

# Construcción de modelos combinando la econometría y el aprendizaje de máquina para el avalúo catastral de predios urbanos y rurales

Zhindón Martín<sup>1</sup>, Quevedo Sebastián<sup>2</sup>, Angamarca Pablo<sup>3</sup>, Córdova Federico<sup>1</sup>

1 Centro de Investigación/Unidad Académica de Ingeniería, Industria y Construcción/Universidad Católica de Cuenca  
Av. de las Américas/Azuay/Ecuador  
mgzhindonm@ucacue.edu.ec  
nfcordovag@ucacue.edu.ec

2 Ingeniería de Sistemas/Facultad de Tecnologías de la Información/ Universidad Católica de Cuenca sede Azogues  
Av. Ernesto Cheguevara/Cañar/ Ecuador  
asquevedo@ucacue.edu.ec

3 EVOTECH  
Panamericana Sur/Cañar/Ecuador  
pabloangamarca@gmail.com

## RESUMEN

La emisión de las Normas Técnicas para Catastro de Bienes Inmuebles Urbanos y Rurales del Ecuador (NTCBIUR), establecen las directrices para la elaboración de modelos de valoración de bienes inmuebles basados en la econometría. En este contexto este trabajo consiste en la construcción de modelos combinando la econometría y técnicas de aprendizaje de máquina (ML), por sus siglas en inglés, para el modelado eficiente de relaciones entre variables económicas complejas, para el avalúo de predios urbanos y rurales. La econometría es una técnica que ha sido empleada de manera convencional para representar relaciones entre variables económicas y realizar predicciones, y su enfoque ha sido la validación de modelos que parten de asumir que los datos provienen de modelos estocásticos, por otro lado existen técnicas modernas que se han desarrollado para la manipulación de los datos, los cuales no analizan la naturaleza de donde provienen los datos, ni el grado de relación que pueda existir entre variables, éstas técnicas emplean modelos a través de algoritmos que permiten realizar predicciones a partir de los datos, y no se centran en analizar los modelos sino la precisión de sus predicciones. La combinación de estas dos técnicas se realiza utilizando el modelo convencional empleado en la econometría para representar relaciones entre variables y el uso de técnicas empleadas en ML, para llegar a establecer los parámetros de éste modelo. Este documento es un reporte de resultados preliminares de las metodologías y técnicas empleadas para la actualización de los avalúos catastrales de los bienes inmuebles del cantón Déleg.

Palabras clave: **econometría, aprendizaje de máquina, datos, modelos, predicciones.**

## ABSTRACT

*The Technical Standards for Cadastre of Urban and Rural Real Estate of Ecuador establish guidelines for the elaboration of models of property valuation based on econometrics. In this context, this work consists of the construction of models combining econometrics and machine learning techniques, for the efficient modeling of relations between complex economic variables, for the valuation of urban and rural properties. Econometrics is a technique that has been used conventionally to represent relationships between economic variables and make predictions, and its focus has been the validation of models that assume the data come from stochastic models, on the other hand there are modern techniques that have been developed for the manipulation of data, which do not analyze the nature of where*

*the data come from, nor the degree of relationship that may exist between variables, these techniques use models through algorithms that allow to make predictions from the data , And do not focus on analyzing the models but the accuracy of their predictions. The combination of these two techniques is done using the conventional model used in econometrics to represent relationships between variables and the use of techniques used in ML, in order to establish the parameters of this model. This document is a report of preliminary results of the methodologies and techniques used for the updating of the cadastral valuation of canton Déleg.*

*Keywords: econometrics, machine learning, data, models, predictions.*

## I. INTRODUCCIÓN

En el Ecuador la Dirección Nacional de Avalúos y Catastros (DINAC) se institucionalizó en el año de 1966 con el fin de crear un inventario centralizado de las propiedades rurales (Lopez Chavez, n.d.), recientemente y luego de cincuenta años, la DINAC emite las NTCBIUR, En ésta se establecen los criterios técnicos y las metodologías para su aplicación a nivel nacional. La NTCBIUR obliga y faculta a los Gobiernos Autónomos Descentralizados Municipales (GADM) a emplear modelos econométricos para la valoración de bienes inmuebles (Dirección Nacional de Avalúos y Catastros, 2016). Por esta razón es necesario que los GADM analicen sus métodos de valoración que emplean actualmente y reformulen o crear nuevos modelos acordes a la normativa.

Los avalúos de bienes inmueble urbanos o rurales son establecidos considerando los precios comerciales actualizados en la zona, y realizados con conocimientos científicos, técnicos o artísticos (Contraloría General del Estado, 2016).

El caso de estudio seleccionado es el cantón de Déleg<sup>1</sup> por encontrarse en proceso de renovación de su plataforma tecnológica de gestión de catastro de bienes inmuebles, y al igual que el resto de cantones, sus GADM tiene que realizar la actualización de su catastro acorde a la NTCBIUR.

La metodología actual de valoración con la que cuenta el GADM del cantón Déleg, fue adoptada en el año 2005, mediante convenio de transferencia tecnológica a través de la Asociación de Municipalidades del Ecuador (AME). A través de este convenio entró en funcionamiento el sistema de avalúos catastrales, que automatiza la valoración de los bienes inmuebles. El convenio se limitó a la instalación del sistema y a la capacitación técnica con respecto a la herramienta tecnológica para su mantenimiento, pero no se transmitió el fundamento de la metodología que emplea para la valoración. Debido a esto la valoración ha sido guiada y se ha llevado a cabo mediante la herramienta tecnológica compuesta de un sinnúmero de indicadores, variables, formularios, pero sin conocer cuál es su fundamento, prácticamente, el instrumento es el que determina la valoración, sin un criterio técnico o científico, y mucho menos considerando la realidad del territorio. Este desconocimiento hace que la evaluación del método y su análisis se vuelva extremadamente complejo al momento de determinar si se ajusta a las metodologías propuestas por la NTCBIUR, siendo muchas veces la exploración del código fuente del aplicativo la única manera de comprender en alguna medida su funcionamiento.

Debido a la dificultad, la falta de conocimiento técnico o científico de los métodos de valoraciones y a la necesidad de adaptarse a la nueva normativa de NTICURB, el propósito de este trabajo es crear modelos para la valoración de bienes

combinando la econometría y una rama de la inteligencia artificial conocida como aprendizaje de máquina.

Las computadoras actualmente se encuentran involucradas en la mayoría de las transacciones, entre estas las económicas, permitiendo capturar éste tipo de datos (Varian, 2014). Una técnica común utilizada para manipular y analizar estos datos es la econometría. La econometría en sus primeras definiciones se conceptualiza como la unificación de las matemáticas, la estadística y la teoría económica, para comprender las relaciones cuantitativas de la vida económica (Frisch & Waugh, 1933). Una definición moderna establece que es la aplicación de las matemáticas estadísticas, herramientas de la inferencia estadística para la medición empírica de relaciones postuladas por una teoría de base. Las técnicas econométricas han sido empleadas en diversos campos como la política, la salud, transporte, ingeniería entre otras (Greene, 2012). De igual manera la NTICURB establece que la valoración de los bienes se puede realizar mediante la aplicación de éstos principios econométricos.

Por otro lado otra técnica utilizada para analizar estos datos es ML. El término ML fue introducido por primera vez por Samuel, (1959) quien lo definió como el campo de estudio en donde las computadoras son capaces de aprender sin la necesidad de ser programadas explícitamente. Posteriormente apareció una definición contemporánea y completa que indica que un programa de computadora aprende de la experiencia E con respecto a una tarea T y una medida de rendimiento P, si su rendimiento en la tarea T, medido por P, mejora con la experiencia E (Mitchell, 1997).

ML es una rama de la Inteligencia Artificial que trata con el desarrollo de algoritmos que pueden ser utilizados por las computadoras para el aprendizaje automático de patrones en los datos (Satnam, 2009).

ML y la econometría buscan patrones que puedan ser extraídos de los datos. El análisis de datos en la econometría entre otros busca predecir u obtener información de los datos mientras que ML esta principalmente enfocado en la predicción (Varian, 2014).

Las dos técnicas utilizan las matemáticas y la estadística para el tratamiento y análisis de los datos. La diferencia radica en que la econometría asume que los datos a ser analizados provienen de modelos estocásticos y éstos se utilizan para estimar parámetros de un modelo cuya validación se hace de acuerdo a pruebas de ajuste y análisis de residuos. El enfoque de ML por otro lado es encontrar un algoritmo que opere sobre entradas para predecir salidas, la validación del algoritmo se mide por la precisión en su predicción (Breiman, 2001).

Este documento es un artículo de reporte de caso. En éste se proporcionan detalles técnicos, metodológicos y resultados

<sup>1</sup> Déleg, cantón perteneciente a la provincial del Cañar.

preliminares en el proceso de la creación de un modelo matemático que combina la econometría y el aprendizaje de máquina, para el avalúo catastral de predios. No se abordan los criterios de selección de variables, las relaciones entre éstas, y tampoco su grado de influencia al momento de establecer avalúo.

El objetivo es crear un modelo combinando la econometría y ML ajustados a la NTICURB para la valoración de predios mediante el desarrollo de un algoritmo de aprendizaje personalizado que sea capaz de predecir el avalúo de un terreno en función de sus características físicas y económicas.

## II. MÉTODO

A pesar de que las técnicas convencionales estadísticas y de econometría funcionan correctamente, existen otras que permiten formas más efectivas de expresar relaciones económicas complejas, como ML. En este contexto tanto las técnicas de la econometría y minería de datos se encuentran fundamentadas en las matemáticas y la estadística, la diferencia es que la primera posee un componente adicional que es la teoría económica mientras que la segunda se fundamenta en la implementación de algoritmos. En este trabajo se utilizarán de manera combinada estas dos técnicas para crear un modelo que permita relacionar características de los predios, y a través de estos realizar predicciones sobre los avalúos.

Los parámetros que definen la ecuación que más se ajusta a los datos son obtenidos del entrenamiento del algoritmo de ML. Para verificar la validez de los parámetros obtenidos, las muestras se dividen en dos conjuntos, el primero correspondiente a datos de entrenamiento, el segundo a datos de prueba. Previo a la aplicación se realizará un proceso de escalado de las variables independientes, para finalmente encontrar la ecuación que mejor determine el precio del bien a valorar.

### A. COMBINACIÓN DE LA ECONOMETRÍA Y ML

Aunque la econometría puede ser empleada para hacer predicciones su principal aplicación es la estimación y análisis de modelos, su principal herramienta es el modelo de regresión múltiple y es considerado el único instrumento para realizar investigaciones empíricas. El modelo de regresión múltiple se utiliza para encontrar las relaciones entre las variables independientes o explicativas y la variable dependiente. A continuación el modelo de regresión lineal múltiple está definido por la ecuación [1]:

$$\hat{y} = f(x_1, x_2, \dots, x_k) + \varepsilon \quad (1)$$

$$= x_1\beta_1 + x_2\beta_2 + \dots + x_k\beta_k$$

donde  $\hat{y}$  es la variable explicada;  $x_i$  son las variables explicativas,  $\varepsilon$  es el error y  $\beta$  son los parámetros del modelo (Greene, 2012). La estimación de los parámetros se realiza mediante los mínimos cuadrados, que en su forma más simple es la de Mínimos Cuadrados Ordinarios (MCO), éste estima los parámetros como los valores que minimizan la suma de los errores  $\varepsilon$  entre los valores observados y los que predice el modelo, ecuación [2]:

$$\varepsilon = \sum_I (y - \hat{y})^2 \quad (2)$$

donde  $y$  es el valor observado y  $\hat{y}$  la predicción (Abdi, 2003). El enfoque de MCO para minimizar el error en un modelo de regresión lineal simple son las derivadas, mientras que para la regresión múltiple es el álgebra de matrices (Abdi, 2003). Éste es el enfoque convencional que utiliza la econometría para

estimar los parámetros del modelo. La combinación entre la econometría y ML consiste en reemplazar éstos dos métodos de estimación de parámetros convencionales, por un método de ajuste de aproximación numérica utilizado en ML denominado Algoritmo de Descenso de Gradiente (ADG).

La pendiente de gradiente es un método iterativo que trata de encontrar los mejores valores de estimación. En cada paso se hace una aproximación y esta se va mejorando sucesivamente conforme avanzan las iteraciones (Abdi, 2003). El ADG es la implementación de éste método utilizando un lenguaje de programación específico. Las redes neuronales consideradas parte de ML, constituyen una aplicación contemporánea del empleo de este método.

### B. ML Y REGRESIÓN MULTIVARIADA

Los algoritmos supervisados de regresión, buscan predecir resultados como una salida continua, para lo cual realizan un mapeo de los datos de entrada a una función que produzca esta salida continua, estableciendo una relación entre los datos de entrada y una salida conocida. En este caso la función de mapeo es la ecuación [3], que corresponde al modelo de regresión múltiple con notación empleada en ML (Ng, n.d.).

$$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \dots + \theta_k x_k \quad (3)$$

El algoritmo implementado aprende los parámetros  $\theta$  a partir de los datos, e intenta minimizar la función de costo, ecuación [4].

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta} x^{(i)} - y^i)^2 \quad (4)$$

En ML, la función de costo es análoga a la función del error que usa la econometría. ADG se encara de encontrar los parámetros  $\theta$  que minimicen la función de costo.

### C. EXTRACCIÓN, TRANSFORMACIÓN Y CARGA DE DATOS

Una de las actividades más importantes en esta etapa es la selección de variables que en conjunto sean determinantes en el avalúo del predio, ya que la precisión de las predicciones del modelo depende los parámetros que el algoritmo aprenda de los datos de entrenamiento. Las características seleccionadas del predio para su valoración son: servicios básicos, poblaciones cercanas al predio, vías, riego, topografía, drenaje, erosión, forma del predio, tipo de riesgos y área.

Una vez seleccionadas las características se procede a extraer de la base de datos del sistema de catastros, los predios y sus atributos utilizando el lenguaje SQL. Con la finalidad de eliminar registros atípicos, los predios a cargar en el algoritmo ML serán los que cumplan los siguientes criterios: precio > \$5000 y área < 77 Ha.

Además es necesario realizar un proceso de transformación de todos los atributos cualitativos para que puedan ser cuantificados y utilizados para cargar y entrenar el algoritmo ML. Las tablas I y II muestran la codificación de atributos ordinales y nominales respectivamente.

**TABLA I**  
Codificación de datos Ordinales

FORMA DEL PREDIO	CODIFICACIÓN
Regular	3
Irregular	2
Muy irregular	1

**TABLA II**  
CODIFICACIÓN DE DATOS NOMINALES

TIPO DE RIESGO	CODIFICACIÓN
Deslaves	1
Hundimiento	1
Volcánico	1
Inundación	1

#### D. VALIDACIÓN

Luego de la fase de entrenamiento, el resultado que devuelve el algoritmo es el modelo con los parámetros resultantes de la etapa de aprendizaje. La validación de éste se realiza utilizando el conjunto de datos de prueba, y su validez se establece en base a la precisión con la que realiza las predicciones.

### III. RESULTADOS

A continuación se presenta el ADG implementado con el lenguaje de programación Octave<sup>2</sup>, éste se utilizó para encontrar los parámetros  $\theta$  del modelo para la predicciones de avalúos.

```

for i = 1:iteraciones
    predicciones = X * theta;
    error = predicciones - y;
    delta_aux = zeros(length(theta),1);
    delta_aux = X' * error;
    delta = (alpha * (1/m)) .* delta_aux;
    theta = theta - delta;
end

```

Luego de haber desarrollado el algoritmo ML, se obtuvieron 7090 predios rurales de la base de datos. Éstos contienen todas las características del predio en función de las cuales se realiza el avalúo del inmueble, además constituyen los registros válidos, luego de la eliminaron de los registros atípicos. El 50% de éstos se utilizaron para el entrenamiento del algoritmo y el restante se utilizó para realizar la validación de las predicciones.

La Figura 1 permite visualizar como se minimiza la función de costo sucesivamente conforme el ADG avanza en cada una de las iteraciones.

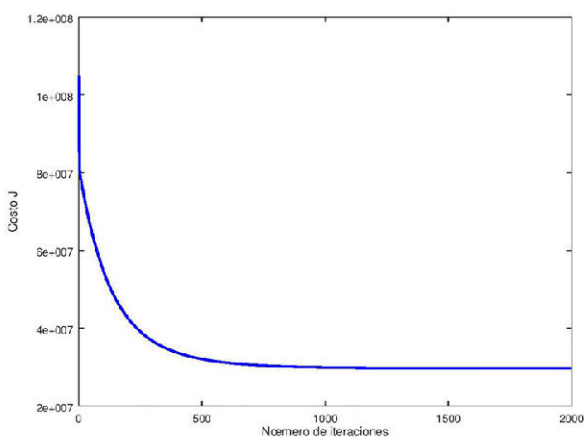


Figura 1. Gráfica que muestra cómo se minimiza el costo J en función de los parámetros  $\theta$  ajustados sucesivamente en cada iteración.

La tabla III contiene una muestra de diez valores correspondientes al conjunto de datos de prueba, para la evaluación del modelo.

**TABLA III**  
Comparación entre predicciones de avalúos por el modelo y avalúos reales

PREDICCIONES	REALES
\$19260.66	\$16945.70
\$5970.07	\$5603.54
\$20808.56	\$24664.53
\$18911.94	\$16088.93
\$20860.55	\$35747.08
\$7360.25	\$6183.98
\$17951.55	\$7108.11
\$21601.58	\$43160.50
\$21585.09	\$25595.46
\$17800.97	\$6438.69

### IV. CONCLUSIONES

La combinación de las técnicas econométricas y de aprendizaje de maquina es posible debido a que las dos se encuentran fundamentas en las matemáticas y la estadística, de esta manera es posible construir modelos que permitan establecer relaciones complejas para la valoración entre las variables independientes, combinando el aprendizaje de máquina para la valoración de predios. El algoritmo ML aprenden del conjunto de muestras de entrenamiento en lugar de programar fórmulas para una tarea específica.

La figura I, muestra que el funcionamiento de ADG es correcto, ya que en cada iteración reduce el costo que es equivalente a la reducción del error cuadrado. A pesar de que el funcionamiento del algoritmo es correcto, si se observa la tabla III, se puede observar que en esa pequeña muestra existen errores muy notables entre las predicciones y el valor real. Esto se debe a que el propósito inicial fue la construcción del modelo, por esta razón, la selección de variables no se realizó con el análisis adecuado, ya que si bien es cierto una variable puede estar relacionada directamente con el avalúo, al combinarla con otras variables puede perder su valor predictivo.

Pese a que el modelo fue probado con datos de predios rurales, éste se puede emplear para la valoración de predios urbanos, sin la necesidad de modificación alguna, ya que al tratarse un algoritmo de ML, lo único que necesita son los datos de entrenamiento, para a través de estos aprender los parámetros para cualquier valoración con cualquier número de variables que representen cualquier característica de los predios.

La metodología que se propone no se encuentra dentro de los enfoques y principios aplicados a la valoración que contempla la normativa NTCBIUR, esto no significa que el modelo no se pueda aplicar, ya que la norma en uno de sus apartados establece que en el caso de emplear otro método de valoración no contemplado, éste puede ser puesto a consideración ante las autoridades competentes (Dirección Nacional de Avalúos y Catastros, 2016).

Con la presentación de estos resultados preliminares y conociendo que el modelo funciona la siguiente etapa es mejorar

<sup>2</sup> <https://www.gnu.org/software/octave/>

la selección de las características de los predios que intervienen en el avalúo.

Este trabajo se puede replicar a otros municipios del Ecuador, ya que se conoce que sus avalúos son en función de la misma herramienta tecnológica implementada y transferida por la AME.

El aporte fundamental de este trabajo al GADM de Déleg es el componente científico, fundamental para el proceso de valoración de acuerdo a la NTCUR, esto marca el enfoque y la diferencia entre el desarrollo de este proyecto y el que pudiera prestar cualquier consultora en tecnología.

## REFERENCIAS

- Abdi, H. (2003). Least Squares., 1-4.
- Breiman, L. (2001). Statistical Modeling: The Two Cultures. *Statistical Science*, 16(3), 199-231.
- Contraloría General del Estado, E. (2016). Reglamento administración y control de bienes del sector público.
- Dirección Nacional de Avalúos y Catastros, E. (2016). Normas técnicas nacionales para el catastro de bienes inmuebles urbanos - rurales y avalúos de bienes; operación y cálculo de tarifas por los servicios técnicos de la Dirección Nacional de Avalúos y Catastros. Retrieved from <http://www.habitatyvivienda.gob.ec/wp-content/uploads/downloads/2016/08/Acuerdo-Ministerial-No-0029-16-Normas-Tecnicas-Nacionales-para-el-Catastro-de-Bienes-Inmuebles-Urbanos-Rurales-y-Avaluos-de-Bienes-Operacion-y-Calculo-de-Tarifas-de-la-Dinac.pdf>
- Frisch, R., & Waugh, F. V. (1933). Partial time regressions as compared with individual trends. *Econometrica: Journal of the Econometric Society*, 387-401.
- Greene, W. H. (2012). *Econometric Analysis*. Prentice Hall. <https://doi.org/10.1198/jasa.2002.s458>
- Lopez Chavez, J. R. (n.d.). Historia y evolución del catastro en el Ecuador. Retrieved from [http://www.catastrolatino.org/documentos/bogotanov2005/ponencia\\_ecuador.pdf](http://www.catastrolatino.org/documentos/bogotanov2005/ponencia_ecuador.pdf)
- Mitchell, T. (1997). *Machine Learning*.
- Ng, A. (n.d.). *Machine Learning* | Coursera. Retrieved June 8, 2017, from <https://www.coursera.org/learn/machine-learning/home/welcome>
- Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3), 210-229. <https://doi.org/10.1147/rd.33.0210>
- Satnam, A. (2009). *Collective intelligence in action*.
- Varian, H. (2014). *Big Data: New Tricks for Econometrics*.