

ANÁLISIS DEL MERCADO DE TIERRAS CON TÉCNICAS COMPLEJAS

Luis M. Vilches-Blázquez

Geógrafo, máster en Dirección de Proyectos GIS y Ph.D. en Ingeniería Geográfica. Dirección de Ordenamiento de la Propiedad y Mercado de Tierras de la Unidad de Planificación Rural Agropecuaria (UPRA). Ministerio de Agricultura y Desarrollo Rural.
luis.vilches@upra.gov.co

RESUMEN

La tierra es un activo valioso, sinónimo de propiedad, factor de producción, además de trabajo y capital. Asimismo, como cualquier otra propiedad, tiene un valor, ya que da lugar a unas ganancias tangibles e intangibles futuras que definen su valor de cambio en un mercado, es decir, un precio.

En trabajos existentes se ha identificado la necesidad de estimación de precio de la tierra. Sin embargo, en países como Colombia, dado el vacío informativo, esto supone una dificultad añadida. Ante esta situación, en este trabajo se recoge un análisis exploratorio de las características del mercado de tierras colombiano mediante la aplicación de técnicas complejas sobre el conjunto de variables presentes en los datos actuales.

Palabras clave: *machine learning*, datos, preprocesamiento, modelado, visualización.

ABSTRACT

Land is a valuable asset and a concept very related to property, production, work and money. Likewise, land has a value since it (land) produces profits for a long time. These profits define its value in the market, that is, a price.

Several proposals appear in the state-of-the-art for estimating the price of rural land. However, this estimating process is very difficult in countries as Colombia, where there is a lack of information for performing the aforementioned process. Taking into

account this scenario, this paper describes a set of activities performed for applying diverse complex techniques on different current datasets.

Keywords: *machine learning, data, pre-processing, modelling, visualization.*

1. Introducción

La tierra es uno de los activos más valiosos; se caracteriza por incluir elementos tales como superficie, espacio, suelo, suministro de alimentos, agua y una base para el desarrollo urbano e industrial (Verheye, s. f.). Este activo posee tres características que lo particularizan: escasez, inmovilidad física y durabilidad (Dasso, Shilling y Ring, 1995). Así, según Reydon *et al.* (2014), la escasez de tierra no es solo una consecuencia de su escasez física, sino también de la escasez de los productos que de ella emanan. Sin embargo, considerando el factor inmóvil que hace que no pueda ser reproducida, la escasez económica de la tierra es causada por la baja elasticidad de la producción y la sustitución, lo que permite la apropiación privada por parte de algunos agentes. No obstante, el desarrollo de tecnologías que aumentan su productividad, así como medidas administrativas como las reformas agrarias, pueden modificar sustancialmente el nivel de escasez de tierras en una región. Adicionalmente, la tierra está constantemente bajo la amenaza de la degradación, principalmente como resultado de múltiples presiones sobre la misma, producto de la agricultura intensiva, la minería, el manejo inadecuado del suelo y la presión de la población, entre otros factores.

Ante este escenario, resulta evidente que la tierra es sinónimo de propiedad y es un factor de producción, además de trabajo y capital. También, es un hecho que en torno a la tierra se conforman ciertas reglas de juego, derivadas de su necesidad, utilización, etc., por parte de las sociedades. Este conjunto de características permite alinear la tierra con el mercado.

Al trasladar este contexto a la tierra, y particularmente a sus mercados, se puede apreciar que, según Polanyi (citado por Delahaye, 1997), «lo que llamamos la tierra es un elemento de la naturaleza que está enlazado con las instituciones humanas».

La más extraña de todas las empresas de nuestros antepasados ha sido «separar la tierra del hombre» y «organizar la sociedad de manera a satisfacer las exigencias de un mercado de inmuebles». Así, la tierra, constituida en mercancía ficticia, se intercambia en el mercado en función de su costo marginal de uso, lo cual pone de manifiesto que, como cualquier otra propiedad, esta tiene un valor, ya que da lugar a unas ganancias tangibles e intangibles futuras que definen su valor de cambio en un mercado (Reydon *et al.*, 2014).

En los trabajos existentes en la literatura relacionada se identifica una necesidad de estimación del precio de la tierra (Choumert y Phélinas, 2015; Muñoz y Cardona, 2013; Soto, 2005), así como una creciente prevalencia de las estimaciones hedónicas de precios de la tierra utilizando múltiples variables (Choumert *et al.*, 2015; Larson, 2015; Allan *et al.*, 2014; Reydon *et al.*, 2014; Caballero, 2006; y Domingo *et al.*, 2005).

En América Latina y el Caribe, la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) ha pretendido observar el dinamismo que tiene el mercado de tierras, establecer sus dimensiones y las tendencias a través del manejo de información válida y confiable (Soto y Gómez, 2012). No obstante, en algunos escenarios, como es el caso de Colombia, el vacío informativo referente a la tenencia de la tierra imposibilita la obtención de información precisa sobre las transacciones, características de los agentes, superficies, montos de las transacciones y las características de las tierras, entre otros factores, a pesar de los esfuerzos y progresos que están realizando entidades como la Unidad de Planificación Rural Agropecuaria (UPRA).

Ante esta situación, este artículo aborda un trabajo exploratorio de aplicación de técnicas complejas, concretamente algoritmos de *machine learning* (aprendizaje automático) con el objetivo de desgranar las características y relaciones existentes en los conjuntos de datos con los que el mencionado grupo interactúa de forma habitual.

El aprendizaje automático es el área de la informática que «da a las computadoras la capacidad de aprender sin ser programadas de forma explícita» (Samuel, 2013), evolucionado a partir del estudio de reconocimiento de patrones y la teoría del

aprendizaje computacional en la inteligencia artificial; de este modo, explora el estudio y construcción de algoritmos que permiten aprender de los datos y hacer predicciones sobre los mismos (Kohavi y Provost, 1998).

La decisión de trabajar con algoritmos de *machine learning* se basa en el hecho de que en las últimas dos décadas este tipo de aprendizaje se ha convertido en uno de los pilares de las tecnologías de la información y con ello, en una parte central, aunque generalmente escondida, de nuestra vida. Además, con la creciente cantidad de datos disponibles, hay buenas razones para creer que el análisis inteligente de datos estará aún más presente como ingrediente necesario para el progreso tecnológico (Smola y Vishwanathan, 2010). De esta manera, se decide aprovechar la prominencia de los algoritmos de aprendizaje automático para identificar patrones y comportamientos en los datos actuales sobre el mercado de tierras rurales en Colombia.

Este artículo se estructura de la siguiente manera: la sección 2 presenta unos antecedentes y diversos trabajos relacionados con el marco de este trabajo; en la sección 3 se describe el detalle del trabajo realizado para el análisis del mercado de tierras con técnicas complejas; finalmente, la sección 4 recoge las principales conclusiones de este trabajo.

2. Antecedentes y trabajo relacionado

2.1. Antecedentes

Múltiples trabajos de investigación buscan desentrañar el valor de las tierras desde sus estructuras. Estos pueden ser agrupados en tres principales líneas de investigación (Larson, 2015):

- El **enfoque residual** intenta explotar la prevalencia de datos del valor de las propiedades, cantidades de estructuras y medidas de costes de estructura. Su principal problema es que el valor de la serie de estructuras se calcula utilizando el costo de reposición de las estructuras, asumiendo que el valor de reposición es una buena estimación del valor de mercado. Esta suposición se viola en áreas en declive. Entre los trabajos asociados con este enfoque destacan las propuestas de Davis y Heathcote (2007) y Davis (2009).
- El **enfoque basado en las transacciones espaciales**, desarrollado en los trabajos de Haughwout, Orr y Bedoll (2008) y Nichols, Oliner y Mulhall (2013), resulta ideal únicamente en lugares con un gran número de ventas de tierra o cuando las estructuras son escasas y, por tanto, conforman un componente relativamente pequeño del valor total de la propiedad. Sin embargo, el número de transacciones de tierra cae sustancialmente en el mercado de la vivienda y la venta únicamente de tierras cerca de los centros de las ciudades, donde los valores son los más altos, a menudo son poco frecuentes.
- El tercer enfoque, el **hedónico**, desarrollado hace más de setenta años para evaluar los problemas de calidad del producto, propone un método que estima una función que relaciona el precio de los productos con los diferentes atributos que posee (precio implícito). Esta propuesta se basa en dos pilares (Rosen, 1974): en primer lugar, en el hecho de que el producto tiene un precio; en segundo lugar, que tiene características o atributos medibles que definen los denominados precios hedónicos o implícitos. Este enfoque implica que las tierras resultan un bien heterogéneo que contiene un conjunto de características $Z = (z_1, \dots, z_k, \dots, z_K)$ vendidas en su conjunto.

No obstante, las propiedades se distinguen unas de otras a través de sus características intrínsecas y extrínsecas.

Esta técnica tradicional de estimación de precios implícitos tiene la ventaja de que todas las transacciones de propiedades que incluyen la tierra pueden ser utilizadas para estimar parámetros de precios, por lo que los problemas de selección y la escasez de datos resulta un problema menor, así que permite estimar precios en zonas geográficas más desagregadas.

2.2. Trabajo relacionado

La determinación de los precios de la tierra es un tema actual que posee un amplio trabajo e interés en la literatura relacionada, como Vilches-Blázquez (2016) describe detalladamente y cuya síntesis se muestra en la tabla 1. Esta proliferación de propuestas se produce como resultado de las particularidades del precio de la tierra, derivadas del hecho de que esta es un factor productivo inmóvil, durable y no reproducible. En consecuencia, como se mencionaba, el precio de la misma está influenciado por la dinámica de la oferta y demanda del mercado. Entre los trabajos analizados (ver tabla 1), desde una perspectiva general, se concluye que no existe una metodología común que pueda establecerse como referencia para el análisis y monitoreo de los precios de la tierra rural; que resulta fundamental poseer datos de calidad y con el mayor nivel de desagregación para poder llevar a cabo diferentes escenarios de análisis; que las variables que ayudan al establecimiento de los precios de la tierra resultan diversas, aunque con elementos comunes; y que, derivado de la heterogeneidad presente, no existe un método o una serie de métodos comunes para aplicar en cualquier escenario que permitan o contribuyan al análisis y monitoreo de los precios de la tierra rural.

Con respecto a las características particulares de los trabajos analizados (para más detalle, se recomienda ver Vilches-Blázquez, 2016), se destacan las propuestas que tratan de abordar la dinámica de los precios de la tierra desde una perspectiva nacional (Bailey y Lawson, 2015; Larson, 2015; Rey *et al.*, 2015; Allan y Kerr, 2014; Nickerson *et al.*, 2012; Baland y Robinson, 2012; USAID, 2010; Davies *et al.*, 2007;

Caballero, 2006; Soto, 2005; Ministerio de Agricultura, Alimentación y Medio Ambiente (s. f.)), aunque también aparecen propuestas a nivel provincial/estado (Reydon *et al.*, 2014; Osano *et al.*, 2011; Choumert y Phélinas, 2015) o ciudad (Palacio *et al.*, 2007; Domingo y Fulleros, 2005). En lo que respecta a las técnicas, destaca sobremanera la utilización de los precios hedónicos entre los trabajos analizados (Choumert y Phélinas, 2015; Larson, 2015; Allan y Kerr, 2014; Reydon *et al.*, 2014; Caballero, 2006; Domingo y Fulleros, 2005). Las fuentes de información consideradas en las propuestas analizadas, diversas y heterogéneas, se vinculan a las características de cada estudio realizado; sin embargo, destacan las fuentes relacionadas con censos agrarios e información estadística de diversa índole. Por su singularidad, merece destacarse el caso chileno (Baland y Robinson, 2012), cuya propuesta utiliza información del periódico *El Mercurio*, del periodo 1956-1960.

Al igual que sucede con las fuentes de información, las variables utilizadas en las diferentes propuestas presentadas en este documento resultan muy diversas, como puede comprobarse en la tabla 1. Se identifican variables comunes entre los diferentes acercamientos, que se clasifican en físicas y no físicas. Entre las primeras aparecen variables como la topografía, la calidad de los suelos, el clima y el agua, mientras que entre las no físicas se incluyen los precios, la distancia, las rentas y la infraestructura.

Por último, los métodos empleados, en consonancia con lo mencionado con anterioridad, resultan muy diversos y propios de cada propuesta. No obstante, como elemento común se identifica el índice de Laspeyres, utilizado por Domingo y Fulleros (2005) y el Ministerio de Agricultura, Alimentación y Medio Ambiente (s. f.). A pesar de la diversidad de trabajos presentes en el estado del arte para el análisis y definición de precios de la tierra rural, no se encuentra en la literatura ningún acercamiento encaminado a realizar dicho análisis o predicciones mediante la aplicación de técnicas complejas, específicamente mediante algoritmos de *machine learning*.

3. Análisis del mercado de tierras con técnicas complejas

En esta sección se recoge el detalle de la exploración de la aplicabilidad de los algoritmos de aprendizaje automático con el objetivo de lograr un mejor entendimiento y comprensión de las relaciones y patrones implícitos en los conjuntos de datos asociados con las dinámicas del mercado de tierras rurales en Colombia. Para la realización de este trabajo exploratorio se seleccionaron, por recomendación del grupo de expertos de Mercado de Tierras de la UPRA, dos municipios del departamento del Tolima: Ibagué y Ortega. En la figura 1 se muestra el contexto de la zona de estudio de este trabajo.

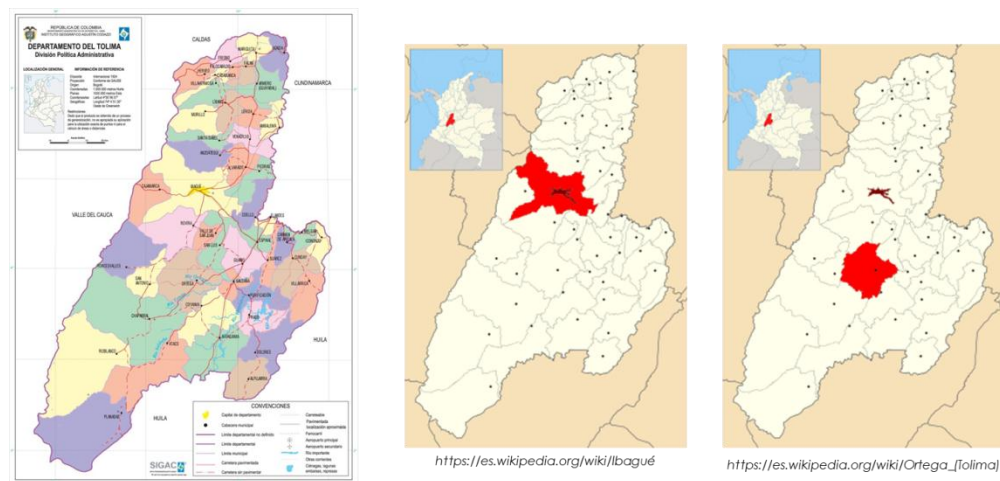


Figura 1. Zona de estudio

Tabla 1. Trabajos relacionados sobre mercado de tierras y precios

Autor(es)	Localización	Escala	Enfoques	Fuentes de información	Variables	Método
Reydon et al. (2014)	Brasil	Estado de Maranhão, zonas homogéneas (agregación de municipios)	Precios hedónicos	-	Físicas (suelo, clima y topografía), productivas (sistema de producción, ubicación y acceso), infraestructura de la propiedad (cercas, edificios) y las expectativas (situación regional e inversiones locales)	<ul style="list-style-type: none"> - Método de Ward - Mínimos cuadrados ordinarios - <i>Forward stepwise technique</i>
Larson (2015)	Estados Unidos	Nacional (enfoque microparcels a nivel de sección censal o inferior)	Precios hedónicos	Censo agrícola del Departamento de Agricultura de Estados Unidos (USDA)	Tipo de tierra, ecosistemas existentes en la tierra y el nivel de la zona de la urbanización, que refleja tanto los costos de transporte como el valor de opción de desarrollo de la tierra	-
Caballero (2006)	Colombia	Nacional	Precios hedónicos	-	Catastro, zonas físicas homogéneas, clima, pendiente del terreno, valor potencial del suelo, uso del suelo,	<ul style="list-style-type: none"> - Métodos econométricos - Coeficiente de Gini

					disponibilidad de fuentes de agua	
Bailey y Lawson (2015)	Reino Unido	Nacional	-	-	-	-
Domingo y Fulleros (2005)	Filipinas	Nacional (2 ciudades)	Precios hedónicos	<i>National Statistics Office Land Registration Administration Records of provincial/city/municipality assessors Bureau of Internal Revenue Housing and Land Use Regulatory Board National Mapping and Resource Information Authority Land Management Bureau Philippine Exports Zone Authority</i>	No explícitas, aunque se mencionan las siguientes: <i>trends of CPI rental prices, derived rental prices of commercial/industrial properties or prices of construction materials plus prices of labour</i>	<ul style="list-style-type: none"> - REPI (Real Estate Price Index) - Laspeyres index
Palacio <i>et al.</i> (2007)	México	San Luis Potosí, Coahuila, Sinaloa, Guanajuato y México (15 % de las	-	Procuraduría Agraria Fideicomisos Instituidos en Relación con la Agricultura (FIRA)	Cultivos, niveles de productividad de cultivos, precios de la tierra, ubicación, ganancias y rentas agrícolas	-

		entidades del país)				
Choumert y Phélinas (2015)	Argentina	Junín (provincia de Buenos Aires) y San Justo (provincia de Santa Fe)	Precios hedónicos	Muestra aleatoria que comprende 186 agricultores que poseen o cultivan 338 parcelas de tierra (julio-agosto de 2011)	Precios, calidad de suelos, ubicación y accesibilidad (carreteras)	- <i>Box-Cox transformation</i> - <i>Variance Inflation Factors (VIF)</i>
Allan y Kerr (2014)	Nueva Zelanda	Nacional	Precios hedónicos	Base de datos de <i>Quotable Value New Zealand (QVNZ)</i>	Número de ventas, precio de venta, la superficie terrestre vendida por categoría de uso de la tierra, precios de productos agrícolas	<i>Ricardian model</i>
Ministerio de Agricultura, Alimentación y Medio Ambiente (s. f.)	España	Nacional	-	Encuesta estadística	Precios de transacciones u operaciones de compra-venta, cultivos y aprovechamientos	- Precios medios ponderados - <i>Laspeyres index</i>

Rey <i>et al.</i> (2015)	Colombia	Nacional (1:100.000)	-	<p>Mapa de capacidad de uso,</p> <p>Mapa de pendientes (DEM a 30 metros),</p> <p>Mapa de inundaciones,</p> <p>Mapa de clasificación climática Caldas Lang.,</p> <p>Mapa avalúos catastrales,</p> <p>Mapa agroclimático,</p> <p>Mapas de isócronas,</p> <p>Mapa de distribución de la propiedad,</p> <p>Mapa de tradición predial,</p> <p>Mapa de oferta hídrica</p>	<p>Características físicas del suelo, aguas, vías, normas, tipologías de condicionamientos legales para el mercado tierras, estado de tenencia, distribución por tamaño de predios, distancia a centros de distribución y consumo</p>	<p>Análisis geoespaciales</p>
USAID (2010)	Ecuador	Nacional	-	<p>Instituto Nacional de Desarrollo Agrario</p> <p>Censo Agrario Nacional</p> <p>FAO</p> <p>Banco Mundial</p>	-	-
Nickerson <i>et al.</i> (2012)	Estados Unidos	Nacional	-	<p>USDA</p> <p>(<i>United States Department of Agriculture</i>)</p>	<p>Factores macroeconómicos (facilidad de crédito, inversiones, compras de tierras) y microeconómicos (uso de la tierra, calidad de suelo, distancia,</p>	<p>- Rent-to-value (RTV) ratio.</p> <p>- Population-Interaction Zones for Agriculture</p>

					subsidios, infraestructura, hogares rurales, servicios, presencia de recursos naturales)	(PIZA) measure
Baland y Robinson (2012)	Chile	Nacional	-	<i>El Mercurio</i> (periódico de Chile): 1956-1960. Índice de precios de la tierra (Comité Interamericano de Desarrollo Agrícola)	Anuncios sobre venta de tierras: tamaño, precio, provincia de localización, índice de precios al consumo	Modelo de clientelismo
Osano <i>et al.</i> (2011)	Sudáfrica	Provincia Occidental del Cabo	-	Datasets nacionales (<i>Farmer's Weekly Journal</i> , SANParks, WWF-SA, South African National Biodiversity Institute, Agricultural Research Council y Statistics South Africa), provinciales (Estate Agents, Western Cape Department of Agriculture y Cape Property Services) y locales (FFI, Parker, R MSc Project, Landowners)	Identificador de la propiedad (nombre e identificador único), tamaño, precio, año de venta, % de transformación, PIB, vegetación, suelos y fertilidad, clima, topografía, producción	- Correlación entre variables - <i>Generalized Linear Model</i>
Davies <i>et al.</i> (2007)	Australia	Nacional	-	Australian Bureau of Agriculture, Resource Economics Farm Survey Report.	Valor de mercado, tierras de secano o regadío, ganado, opción de	-

					compra, presupuesto, alquiler	
Soto (2005)	América Latina	País	-	-	Ubicación, producción, rentas de producción, rentas de usos alternativos, superficie de tierra, salario agrícola, costo de arriendo, tecnología, calidad del suelo, acceso al agua, precio de venta de productos, costos transporte, distancia, impuestos, PIB, subsidios	Modelo analítico para transacciones con cambio de uso de la tierra y entre productores agrícolas

La metodología empleada, inspirada en la propuesta de Cuesta (2013), tiene tres grandes actividades: preprocesamiento (obtención, limpieza, normalización y transformación de los datos), modelado (modelado predictivo, implementación y validación del modelo) y visualización (visualización e interpretación de resultados). La figura 2 muestra una abstracción de las actividades comprendidas en la metodología aplicada en este trabajo.

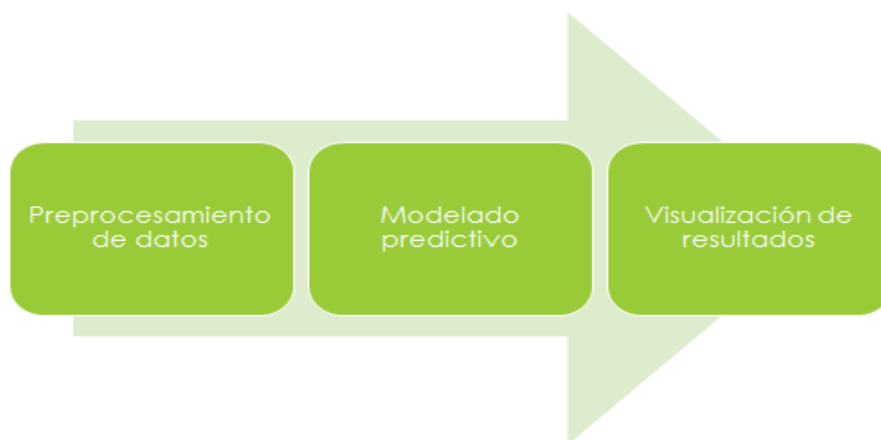


Figura 2. Metodología empleada

A continuación, se presenta el detalle del trabajo realizado para el desarrollo del trabajo exploratorio.

3.1. Preprocesamiento

Esta actividad se centra en cómo obtener, limpiar, normalizar y transformar los datos en un conjunto de datos óptimo, tratando de evitar cualquier posible problema de calidad de los mismos como valores no válidos, ambiguos, fuera de rango o faltantes. Este proceso suele requerir mucho tiempo en un proyecto y resulta fundamental llevarlo a cabo, ya que analizar datos que no han sido cuidadosamente preparados puede conducir a resultados altamente engañosos (Cuesta, 2013).

3.1.1. Obtención de datos

En este trabajo se trata con información recogida en 13 archivos diferentes, almacenados en formato Microsoft Excel, procedentes de diversas organizaciones, tales como la Superintendencia de Notaría y Registro (SNR), el Instituto Geográfico Agustín Codazzi (IGAC) y la Unidad de Planificación Rural Agropecuaria (UPRA).

- De la SNR, se trabaja con información relacionada con estadísticas nacionales de la Unidades Operativas de Catastro y Oficina de Registro de Instrumentos Públicos (UOC-ORIP) a nivel nacional, para los años 2012 al 2015.
- El IGAC aporta a este trabajo los archivos con información catastral a nivel nacional recogida en R1 y R2 de 2014. Específicamente para este trabajo se aborda la información de los municipios de Ibagué y Ortega.
- La UPRA proporciona para este proyecto diferentes conjuntos de datos asociados con el mapa edafoclimático y el potencial de suelos.

En el análisis de estos datos se identifican diversas variables susceptibles de ser utilizadas en la actividad de procesamiento realizada en este trabajo. Así, se identifican variables como el destino económico, avalúo, código ORIP, tipo potencial de suelo, nivel de ruralidad, etc.

Adicionalmente se destaca que, a pesar de su importancia, las fuentes no tienen información del componente espacial, lo que imposibilita tener esta caracterización en el territorio.

3.1.2. Limpieza, normalización y transformación de datos

Ante la diversidad de la procedencia de los conjuntos de datos con los que se trata en este trabajo, aun cuando estos provienen de entidades oficiales responsables de la generación de los mismos, se lleva a cabo un proceso de limpieza de estos para obtener materia prima adecuada.

Para este proceso se utiliza el *software* libre OpenRefine¹, herramienta originalmente creada por Google para el manejo de bases de datos que permite limpiar conjuntos de datos, exportarlos en diferentes formatos, así como arreglar y manejar las bases para un mejor uso. Actualmente el proyecto no es financiado por Google, es un proyecto abierto. Los archivos que OpenRefine permite pueden tener las extensiones TSV, CSV, XML, JSON, XLS e incluso Google Spreadsheets, entre otros. Este *software* también permite transformar archivos de cualquiera de estos formatos a otro.

La utilización de este *software* permitió llevar a cabo un proceso de limpieza y armonización de los diversos conjuntos de datos tratados, resolviendo problemas relacionados con incorrecciones ortográficas. No obstante, se identificaron otros problemas de calidad relacionados con la presencia de fechas erróneas, registros en blanco o diferencias en topónimos (figura 3). Diversas inconsistencias asociadas a los problemas de toponimia fueron solucionadas durante la realización de este trabajo. Sin embargo, se requiere la revisión en profundidad por parte de expertos para que estos contribuyan a su estandarización, lo que posibilitaría una armonización exhaustiva de los diversos conjuntos de datos en lo que concierne a este problema.

¹ <<http://openrefine.org/>>.

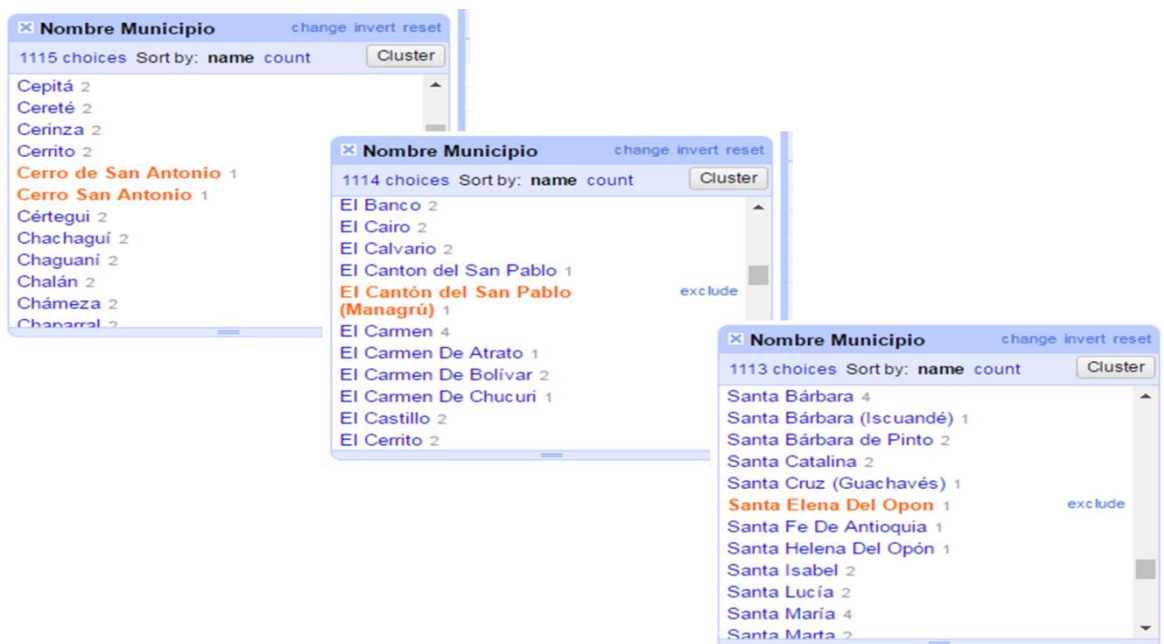


Figura 3. Problemas en topónimos

Tras el proceso de limpieza aplicado sobre los conjuntos de datos originales, se procedió a la transformación de los mismos (almacenados en archivos de Microsoft Excel) mediante la construcción de una base de datos MySQL, con el objetivo de poder conectar y consultar los múltiples archivos de una manera centralizada utilizando consultas SQL (Structured Query Language). La construcción de esta base de datos permite poder consultar de forma conjunta las múltiples variables recogidas en los diversos archivos tratados y que estas puedan ser utilizadas en la actividad de modelado.

3.2. Modelado

El modelado predictivo es un proceso utilizado en el análisis de datos para crear o elegir un modelo estadístico que intenta predecir mejor la probabilidad de un resultado (Cuesta, 2013). En este trabajo, se utilizó una variedad de modelos asociados a la implementación de diferentes algoritmos, agrupados en tres categorías: clasificación (árbol de decisión y *Naïve Bayes*), regresión (lineal y no-lineal) y *clustering* (*K-Means*).

Asimismo, se llevó a cabo una evaluación de los diferentes modelos propuestos que ayudó a asegurar que el análisis realizado no fuera demasiado optimista o excesivamente ajustado. Así, en este trabajo se validó el modelo de dos maneras diferentes: validación cruzada (subconjuntos de igual tamaño; se prueba el modelo predictivo para estimar cómo se va a realizar en la práctica) y *hold-out* (se dividen los datos al azar en tres subconjuntos: conjunto de entrenamiento, conjunto de validación y conjunto de prueba).

Con respecto a la interpretación de los resultados, en primer lugar, las implementaciones realizadas de los algoritmos de clasificación demuestran la pertinencia de aplicación del algoritmo «árboles de decisión» (tablas 2 y 3) sobre la propuesta de *Naïve Bayes* (tabla 4) en el conjunto de datos considerados.

Tabla 2. Resumen de resultados del algoritmo «árboles de decisión» sobre datos de Ortega (Tolima)

	Categoría predominante tamaño de la propiedad	Piso térmico	clase agrológica	Categoría predominante tradición	Pendiente	Destino económico
Number of leaves	149	15	10	166	22	34
Size of the tree	216	21	18	245	34	67
Correctly classified instances	46,12 %	99,78 %	99,81 %	61,11 %	100 %	100 %
Incorrectly classified instances	53,88 %	0,219 %	0,188 %	38,89 %	0 %	0 %

Tabla 3. Resumen de resultados del algoritmo «árboles de decisión» sobre datos de Ibagué (Tolima)

	Categoría predominante tamaño de la propiedad	Piso térmico	clase agrológica	Categoría predominante tradición	Pendiente	Destino económico
Number of leaves	212	17	33	152	39	17
Size of the tree	271	23	50	197	61	33
Correctly classified instances	55,69 %	99,41 %	99,80 %	84,78 %	100 %	100 %
Incorrectly classified instances	44,31 %	0,58 %	0,19 %	15,21 %	0 %	0 %

No obstante, merece ser resaltado que las implementaciones realizadas muestran que este algoritmo («árboles de decisión») no se puede aplicar con cualquier tipo de variable, puesto que los resultados obtenidos reflejan un funcionamiento desigual, pues ofrecen comportamientos ampliamente satisfactorios en los casos de las variables piso térmico, clase agrológica, pendiente y destino económico, y resultados mejorables en los casos de categoría predominante tamaño de la propiedad y categoría predominante tradición. Los detalles de los resultados ofrecidos por el algoritmo se recogen en las tablas 2 y 3.

Tabla 4. Resumen de resultados del algoritmo Naïve Bayes sobre datos de Ortega (Tolima)

	Piso térnico	Potencialidad de suelos para riego y drenaje	Destino económico
Correctly classified instances	87,65 %	74,06 %	98,89 %
Incorrectly classified instances	12,35 %	25,94 %	1,11 %

Con respecto a los algoritmos de regresión lineal y no lineal (tabla 5), estos se utilizaron para realizar predicciones utilizando los datos de partida considerados. En este sentido, resulta fundamental tener un conjunto de datos de calidad, completo y con un amplio recubrimiento temporal, ya que un conjunto de datos con estas características producirá resultados más confiables en la implementación de los algoritmos.

Tabla 5. Resumen de resultados del algoritmo regresión lineal y no lineal

	Regresión lineal				Regresión no lineal	
	Prueba 1	Prueba 2	Prueba 3	Prueba 4	Prueba 1	Prueba 2
Attributes	12	11	21	21	21	21
Correlation coefficient	0,793	0,7868	0,7042	0,8787	0,9997	0,9998
Localización	Ortega	Ortega	Ortega	Ibagué	Ortega	Ibagué

Los trabajos exploratorios realizados con el algoritmo de regresión lineal muestran que la implementación de este algoritmo puede resultar de gran utilidad para lograr mecanismos predictivos sobre los precios del mercado de tierras en Colombia. En las pruebas realizadas con el algoritmo de regresión lineal, resumidas en la tabla 5, las pruebas 1 y 2 se realizaron con variables numéricas y nominales, mientras que las pruebas 3 y 4 únicamente consideraron variables numéricas para el establecimiento del modelo lineal predictor del valor del avalúo.

Por otro lado, la implementación del algoritmo de regresión no lineal sobre los datos considerados sirvió para poner de manifiesto que los datos del municipio de Ortega presentan mayor diversidad que los registrados en Ibagué, consecuencia de que en el primer caso (Ortega), la implementación de este algoritmo produjo 37 modelos lineales, mientras en el segundo (Ibagué) resultaron 12 modelos. Los diferentes modelos lineales resultantes sirven para predecir resultados a cada uno de los niveles establecidos por el algoritmo, con lo que, utilizados con datos de calidad, resultan de gran utilidad y confiabilidad para predecir a diferentes niveles de detalles.

Finalmente, el algoritmo de *clustering* (*k-means*) se utiliza en este trabajo exploratorio para agrupar conjuntos de instancias con características similares, lo cual puede utilizarse para establecer políticas públicas, por ejemplo, de intervención ante la identificación de una serie de características comunes que afecten a una zona o sector y que, por tanto, precisen un cambio de la situación.

En la tabla 6 se recoge el detalle resultante de la implementación del mencionado algoritmo tomando como referencia diferentes agrupamientos (2, 5 y 10 clústeres).

Tabla 6. Resumen de resultados del algoritmo *k-means*

Localización	K=2	K=5	K=10
Ortega	58.184 (39 %)	18.494 (12 %)	17.116 (11 %)
	91.816 (61 %)	47.582 (32 %)	17.650 (12 %)
		37.318 (25 %)	9256 (6 %)
		23.265 (16 %)	9974 (7 %)
		23.341 (16 %)	16.637 (11 %)
			20.619 (14 %)
			12.310 (8 %)
			12.012 (8 %)
			25.861 (17 %)
			8565 (6 %)
Ibagué	85.582 (57 %)	34.173 (23 %)	19.381 (13 %)
	64.418 (43 %)	54.275 (36 %)	10.833 (7 %)

	22.941 (15 %)	18.043 (12 %)
	18.511 (12 %)	12.288 (8 %)
	20.100 (13 %)	16.130 (11 %)
		27.230 (18 %)
		11.789 (8 %)
		13.698 (9 %)
		8837 (6 %)
		11.771 (8 %)

3.3. Visualización

En el paso final del trabajo exploratorio con los algoritmos de *machine learning*, se presentaron diversas visualizaciones de los resultados obtenidos y se recogió la interpretación de los mismos.

De forma particular, en este artículo se recoge la visualización de dos resultados producto de la implementación del algoritmo de árboles de decisión (figuras 6 y 7) sobre la variable pendiente. Así, en el caso del árbol de decisiones del municipio de Ortega presentado en la figura 6, se observa que para llegar a la clasificación de los diferentes niveles de pendiente se parte de la variable piso térmico, la cual a través de los diferentes pisos (cálido, templado y frío) clasifica los datos, para pasar a un segundo nivel que toma como referencia el código del mapa agroclimático. Conforme a los diversos códigos presentes en los datos aparecen las primeras clasificaciones de pendientes (por ejemplo, cuando aparece el código del mapa agroclimático superior a 65, la pendiente está entre el 50 y el 75 %; cuando se recogen códigos inferiores al 65, la pendiente se encuentra entre el 25 y el 50 %). Adicionalmente, en algunos casos, el árbol de decisiones considera la variable clase agrológica, asociada con el código del mapa agroclimático, para incorporar un nivel de clasificación adicional y llegar a la pendiente correspondiente, producto de la clasificación del conjunto de las instancias. Por el contrario, en la figura 7, la clasificación mostrada por el árbol de decisiones parte de la variable clase agrológica en los datos del municipio de Ibagué, la cual sirve

para mostrar un primer nivel de clasificación de la información, que se combina en un segundo nivel con las variables categoría predominante rendimiento hídrico y código mapa agroclimático. La combinación de estas variables permite llegar a diversos valores de pendiente. Cuando esto no resulta suficiente, el algoritmo incluye la variable piso térmico para continuar la clasificación de los datos considerados y llegar a los niveles de pendiente, que fue la variable de referencia considerada para la implementación del algoritmo.

Por otro lado, con respecto a la regresión lineal, se calculó una predicción del valor de los avalúos en el municipio de Ibagué (Tolima) hasta 2025, mostrado en la tabla 7, producto de la aplicación del siguiente modelo de regresión lineal:

$$\text{Avalúo rural} = 242.988.866.472,5279 * \text{año} - 488.448.007.618.918,5024$$

Tabla 7. Cálculo de proyección-avalúos de Ibagué

MUNICIPIO	AÑO	AVALÚO RURAL (billones de \$)
Ibagué	2025	3,60
Ibagué	2024	3,36
Ibagué	2023	3,12
Ibagué	2022	2,88
Ibagué	2021	2,63
Ibagué	2020	2,39
Ibagué	2019	2,15
Ibagué	2018	1,90
Ibagué	2017	1,66
Ibagué	2016	1,42
Ibagué	2015	0,98
Ibagué	2014	0,98
Ibagué	2013	1,10
Ibagué	2012	0,25
Ibagué	2011	0,13

Finalmente, con el objetivo de reflejar la importancia de la visualización de resultados y, dentro de ella, la pertinencia de aplicar saturación (inclusión de ruido) en la visualización de los mismos, se presentan las figuras 4 y 5. En la figura 4 se muestra el agrupamiento de instancias considerando la variable avalúo (eje X) y área de terreno (eje Y). Como se puede observar, en la figura mencionada, destaca la presencia de instancias de los clústeres 1 (color rojo) y 9 (color gris). Esta figura muestra como las propiedades se caracterizan por estar concentradas en bajos valores de avalúos y con escasas dimensiones.

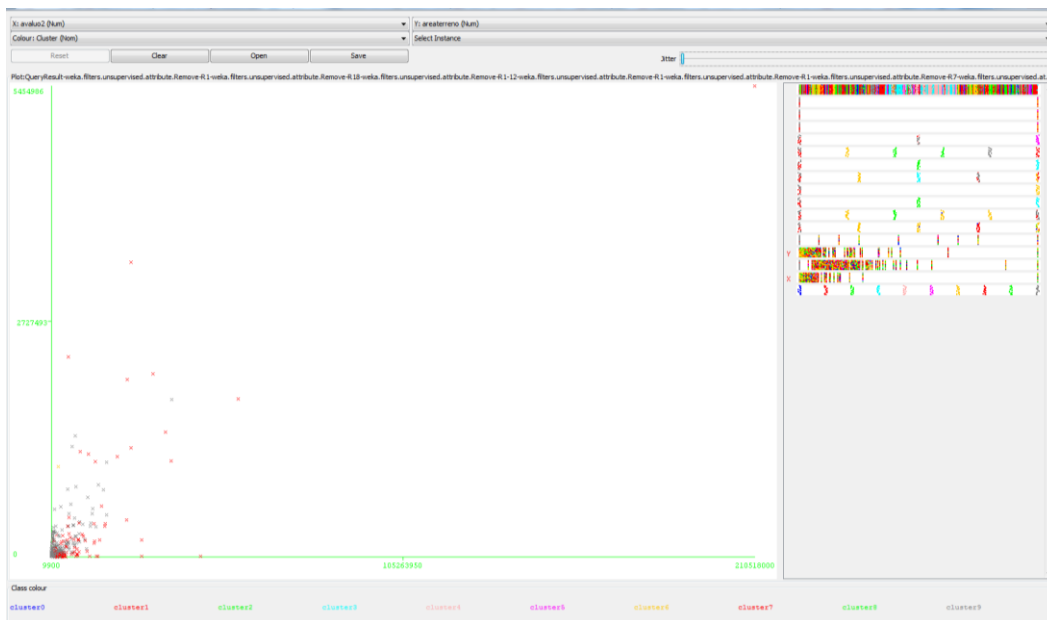


Figura 4. K-means (k=10)-avalúo=X; área de terreno=Y-Ortega (Tolima)

Sin embargo, al añadir ruido a la imagen resultante (figura 5), se comprueba una mayor presencia de los diferentes clústeres, caracterizados por bajos avalúos, pero donde el tamaño de terreno se ve incrementado.

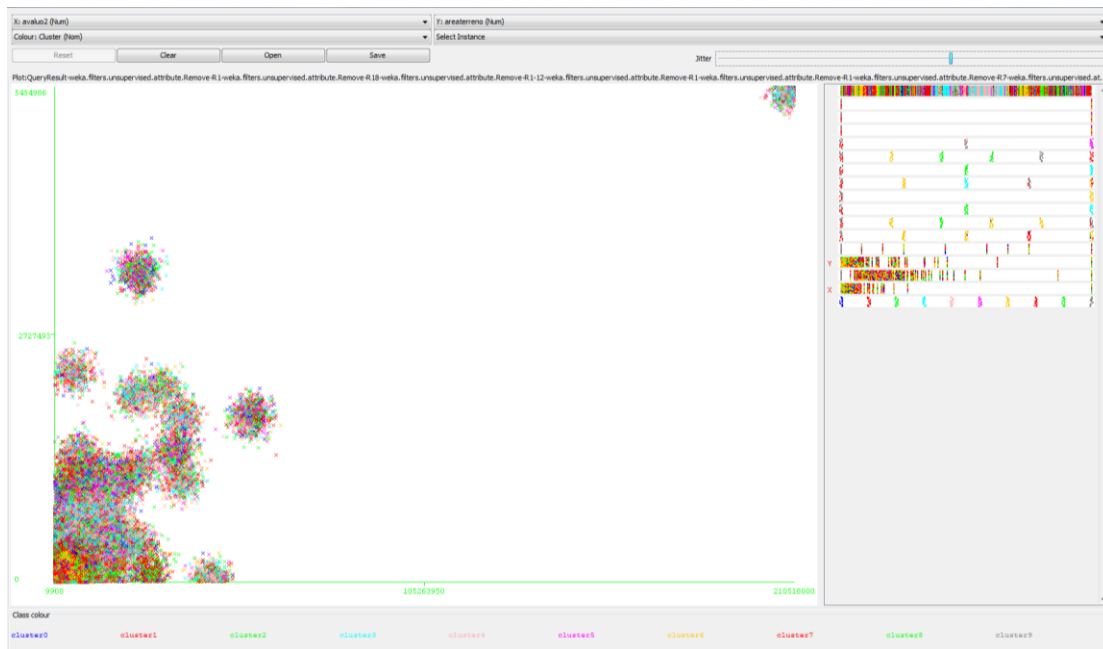


Figura 5. K-means (k=10)-avalúo=X; área de terreno=Y (con ruido)-Ortega (Tolima)

4. Conclusiones

El trabajo descrito en este artículo recoge la actividad exploratoria realizada para comprobar la aplicabilidad de algoritmos de *machine learning* en el conjunto de datos con los que el grupo de Mercado de Tierras de la UPRA interactúa de forma habitual.

De esta manera, se aborda un importante trabajo de preprocesamiento de los datos originales, modelado a través de la implementación de algoritmos de *machine learning* y, finalmente, la visualización de diversos resultados. El objetivo de estas actividades es permitir una aproximación a un mejor entendimiento y comprensión de las relaciones y patrones implícitos en los conjuntos de datos asociados con las dinámicas del mercado de tierras rurales en Colombia. No obstante, se resalta que este trabajo es exploratorio y experimental y, por tanto, los resultados y estimaciones que se presentan no son representativos de la realidad

nacional ni regional, por lo que deben entenderse como indicativos y no concluyentes.

Profundizando en el detalle del trabajo realizado, la actividad de preprocesamiento permitió la limpieza y normalización de diversos conjuntos de datos originales, así como la identificación de diferentes problemas con respecto a la calidad de las fuentes de información consideradas. Además, este trabajo permitió la creación de una base de datos (con datos normalizados y limpios, a pesar de la persistencia de problemas de calidad), lo que hizo posible acceder y consultar de manera conjunta múltiples variables.

Con respecto a la actividad de modelado, se implementaron diversos algoritmos agrupados en distintas categorías: clasificación (árbol de decisión y *Naïve Bayes*), regresión (lineal y no-lineal) y *clustering* (*k-means*). Estos algoritmos se probaron con datos y variables de los municipios de Ortega e Ibagué (Tolima).

Finalmente, en este documento se recogen diversas visualizaciones de los resultados de los algoritmos implementados con el propósito de ayudar a la comprensión de los resultados obtenidos.

Bibliografía

Allan, C. y Kerr, S. (2014). «Examining Patterns in and Drivers Rural Land Value. Motu Working Paper DRAFT». En *Motu Economic and Public Policy Research*, no. 19. Nueva Zelanda.

Bailey, I. y Lawson, A. (2015). «Market Survey». En *UK Agricultural Land 2015*. Savills World Research, UK Rural.

Baland, J. y Robinson, J. (2012). «The Political Value of Land: Political Reform and Land Prices in Chile». En *American Journal of Political Science*, vol. 56, no. 3, pp. 601-619.

Caballero, Y. (2006). «Valorización de la tierra rural en Colombia». En *Civilizar. Ciencias Sociales y Humanas*, no. 10. ISSN: 1657-8953.

Choumert, J. y Phélinas, P. (2015). «Determinants of agricultural land values in Argentina». En *Ecological Economics*, vol. 110, pp. 134-140.

Cuesta, H. (2013). *Practical Data Analysis*. Birmingham: Packt. ISBN 978-1-78328-099-5.

Dasso, J.; Shilling, J. y Ring, A. (1995). *Real Estate*. 12.ª ed. Nueva Jersey: Englewood Cliffs, Prentice Hall.

Davies, L.; Mortensen, K. y Patton, D. (2007). «Leasing land: calculating a rental». En *PRIMEFACT: Profitable & Sustainable Primary Industries*. Primefacts 338.

Davis, M. (2009). «The price and quantity of land by legal form of organization in the United States». En *Regional Science and Urban Economics*, no. 39, vol. 3, pp. 350-359.

Davis, M. y Heathcote, J. (2007). «The price and quantity of residential land in the United States». En *Journal of Monetary Economics*, no. 54, vol. 8, pp. 2595-2620.

Delahaye, O. (1997). «Renta y mercado de la tierra agrícola: algunas indicaciones de los casos venezolano y chileno». En *FAO on the Internet (1998)*. Recuperado de <http://www.nzdl.org/gsdImod?e=d-00000-00---off-0fi1998--00-0---0-10-0---0---0direct-10-4-----0-11--11-en-50---20-about---00-0-1-00-0--4---0-0-11-10-0utfZz-8-00&a=d&c=fi1998&cl=CL1.1&d=HASH0e7f2a00be129311060edd>.

Domingo, E. y Fulleros, R. (2005). «Real estate price index: a model for the Philippines. In Real Estate Indicators and Financial Stability». En *Proceedings of a Joint Conference Organised by the Bank for International Settlements and the International Monetary Fund*.

El Mercurio. (1956-1960). Chile.

Haughwout, A.; Orr, J. y Bedoll, D. (2008). «The price of land in the New York metropolitan area». En *Federal Reserve Bank of New York: Current Issues in Economics and Finance*, no. 14, vol. 3.

Kohavi, R. y Provost, F. (1998). «Glossary of terms». En *Machine Learning*, no. 30, pp. 271-274.

Larson, W. (2015). «New Estimates of Value of Land of the United States Paperback». En *Bureau of Economics Analysis*. ISBN-10: 1511723777.

Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente. (s. f.). *Precios medios anuales de las tierras de uso agrario: informe metodológico estandarizado*. España: Mapama. Recuperado de http://www.magrama.gob.es/es/estadistica/temas/estadisticas-agrarias/01044_PRECIOS_MEDIOS_ANUALES_DE_LAS_TIERRAS_DE_USO_AGRARIO_tcm7-349732.pdf.

Ministerio de Agricultura, Alimentación y Medio Ambiente. (s. f.). *Precios medios anuales de las tierras de uso agrario (metodología)*. España: Magrama. Recuperado de http://www.magrama.gob.es/es/estadistica/temas/estadisticas-agrarias/preciosdelatierrametodologiaparaeurostatmy2016_tcm7-201078.pdf.

Muñoz, J. y Cardona, H. (2013). «El precio de la tierra: estado del arte de las metodologías de valoración de predios rurales y su aplicación en Colombia». En *Suma de Negocios*, vol. 4, no. 1. ISSN: 2215-910X.

Nichols, J.; Oliner, S. y Mulhall, M. (2013). «Swings in commercial and residential land prices in the United States». En *Journal of Urban Economics*, 73, pp. 57-76.

Nickerson, C.; Morehart, M.; Kuethe, T.; Beckman, J.; Ifft, J. y Williams, R. (2012). «Trends in U.S. Farmland Values and Ownership». En *Economic Research Service Economic Information Bulletin*, no. 92. United States Department of Agriculture.

Osano, P.; Rouget, M.; Turpie, J.; Thuiller, W. y Balmford, A. (2011). «Estimating Land Prices and Opportunity Costs of Conservation in a Megadiversity Country». En *African Technology Policy Studies Network*, no. 58. ISBN: 978-9966-030-10-8.

Palacio, V.; Montesillo, J. y Santacruz de León, E. (2007). «Análisis del mercado de tierras en México». En *Revista Textual* (Chapingo), vol. 49, pp. 1-40.

Rey, D.; Lagos, L.; Romero, O. y Forero, O. (2015). *Determinación de los elementos a tener en cuenta en la elaboración de una zonificación de precios de la tierra rural en Colombia escala 1:100.000* (sic). Documentación de la construcción de los elementos por tener en cuenta para la elaboración de la zonificación de precios. Versión 8. Unidad de Planificación Rural Agropecuaria (UPRA).

Reydon, B.P.; Agurto Plata, L.E.; Sparovek, G.; Burstein Goldszmidt, R.G.; Santos Telles, T. (2014). «Determination and forecast of agricultural land prices». En *Nova*, vol. 24, no. 2, pp. 389-408. ISSN 0103-6351. Recuperado de <<http://dx.doi.org/10.1590/0103-6351/1304>>.

Rosen, S. (1974). «Hedonic prices and implicit markets: product differentiation in pure competition». En *Journal of Political Economy*, vol. 82, no. 1, pp. 34-55.

Samuel, A. (1959). «Some Studies in Machine Learning Using the Game of Checkers». En *IBM Journal*, vol. 3, no. 3, pp. 210-229. doi:10.1147/rd.33.0210.

Smola, A. y Vishwanathan, S. (2010). *Introduction to Machine Learning*. Cambridge: Cambridge University Press. ISBN 0 521 82583 0.

Soto, R. (2005). *El precio de mercado de la tierra desde la perspectiva económica*. Santiago de Chile: Cepal.

Soto, F. y Gómez, S. (Eds.). (2012). *Dinámica del mercado de la tierra en América Latina y el Caribe: concentración y extranjerización*. Santiago: FAO. ISBN 978-92-5-307117-3.

USAID - U.S. Agency for International Development. (2010). *Ecuador: property rights and resource governance profile 1*. Recuperado de < https://www.land-links.org/wp-content/uploads/2016/09/USAID_Land_Tenure_Ecuador_Profile.pdf >

Verheye, W. (s. f.). «The value and price of land». En *Land use, land cover and soil sciences, vol. III, Encyclopedia of Life Support Systems (EOLSS)*.

Vilches-Blázquez, L.M. (2016). *Identificación de las técnicas complejas aplicables a la definición de precios del mercado de tierras rurales agropecuarias en Colombia*. Versión 4. Unidad de Planificación Rural Agropecuaria (UPRA).

5. ANEXO

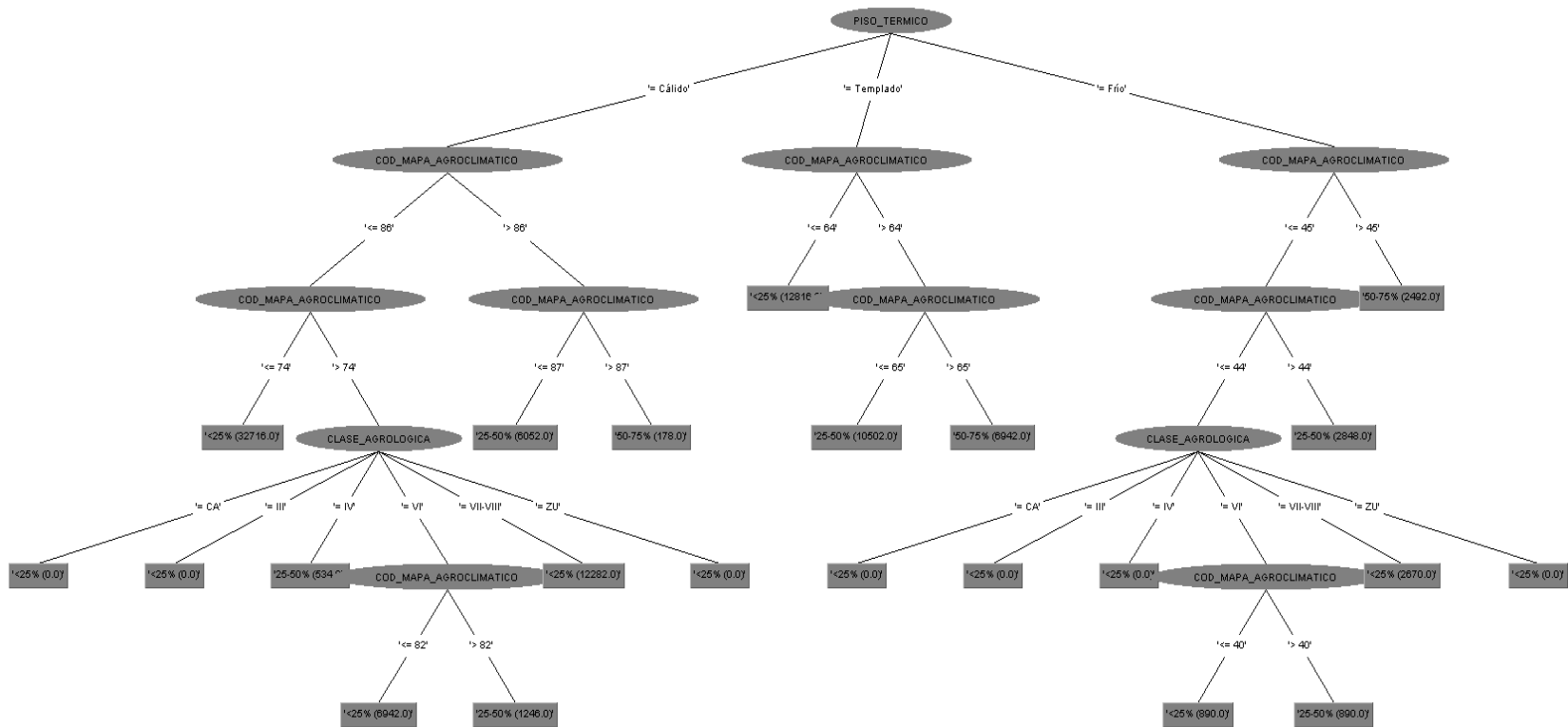


Figura 6. Árbol de decisión de la variable pendiente en Ortega (Tolima)

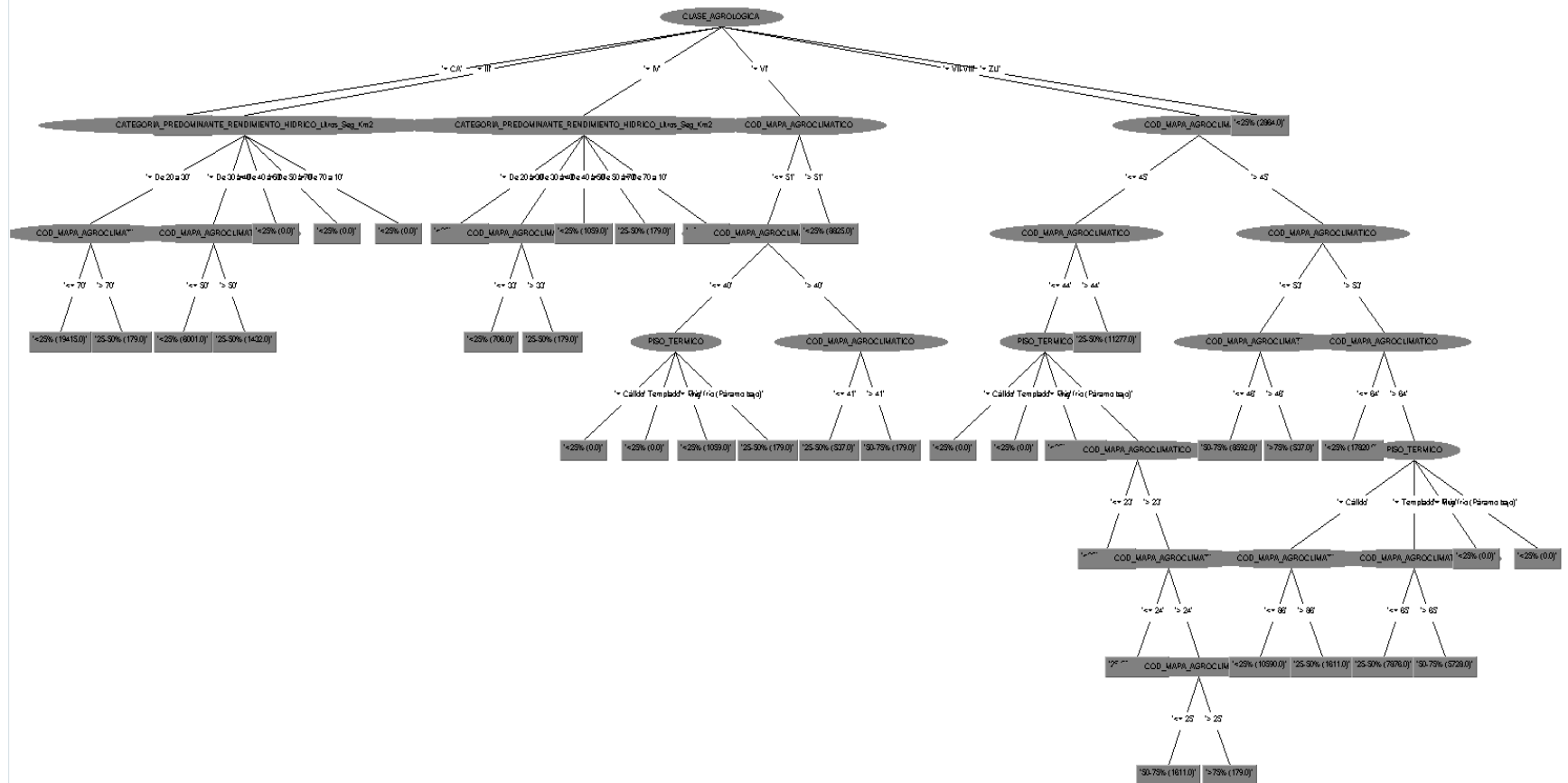


Figura 7. Árbol de decisión de la variable pendiente en Ibagué (Tolima)