X_test)

from numpy import savetxt savetxt('preds_mlp.csv', preds,
delimiter=',') savetxt('truths_mlp.csv', y_test, delimiter=',')

from sklearn.metrics import mean_squared_error
rms = mean_squared_error(y_test, preds, squared=False) print("RMSE:
", rms)

Best solver: sgd
Best activation: tanh Best
max_iter: 10000
R^2: 0.8409693706612211
RMSE: 211062.58129959198
CPU times: user 48.5 s, sys: 28.2 s, total: 1min 16s Wall time: 39 s
```

Experiencia profesional

2008-Actual Facultad de Ingeniería de la U.A.Ch.

Cargo: Jefa de la Unidad de Administración Escolar, Coordinador de la Especialidad en Valuación
Función: Administración del sistema escolar SEGA de posgrado, altas, bajas, inscripciones, horarios, cambios de planes, cambios de campus, admisión de estudiantes de nuevo ingreso, seguimiento escolar de los estudiantes de las Maestrías que se imparten en la facultad.

Coordinación de la Especialidad en Valuación, re diseño curricular, horarios de Maestros, Control de protocolos, vinculación con los Colegios de Valuadores de Chihuahua, Control de tesis o Estudios de Caso.

2012- Actual Facultad de Ingeniería de la U.A.Ch.

Docente en la materia Seminario de Investigación

2000-2001- Instituto Bilingüe México Moderno

Cargo: Docente en la materia de computación

Formación académica

1992-1999 Ingeniero en Sistemas Computacionales

Facultad de Facultad de Ingeniería
Título de Ingeniero en Sistemas Computacionales

2009-2011 Maestría en Administración

Facultad de Contaduría y Administración de la U.A.Ch.
Grado de Maestría

2015 Curso Modelo Educativo de la Universidad

Curso: Programas analíticos por competencias
Curso: Evaluación por competencias
Curso: Valor de oportunidad
Curso: Valuación de Maquinaria y Equipo

2017 Diplomado en Rediseño Curricular (100 horas)

Curso: Proyectos formativos
Curso: Alternativas analíticas en la Valuación de Naves industriales

2017-2018 Especialidad en Valuación

Candidato a grado

2018 Curso: Bases de Datos

Curso: Normatividad y Temas Selectos en SHF
Curso: Casos prácticos de Mayor y mejor uso

2019 Curso : Estrategias para la Redacción y Publicación de Artículos Científicos

Curso: Aprender Normas APA
Curso: Avalúos Residuales, Dinámicos y Estáticos
Curso: Principios y lineamientos técnicos para valuar Maquina y Equipo Agroindustrial
Curso: Análisis de Mercado empleando Modelos de Regresión
Curso: Soluciones para la investigación ELSEVIER
Curso: Valuación Industrial Integral

LV Congreso Nacional de Valuación Chihuahua 2019 “VALUANDO LA TRANSFORMACIÓN EN GRANDE, SU EVOLUCIÓN HACIA UNA VISIÓN GLOBAL”

Curso: Liderazgo e Inteligencia Emocional

2020 Curso: Introducción a la Valuación de Aeronaves

Curso: Los drones, aplicación en la valuación inmobiliaria
Curso: MOOC Moodle para Docentes
Curso: Excel para valuadores
Curso: Enfoque de Ingresos 2020
Curso: hoteles 2020 UACH COVASFER

Otros datos

Manejo de equipo de cómputo y paquetería:

Word, Excel, power point.

Manejo SEGA (Sistema de Gestión Escolar Académica)

Habilidades:

- Trabajo en equipo
- Proactiva
- Organizada
- Trabajo bajo presión
- Disponible
- Responsable

Publicaciones

- Artículo: Las revistas depredadoras FING UACH Año 5 núm. 18 Diciembre –febrero 2019

María de Lourdes Flores Portillo

C. Valle del Ángel #5723

Fracc. Valle del Angel

6144254725 / 614-239-94-41

mlflores@uach.mx

Edo. Civil: Casada