

**Usos y aplicaciones de técnicas de Machine Learning en procesos de gestión catastral
multipropósito**

Daniel Armando Rosero Hernández

Asesor

Edith Johana Morales Liberato

Universidad Nacional Abierta y a Distancia UNAD
Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI
Especialización en Ciencia de Datos y Analítica
2024

Resumen

Esta investigación analiza el potencial del Machine Learning (ML) para optimizar la gestión del catastro multipropósito en Colombia, donde actualmente aproximadamente solo el 9.4% del territorio nacional cuenta con información catastral actualizada, cifra lejana de la meta del 70% establecida en el Plan Nacional de Desarrollo 2022-2026. A través de una revisión bibliográfica exhaustiva, se examinan las aplicaciones del ML en los componentes físico, jurídico y económico del catastro, considerando experiencias internacionales y el contexto colombiano. La investigación se estructura en tres ejes: la descripción de técnicas de ML aplicables al ámbito catastral, el análisis de su potencial contribución en la agilización de procesos, y la propuesta de estrategias para su implementación efectiva. Los resultados sugieren que el ML ofrece oportunidades significativas para automatizar tareas complejas, mejorar la precisión en la identificación de la dinámica inmobiliaria y potenciar la eficiencia en su valoración, contribuyendo así a la modernización del sistema catastral colombiano y al cumplimiento de las metas nacionales de actualización catastral.

Palabras claves: gestión catastral, machine learning.

Abstract

This research analyzes the potential of Machine Learning (ML) to optimize the management of multipurpose cadastre in Colombia, where currently approximately only 9.4% of the national territory has updated cadastral information, a figure far from the 70% goal established in the National Development Plan 2022-2026. Through an exhaustive literature review, ML applications are examined in the physical, legal, and economic components of the cadastre, considering international experiences and the Colombian context. The research is structured along three axes: the description of ML techniques applicable to the cadastral field, the analysis of their potential contribution to streamlining processes, and the proposal of strategies for their effective implementation. The results suggest that ML offers significant opportunities to automate complex tasks, improve accuracy in identifying real estate dynamics, and enhance efficiency in property valuation, thus contributing to the modernization of the Colombian cadastral system and the fulfillment of national cadastral update goals.

Keywords: cadastral management, machine learning.

Tabla de Contenido

Introducción	8
Planteamiento del Problema	10
Justificación	12
Objetivos	14
Objetivo General	14
Objetivos Específicos	14
Marco Conceptual	15
Catastro Multipropósito	15
Aprendizaje Automático	22
Machine Learning en los Componentes del Catastro	27
ML dentro del Componente Físico	27
ML dentro del Componente Jurídico	30
ML dentro del Componente Económico	31
El ML en la Operatividad del Catastro	34
Etapa Preoperativa	35
Caracterización Territorial	36
Diagnostico Catastral	37
Plan de Intervención	40
Etapa Operativa	41
Reconocimiento Predial y Ajuste Cartográfico	42
Barrido Jurídico	48
Consolidación Económica	51

Etapa Postoperativa.....	52
Implementación del Machine Learning en la Gestión Catastral Multipropósito.....	56
Evaluación de las Necesidades.....	56
Adquisición y Preparación de Datos.....	56
Selección y Entrenamiento de Modelos de Machine Learning.....	57
Validación e Integración.....	58
Capacitación del Personal.....	58
Evaluación de Riesgos y Desafíos Éticos.....	60
Conclusiones.....	62
Referencias.....	65

Lista de Figuras

Figura 1 <i>Normatividad Implementación del Catastro Multipropósito.</i>	15
Figura 2 <i>Componentes Esenciales del Catastro Multipropósito.</i>	16
Figura 3 <i>Categorías de Modelos LADM perfil Colombiano</i>	18
Figura 4 <i>Modelo Extendido Catastro Registro.</i>	21
Figura 5 <i>Flujo de Aprendizaje Supervisado.</i>	24
Figura 6 <i>Flujo de Aprendizaje No Supervisado.</i>	25
Figura 7 <i>Actividades de las Etapas de Formación y Actualización Catastral Multipropósito.</i> ..	35
Figura 8 <i>Ficha Reconocimiento Predial IGAC.</i>	44
Figura 9 <i>Detección de Construcciones con Técnicas de ML.</i>	45
Figura 10 <i>Comparación de Cartografía o Vectores Actuales y Redes Neuronales Convolucionales.</i>	46
Figura 11 <i>Aplicación de Método SAM para Generación de Vectores.</i>	47
Figura 12 <i>Ejemplo Información Catastral Histórica.</i>	50

Lista de Tablas

Tabla 1 <i>Variables Catastrales Generales Modelo Antiguo</i>	19
--	----

Introducción

El catastro multipropósito representa uno de los pilares fundamentales para el desarrollo y ordenamiento territorial de Colombia, transformando la concepción tradicional del catastro como una herramienta meramente fiscal hacia un sistema integral que soporte la toma de decisiones en múltiples dimensiones de la gestión territorial. Esta transformación, impulsada desde los acuerdos de paz y materializada en diversas políticas públicas, busca consolidar un censo predial completo y actualizado que sirva como base para la planificación y el desarrollo sostenible del país.

Sin embargo, las cifras actuales revelan un desafío significativo, aproximadamente solo el 9.4% del territorio nacional cuenta con un catastro actualizado bajo el enfoque multipropósito, muy lejos de la meta del 70% establecida en el Plan Nacional de Desarrollo 2022-2026. Esta brecha no solo representa un rezago en términos de la gestión administrativa, sino que tiene implicaciones directas en la capacidad del estado para implementar políticas efectivas de ordenamiento territorial, gestión fiscal y reforma rural integral.

En este contexto, las tecnologías emergentes, particularmente el Machine Learning (ML), surgen como una alternativa prometedora para optimizar los procesos de gestión catastral. El ML, con su capacidad para procesar grandes volúmenes de datos y automatizar tareas complejas, ofrece posibilidades que sugieren agilizar la actualización catastral, mejorar la precisión en la identificación de cambios prediales y potenciar la eficiencia en la correcta identificación y valoración de inmuebles.

La presente investigación se centra en analizar cómo las técnicas de ML pueden contribuir a optimizar los procesos de gestión catastral multipropósito en Colombia. A través de una revisión bibliográfica exhaustiva, se exploran las aplicaciones potenciales del ML en los

componentes físico, jurídico y económico del catastro, considerando tanto las experiencias internacionales como las particularidades del contexto colombiano.

El documento está estructurado en tres ejes principales, primero, una descripción detallada de las técnicas de ML aplicables al ámbito catastral; segundo, un análisis de su potencial contribución en la agilización de los procesos de gestión catastral; y tercero, una propuesta de estrategias para su implementación efectiva. Esta estructura permite abordar tanto los aspectos técnicos como los desafíos prácticos de la implementación del ML en el contexto catastral colombiano.

La relevancia de este estudio radica en su potencial para contribuir a la modernización del sistema catastral colombiano, proporcionando una hoja de ruta para la incorporación de tecnologías avanzadas que permitan alcanzar las metas de actualización establecidas. Además, el análisis considera no solo los aspectos técnicos del ML, sino también las implicaciones organizacionales, éticas y sociales de su implementación en el contexto específico del catastro multipropósito.

En un momento donde la transformación digital y la modernización de la gestión administrativa del estado son cruciales para el desarrollo del país, este estudio busca aportar al entendimiento de cómo las nuevas tecnologías pueden contribuir a superar uno de los desafíos más significativos en la gestión territorial, la actualización y mantenimiento de un catastro multipropósito eficiente que refleje la realidad inmobiliaria de los territorios.

Planteamiento del Problema

Actualmente en Colombia se está implementando la política pública de catastro multipropósito (CM), la cual busca transformar el catastro tradicional, enfocado como una herramienta fiscal, a un censo que sirva para múltiples propósitos, el cual permita un adecuado ordenamiento territorial y efectiva gobernanza basada en datos, desde la institucionalidad se han adelantado esfuerzos para que este cambio de enfoque se convierta en realidad, un avance relevante de esta estrategia es la consolidación de la norma ISO 19152 al perfil colombiano, la cual define el modelo de administración del territorio y la estructuración de la información relacionada, con el objetivo de hacerla interoperable y eficiente.

Toda la implementación de esta nueva política lleva algunos años en desarrollo, su planteamiento en el contexto del país viene desde los acuerdos de paz, donde se pone en evidencia la necesidad de implementar un catastro con múltiples propósitos para garantizar el ordenamiento de la propiedad rural, siendo este, un instrumento para minimizar los conflictos del país (Gobierno Nacional & FARC-EP, 2016), desde ese momento histórico se han realizado adecuaciones institucionales que han llevado a que hoy se esté implementando el CM de manera gradual, habilitando gestores catastrales para que ejecuten de manera descentralizada la denominada gestión catastral multipropósito (Gobierno Nacional, 2019).

La gestión catastral multipropósito tiene como objetivo mantener actualizada la base de datos catastral en las diferentes etapas inmersas, como son formación, actualización y conservación catastral, esto ha resultado desafiante tanto para la cabeza de sector, como es el IGAC, como para los gestores y operadores catastrales, muestra de esto son las recientes cifras del porcentaje de actualización catastral en las que se encuentra el país, según el director del IGAC, actualmente el país se encuentra actualizado solo en un 9,4 % (Marulanda G., 2023).

Entendiendo que actualmente la desactualización catastral, implica serios problemas en materia fiscal, social, política y de seguridad, es pertinente plantear que se deben analizar nuevas alternativas que permitan aumentar los esfuerzos y contribuir a la mayor eficiencia en los procesos de actualización catastral. Implementación de nuevas tecnologías, automatización de procesos y ajuste de metodologías pueden ser elementos que contribuyan a optimizar estos procesos.

Actualmente las técnicas basadas en aprendizaje automático tienen una notoria visibilidad en distintos campos, su capacidad de asimilar procedimientos ha hecho que su aplicación sea posible en cualquier campo que cuente con información para procesar, la habilidad de estos sistemas para realizar tareas específicas de manera autónoma permite dentro de este contexto preguntarse, ¿Cómo pueden las técnicas de Machine Learning (ML) contribuir a optimizar los procesos de la gestión catastral multipropósito?

Justificación

El bajo porcentaje de actualización catastral en el país implica un serio problema a la hora de tomar decisiones en ordenamiento e inversión territorial, adicionalmente pero no menos importante, tiene implicaciones en uno de los puntos más relevantes de los acuerdos de paz, que es la reforma rural integral, la cual depende de un censo claro sobre la realidad de tierras del país, dicho censo corresponde a la actualización catastral (Gonzalez J., 2023). Sin este insumo valioso será difícil dar un dictamen sobre las dimensiones que requiere dicha reforma.

En materia catastral se ha determinado que la cifra más reciente de porcentaje de actualización catastral con enfoque multipropósito corresponde a un 9,4% (Marulanda G., 2023), según el plan nacional de desarrollo 2022-2026 la meta es alcanzar el 70 % de actualización (Gobierno Nacional, 2023), meta que actualmente se encuentra muy lejana. Entendiendo este rezago y que el incumplimiento de las mismas tiene implicaciones en todos los sectores del país, hace aún más notoria la necesidad de que se brinde una discusión sobre el uso de nuevas metodologías y herramientas que puedan contribuir a cumplir con esa meta.

Según estimaciones de la dirección del IGAC el costo de la actualización catastral del país ronda los 4 billones de pesos (Marulanda G., 2023), de los cuales ya se tiene aproximadamente el 27 % comprometido para dicho fin en proyectos de corto y mediano plazo. Considerando la importancia para el país y el compromiso de diferentes fuentes de financiación para llevar a cabo proyectos destinados a actualizar el territorio, permite sentar una línea base sobre la cual se puede brindar un panorama de la implementación de herramientas de vanguardia como el uso de técnicas de ML.

Actualmente el ML nos permite generar herramientas que aprenden de manera autónoma para llevar a cabo una tarea específica, siendo estas ampliamente replicables en aplicaciones que

cuentan con información recurrente o estructurada, adicionalmente es importante mencionar que dentro del ámbito catastral se manejan bases de datos de gran tamaño que contienen diferentes tipos de datos, según cifras del DANE para 2018 existían en Colombia 13.480.729 predios hoy la cifra se estima en más de 16 millones de predios.

El volumen de información asociada al ejercicio de esta actividad depende del área a la que se limite, por ejemplo, si se trabaja con la base de datos catastral del distrito de Bogotá estamos hablando de información alfanumérica, fotografías, documentos y demás, asociadas a cada uno de los 2.587.000 predios del distrito capital. Entendiendo las ventajas que ofrecen técnicas de ML para el procesamiento de grandes cantidades de datos y a su vez generar conocimiento y valor agregado al conjunto de datos, resulta en una alternativa al menos interesante de explorar.

Adicionalmente el ámbito inmobiliario se mantiene en constante cambio, requiere de una actualización continua en sus componentes físico, jurídico y económico, paralelamente los modelos de ML por sus características intrínsecas usan el gran volumen de datos para aprender de ellos y a su vez adaptarse rápidamente a cambios en sus diferentes componentes, actualmente el DNP ha publicado avances en implementación de técnicas de ML para captura de información útil en procesos de actualización catastral (DNP, 2023), pero si se analiza el grado de desactualización claramente los esfuerzos deben aumentarse.

En este estudio se pretende brindar un panorama de los posibles usos y aplicaciones de técnicas de ML en cada uno de los procesos inmersos dentro de la gestión catastral con enfoque multipropósito, esta revisión bibliográfica pretende dar puntos de partida que resulten en proyectos aplicados que contribuyan a la optimización de estos procesos, aportando de manera significativa a la adopción de estas nuevas tecnologías en el ámbito catastral.

Objetivos

Objetivo General

Analizar los usos y aplicaciones de técnicas de Machine Learning en los procesos de gestión catastral multipropósito para optimizar la formación, actualización y conservación de la información catastral.

Objetivos Específicos

Describir las técnicas de Machine Learning aplicables a la gestión catastral multipropósito, identificando sus características y funcionalidades principales.

Examinar la contribución de las técnicas de Machine Learning en la agilización de los procesos de actualización catastral, identificando ventajas y limitaciones.

Proponer estrategias para la implementación efectiva de técnicas de Machine Learning en la gestión catastral multipropósito.

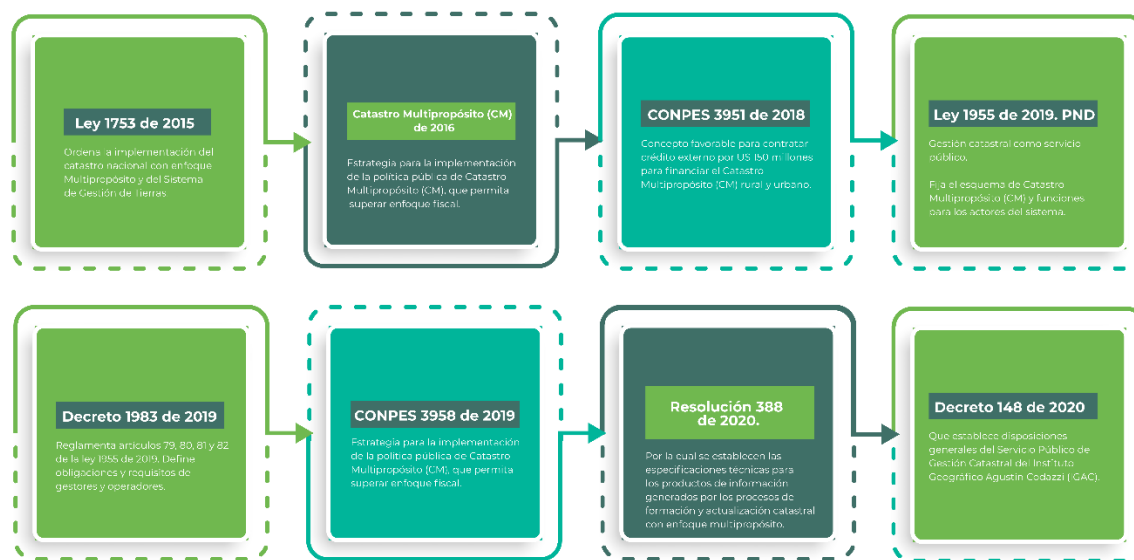
Marco Conceptual

Catastro Multipropósito

Como ya se expuso anteriormente el catastro multipropósito es un eje fundamental para el desarrollo del país, pero no se ha definido explícitamente que conlleva este concepto, una serie de instrumentos normativos permitió consolidar lo que hoy es el catastro multipropósito (Figura 1), el decreto 148 de 2020 lo define en dos instancias, el que hace y el para que, en cuanto al que hace, se puede resumir como el censo de bienes inmuebles a lo largo y extenso del territorio nacional, el para que ya se ha abordado previamente, pero podemos concebirlo de manera general como el insumo fundamental para el adecuado desarrollo y ordenamiento del territorio (Decreto 148 de 2020). Detrás de estas descripciones generales se desglosa el trasfondo técnico y el impacto de su implementación en las diversas políticas públicas. Dentro de este apartado se pretende describir estas especificidades.

Figura 1

Normatividad Implementación del Catastro Multipropósito.

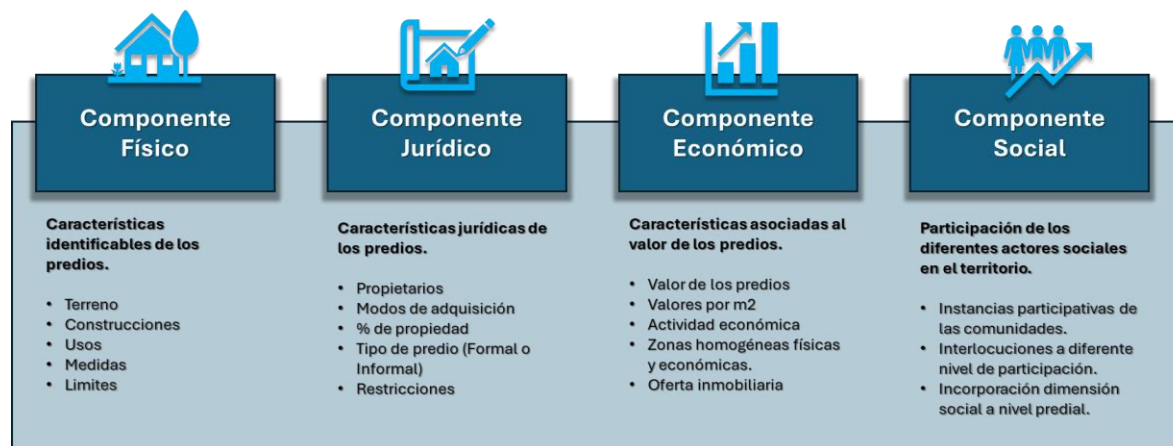


Nota. Línea de Tiempo Normativa Catastro Multipropósito (Catastro Bogotá, 2020)

Catastro es el censo de bienes inmuebles en su componente físico, jurídico y económico, este de estar debidamente actualizado, actualmente se habla de dos catastros, catastro tradicional y catastro multipropósito, el primero tiene un enfoque fiscal, basado netamente en el recaudo a partir del impuesto predial unificado, mientras que el multipropósito busca darle dimensiones adicionales y robustecer el uso de la información catastral en temas de seguridad jurídica en la tenencia de tierra, eficiencia del mercado inmobiliario, desarrollo y ordenamiento del territorio, y demás aplicaciones que potencialicen la mejor toma de decisiones en la asignación de recursos públicos y el fortalecimiento fiscal de los territorios. Adicionalmente es importante resaltar que dentro del enfoque multipropósito se añade la dimensión social, esta busca fortalecer el censo a partir de la efectiva participación de la comunidad (Resolución 1040, 2023).

Figura 2

Componentes Esenciales del Catastro Multipropósito.



Nota. Componentes esenciales para la adecuada gestión multipropósito. Elaboración propia

Dentro de la gestión catastral se enmarcan 3 procesos fundamentales, la formación, actualización y conservación catastral, estos buscan mantener lo mayor actualizada posible la base de datos catastral, la primera se ejecuta cuando en el territorio no existe ninguna

información de base, en este sentido se debe formar la base catastral tanto geográfica como alfanumérica desde cero, en contraparte la actualización catastral, tal como su nombre lo indica, busca actualizar territorios que ya han sido formados y que se encuentran desactualizados con una anterioridad mayor a 5 años (Ley 223 de 1995).

Finalmente, la conservación catastral es la actualización continua de la base de datos catastral, la cual incorpora modificaciones que se realizan en periodos de tiempo más cortos, las cuales generalmente provienen de solicitudes realizadas por usuarios o propietarios, dentro de este proceso se enmarcan las denominadas mutaciones catastrales bajo las cuales se realizan modificaciones específicas que actualizan uno o más predios en uno o más componentes del catastro (Decreto 148, 2020).

Dentro de estos 3 procesos fundamentales y con la adopción del nuevo enfoque multipropósito, emergen 3 métodos que permiten capturar información del territorio, como son, métodos directos, que son los que requieren necesariamente inspección in situ para recolectar la información predial, los métodos indirectos que permiten hacer uso de herramientas a distancia como sensores remotos, bases de datos, modelos, big data y fuentes secundarias para adquirir información y finalmente los métodos declarativos y colaborativos, que pretenden recopilar información mediante la participación de la comunidad (Decreto 148 de 2020, 2020).

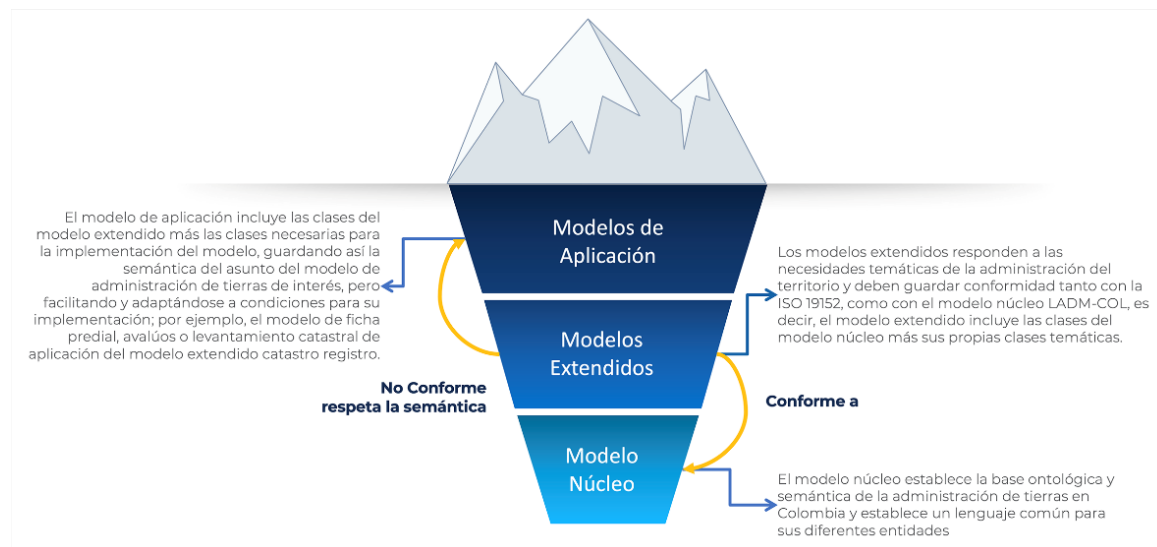
Adicionalmente dentro de la nueva gestión catastral multipropósito se consolida el modelo de administración del territorio para el territorio colombiano, tomando como referencia el estándar ISO 19152 y adoptándolo a la realidad nacional, este modelo permite que la información relacionada a catastro tenga una misma estructura y pueda ser interoperable y consumible, de forma que se le pueda dar un uso efectivo, a su vez mediante los modelos

extendidos y de aplicación se puede adaptar a las necesidades específicas de cada espacio geográfico (IGAC & SNR, 2024).

La adopción de este nuevo modelo de estructuración de la información catastral también sugiere un reto, pues su implementación requiere de una transferencia de conocimiento de las cabezas de sector público hacia los distintos actores que producen y gestionan esta información. En este sentido se habla de un grado de desactualización catastral, pero se deja un poco de lado territorios que se encuentran actualizados, pero que tienen una brecha importante en infraestructura y capacidades para ajustar la información al nuevo modelo de administración del territorio.

Figura 3

Categorías de Modelos LADM perfil Colombiano



Nota. Definición de los modelos de datos para la gestión catastral multipropósito (ICDE, 2023)

Dentro de los procesos de actualización catastral con enfoque multipropósito se deben levantar por los métodos disponibles normativamente una serie de variables que en su conjunto permiten consolidar la base de datos catastral actualizada. Actualmente con el cambio de

paradigma en el catastro tradicional, se han agregado nuevos elementos que se deben recopilar en los procesos de barrido predial. Por esto es importante que se realice un comparativo del estado actual a nivel de información para municipios que no han adoptado el nuevo enfoque multipropósito y cuáles serían los elementos nuevos para incorporar acorde al modelo LADM.

El antiguo modelo de base de datos catastral dispuesta por el Instituto Geográfico Agustín Codazzi se componía de dos tablas principales, la tabla R1 y la tabla R2, estas contenían toda la información general necesaria para describir las características catastrales alfanuméricas de un predio, estas claramente no contemplaban el componente espacial, es decir la representación geográfica de cada predio en su conjunto terreno y construcción que permiten localizar un predio. Algunas de estas variables tienen relaciones uno a muchos, es decir un predio puede tener múltiples elementos de una variable específica, o sencillamente relaciones uno a uno. En la Tabla 1 se enlistan las 24 variables principales que contempla una base de datos catastral. Este modelo anterior no contemplaba un modelo óptimo de estructuración de la información, es también una de las razones de la necesidad en la adopción del nuevo modelo LADM.

Tabla 1

Variables Catastrales Generales Modelo Antiguo.

Variable	Descripción
Departamento	Identificador del departamento en el que se ubica
Municipio	Identificador del municipio en el que se ubica
Numero_Predial	Identificador único del predio
Direccion	Dirección de ubicación del predio
Destino_Economico	Actividad económica predominante
Area_Terreno	Área del lote o terreno
Area_Construida	Área total construida
Numero_Predial_Anterior	Identificador único anterior del predio
Nombre Propietario	Nombre de la persona relacionada
Estado Civil	Estado civil de la persona relacionada

Variable	Descripción
Tipo Documento	Tipo de documento de identificación la persona relacionada
No Documento	No de documento de la persona relacionada
Avaluo	Valor catastral del predio en su conjunto terreno y construcción
Vigencia	Año de vigencia de la información
Zona_Fisica	Identificador de la zona homogénea física asociada
Zona_Economica	Identificador de la zona homogénea geoeconómica asociada
Puntaje	Puntaje o calificación asociada a la construcción
Habitaciones	Número de habitaciones
Baños	Numero de baños
Locales	Número de locales
Pisos	Número de pisos de la construcción
Condición	Tipo de condición en la que se encuentra
Usos	Identificador único usos de la construcción
Matricula Inmobiliaria	Numero de matrícula inmobiliaria para predios formales

Nota. Variables generales dentro de una base de datos catastral. Elaboración propia

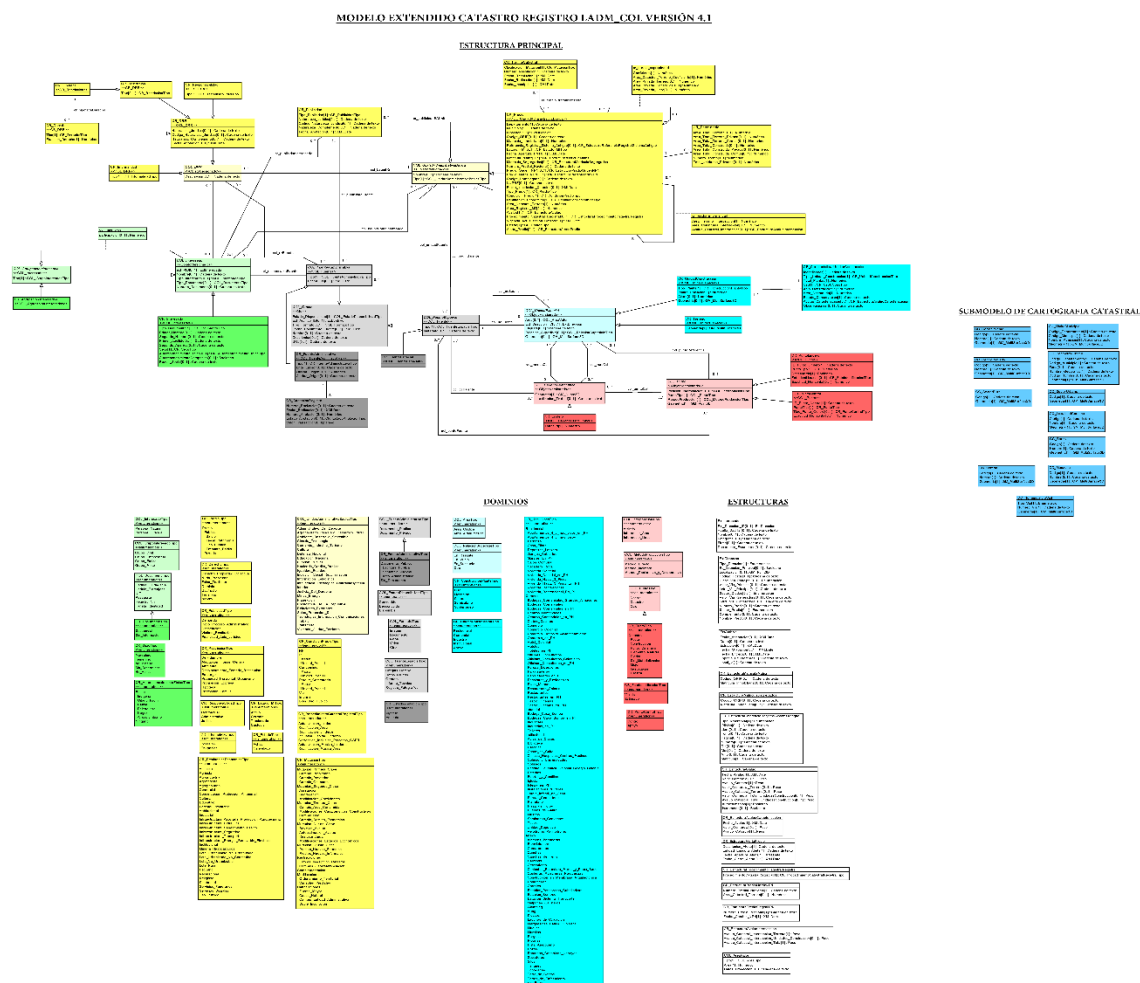
Ahora bien con la definición del nuevo modelo de administración de la información catastral, se habla de un modelo relacional que contempla múltiples capas, las cuales se estructuran conforme a la norma ISO 19152, a partir de la cual se generan modelos extendidos y de aplicación que permiten hacer uso de esta información a las diferentes entidades nacionales. Desde la puesta en vigencia de esta nueva forma de estructurar y hacer interoperable la información se han generado múltiples modelos para los fines misionales de cada entidad, para el caso específico de este estudio se contemplará 2 modelos esenciales, el modelo extendido de catastro registro y el modelo de levantamiento catastral. Claramente la discusión no se centrará en su consolidación en sí, más bien en las variables que contempla y como pueden acorde a su naturaleza ser recopiladas o procesadas de una manera más eficiente por técnicas de ML.

El último modelo vigente extendido de catastro registro que es a partir del cual se generan los modelos de aplicación que están algo más aterrizados a la realidad del ejercicio de recopilación de información in situ, contempla más de 30 clases y más de 100 atributos dentro de

ellas, a su vez estos elementos deben ceñirse a una estructura y dominios específicos (Figura 4). Esto muestra que el cambio de paradigma en el ámbito catastral no solo implica la concepción en sí, sino también un desafío en materia de implementación tecnológica para el adecuado gestionamiento de la información de este tipo.

Figura 4

Modelo Extendido Catastro Registro.



Nota. Modelo extendido catastro registro para la interrelación de información catastral y registral en el país (ICDE, 2024) – para más detalle ver anexo 1

Ahora dimensionemos este volumen de información, el cual debe recopilarse, actualizarse o conservarse a nivel de predio en los 1.123 municipios del territorio nacional. Se estima que en los 114 millones de hectáreas que abarca el país existen más de 16 millones de predios. Esto sugiere un desafío operativo y tecnológico de proporciones considerables, en las secciones posteriores intentaremos vislumbrar técnicas de vanguardia tecnológica que contribuyan a esta labor.

Aprendizaje Automático

Actualmente existen múltiples herramientas y áreas de investigación que permiten hacer uso de grandes conjuntos de datos y aún más importante aprender de ellos, optimizando su recopilación, mantenimiento y valor añadido. El aprendizaje automático (ML) es un campo en el cual se plantea que los sistemas aprendan de manera autónoma sin específicamente requerir una programación o instrucción explícita, revelando patrones y características que tradicionalmente pueden ser complejas de identificar (Caltech, 2024).

Dentro del ML podemos marcar 3 grandes subgrupos de aprendizaje, el supervisado, no supervisado y el aprendizaje por refuerzo. El supervisado se enfoca en trabajar con datos que ya tienen una clasificación y bajo los cuales se pretende generar modelos predictivos de alta eficiencia (MIT, 2020). En contraste el aprendizaje no supervisado es óptimo para trabajar con datos no clasificados o sobre los que se busca identificar, patrones, características o grupo propios de los datos sin proporcionar de manera específica lo que se quiere identificar (Caltech, 2024).

Adicional a estos el aprendizaje por refuerzo busca generar un aprendizaje optimizado basado en la maximización de la recompensa, actuando recurrentemente en distintos escenarios

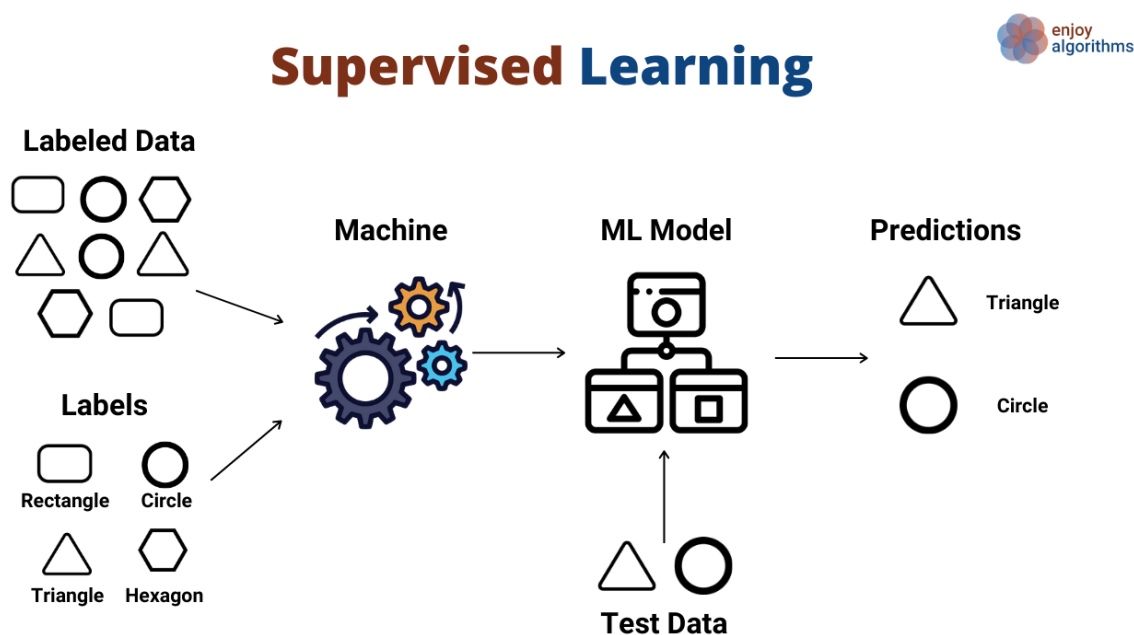
hasta que el sistema de recompensas logre ser el más óptimo, resultando en decisiones más eficientes acorde al escenario y entorno planteados (Sutton & Barto, 2015).

El aprendizaje supervisado es una de las técnicas más utilizadas en Machine Learning. Trabaja con un conjunto de datos previamente etiquetado, lo que significa que el modelo aprende a partir de ejemplos conocidos para predecir futuros resultados. Dentro de este subgrupo se encuentran técnicas como la regresión lineal, regresión logística, árboles de decisión y máquinas de soporte vectorial (SVM).

- **Regresión Lineal y Logística:** La regresión lineal se usa principalmente para predecir valores continuos, mientras que la regresión logística es ideal para la clasificación binaria, recurrentemente usada para clasificar la ocurrencia o no de un evento específico (Osisanwo et al., 2017).
- **Árboles de Decisión y Random Forest:** Los árboles de decisión permiten crear reglas de decisión que segmentan los datos en diferentes grupos. Random Forest combina múltiples árboles de decisión, mejora la precisión y evita el sobreajuste, siendo útil en la predicción del valor de propiedades (Sisman & Aydinoglu, 2022b)
- **Máquinas de Soporte Vectorial (SVM):** Las SVM son técnicas poderosas para la clasificación de datos complejos, especialmente cuando los datos no son linealmente separables (Osisanwo et al., 2017).

Figura 5

Flujo de Aprendizaje Supervisado.



Nota. Flujo de funcionamiento de un proceso de aprendizaje supervisado, con sus entradas y salidas (Kozan, 2021)

El aprendizaje no supervisado se utiliza cuando no se dispone de datos etiquetados. El objetivo es descubrir patrones ocultos o estructuras subyacentes en los datos, el aprendizaje en este tipo de enfoque se basa en la identificación de patrones y relaciones entre el conjunto de datos que permiten a posterior generar una categorización sin ninguna supervisión u orientación externa a los mismos, simplemente por sus características y relacionamiento propio (Edwards et al., 2021)

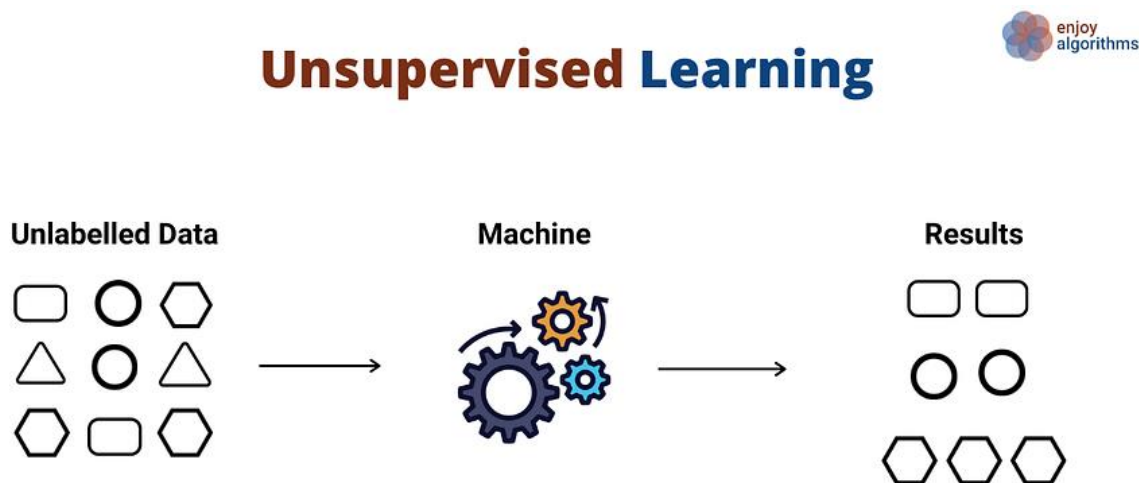
- **Clustering (Agrupamiento):** Una de las técnicas más conocidas es el clustering, la cual permite agrupar información que guarda cierto grado de relación, generando cluster o grupos que pueden revelar patrones relevantes del contexto de los datos. Existen muchos

algoritmos considerados de clustering dentro de los cuales se puede encontrar K-means, clustering aglomerativo y DBSCAN, estos agrupan pero sus diferencias se pueden asociar de forma resumida a lo siguiente, K-means necesita un número fijo de clusters, el clustering aglomerativo genera una jerarquía de clusters, y DBSCAN es muy bueno para detectar clusters de densidad y manejar datos ruidosos o con formas complejas. (Müller & Guido, 2016)

- Análisis de Componentes Principales (PCA): PCA es una técnica de reducción de dimensionalidad que facilita el análisis de grandes volúmenes de datos, permitiendo una menor cantidad de datos que explican la mayor parte de la información contenida en ellos. (Müller & Guido, 2016).

Figura 6

Flujo de Aprendizaje No Supervisado.



Nota. Flujo de funcionamiento de un proceso de aprendizaje supervisado, con sus entradas y salidas (Kozan, 2021)

A diferencia de los otros subgrupos, el aprendizaje por refuerzo es particularmente útil en situaciones en las que se debe tomar una serie de decisiones secuenciales, este tipo de

aprendizaje generar una serie de mecanismos de interacción que le permiten maximizar la recompensa para tomar una decisión adecuada para un problema concreto. (J. S. Park & Park, 2020)

- Q-learning: Es una técnica de aprendizaje por reforzamiento que se utiliza tablas de valor para encontrar la mejor acción a tomar, dada la situación actual (Jang et al., 2019).
- Aprendizaje Profundo por Reforzamiento (Deep Reinforcement Learning): Combina redes neuronales profundas con técnicas de aprendizaje por reforzamiento para manejar entornos complejos y maximizar la eficiencia de la estrategia para tomar una decisión específica (Hester et al., 2018).

Machine Learning en los Componentes del Catastro

Para el presente estudio se ha desarrollado una revisión bibliográfica que busca establecer de manera inicial las posibilidades de aplicación del ML en los distintos procesos inmersos en la gestión catastral multipropósito. Para poder brindar una mirada clara del aporte específico en cada proceso, este documento plantea describir las técnicas de ML aplicables al campo catastral, examinar su aporte específico en los distintos procesos y componentes, y finalmente proponer unas estrategias para que su implementación pueda darse de una forma efectiva. En este apartado se presentan algunos referentes en materia, enfocando su relacionamiento a la aplicabilidad al contexto catastral colombiano, en este caso diferenciado por los componentes esenciales del catastro, físico, jurídico y económico.

ML dentro del Componente Físico

El componente físico puede asociarse al compendio de todos los elementos que son plenamente identificable en terreno, de manera general suelen identificarse por métodos directos, dentro de este grupo podemos encontrar la definición de linderos, que delimitan un predio de otro, ubicación, construcciones, nomenclaturas, etc. Para este estudio dentro de este componente también se contempla toda la información cartográfica, a pesar de que sea una representación del componente físico se toma como un elemento más del mismo.

La literatura muestra que existe una amplia variedad de investigaciones asociadas a variables que hacen parte de este componente, uno de los elementos principales del componente físico es la caracterización de construcciones, García-Retuerta et al. (2018) presentan una alternativa a métodos convencionales basada en redes neuronales convolucionales, partiendo de una base de datos de imágenes sin previa clasificación, los resultados de su estudio mostraron una alta precisión superior al 85%, evidenciando una alta aplicabilidad en clasificación de

construcciones basados en imágenes, Potsiou et al. (2020) presenta un enfoque basado en aplicativos móviles que mapean en 3D los interiores de las construcciones, permitiendo extraer la mayor cantidad de características de las mismas, usando diferentes técnica de ML y Deep Learning para el posicionamiento y reconstrucción de escenas, esto para ser utilizadas en una clasificación automatizada o para una revisión con soporte humano.

En muchas ocasiones la información catastral está asociada a documentos históricos los cuales datan de muchos años atrás o que simplemente no se encuentran en las mejores condiciones, como pueden ser mapas, planos, escrituras, documentos de compraventa y demás elementos que permiten identificar características esenciales de los predios. El ML también tiene aplicaciones para ayudar a comprender de mejor manera este tipo de información, el Deep Learning surge como el método más idóneo, tanto para digitalizar la información (Ignjatić et al., 2018; Mango et al., 2023) o extrayendo de forma masiva elementos esenciales para su comprensión (Oliveira et al., 2017, 2019).

Una amplia gama de estudios está enfocada en la aplicación de técnicas de ML para la representación o corrección de datos del componente físico, como pueden ser linderos, divisiones, construcciones y demás, combinado técnicas de SIG y de digitalización automática, una parte importante dentro de la consolidación de la información física en un proceso de actualización catastral es la conformación de zonas homogéneas físicas, las cuales como su nombre lo indica buscan agrupar porciones de terreno con características similares, Bastidas Méndez & Bonilla Bautista (2018) plantean el uso de árboles de decisión para la determinación de estas a partir elementos físicos como vías, topografía, uso de suelo, disponibilidad de servicios públicos, etc. Los resultados mostraron una buena precisión y una notoria mejoría en términos de tiempo, lo cual es crucial en este tipo de procesos.

Dentro de los procesos de formación y actualización catastral es necesario consolidar una cartografía robusta que represente la realidad, para cumplir con este objetivo los proyectos requieren de una inversión costosa en términos de tiempo y personal. Múltiples estudios muestran las bondades del ML en la extracción de características del terreno, como pueden ser linderos, construcciones o elementos identificables mediante imágenes satelitales o fotografías aéreas tomadas con aeronaves y drones. Fetai et al. (2022) & Hajiheidari et al.(2024) exponen como el uso de técnica de ML pueden potenciar la actualización de cartografía existente y la correcta delimitación del territorio, Yuan et al. (2021) también pone en evidencia el potencial de técnicas de Deep Learning para la segmentación de los distintos elementos del paisaje.

Dentro del ámbito nacional es sumamente relevante mencionar que una causa de la desactualización catastral del país, son las disputas de conflicto armando en las regiones más rurales del territorio, sin acceso a estas zonas cobra aún más sentido el uso de métodos indirectos para obtener una primer mirada con una adecuada precisión de la realidad de esas zonas dispersas y que tanto requieren una adecuada atención desde el estado, Lee & de Vries (2020) muestran un ejemplo de cómo se puede usar herramientas remotas para identificar las relaciones de tenencia de la tierra en zonas poco accesibles, usando información remota combinada con técnicas de ML se pueden obtener resultados valiosos para una adecuada intervención desde los organismos responsables.

La realidad catastral del país muestra que se requiere de una constante y acelerada renovación que garantice una efectiva administración de la información catastral, actualmente el país maneja un catastro representable en 2 dimensiones, catastros más avanzados muestran la necesaria transición a un catastro de 3 dimensiones que contemple de mejor manera la realidad del territorio específicamente en las construcciones, si bien se está lejos de este punto, es

pertinente poner en evidencia vistas a futuro de esta transformación, Park & Guldman (2019) muestran como una representación de la ciudad en 3 dimensiones permite una adecuada administración del territorio, como a partir de esta se puede dimensionar de otra manera el ordenamiento de los territorios y como el catastro como elemento fundamental puede transformarse para dicho fin.

ML dentro del Componente Jurídico

El componente jurídico del catastro puede entenderse como el rubro que brinda seguridad jurídica a todos los procesos catastrales, desde la inscripción de predios en las diferentes condiciones de propiedad, posesión o tenencia, y en la identificación de anomalías en cuanto al debido y correcto registro de los predios en la base de datos catastral. Este componente tiene como insumo principal la normativa catastral vigente, y la diversidad de documentos que soportan la propiedad de un predio, los cuales pueden consistir en escrituras, cartas reales, sentencias, adjudicación y demás títulos que acrediten dicha condición.

Entendiendo que la amplia documentación que debe manejar este componente por lo general viene de manera física y que posteriormente se digitaliza, implica un manejo de información abrumadora, que en términos del desempeño humano está sujeta a la revisión puntual de cada documento. En este apartado se exponen algunos estudios basados en manipulación de documentos mediante técnicas de ML.

Como se mencionó anteriormente el componente jurídico requiere analizar una gran cantidad de documentos, Zhao et al. (2019) exponen una metodología basada en redes neuronales convolucionales para extraer información de documentos que tienen una estructura medianamente definida, su estudio se enfoca en documentos tipo recibos, sobre los cuales se extrae información de clientes, valores, etc. Llevando esto al ámbito catastral, muchas veces es

necesario extraer información de escrituras o certificados de libertad y tradición, si a estos procesos tradicionales es factible aplicarles esta tecnología, se traduciría en una reducción de tiempo sustancial.

bin Abdullah & Iqbal (2022) proponen una metodología basada en procesamiento de lenguaje natural (NLP) y reconocimiento óptico de caracteres (OCR) para el análisis de documentos tanto estructurados como no estructurados, permitiendo usar no solo documentos si no archivos de imagen, dando mayor adaptabilidad a diferentes tipos de entrada, adicionalmente se resalta la capacidad para generar interpretación precisas del contenido de los mismos, una de las desventajas identificadas de este estudio es la disminución de la precisión al trabajar con datos no estructurados, como alternativa Devlin et al. (2018) obtienen mejores resultados con este tipo de datos, arrojando altas precisiones por el mejor entrenamiento de procesamiento de lenguaje natural, también obtiene ventajas en el análisis contextual de la información, ya que usa capas enfocadas en la semántica y el análisis de sentimientos, estos estudios sugieren un apoyo significativo para la mejor toma de decisiones basadas en un gran volumen de datos.

Una parte importante del manejo de gran volumen de información es la disposición de esta bajo una estructura de gestión documental, el método propuesto por Omurca et al. (2023) puede ser útil para la automatización en la clasificación de documentos y su posterior disposición según una estructura, la cual puede complementarse con variaciones de algoritmos de ML (W. Yu et al., 2020) para aportar en todo el proceso de gestión, ordenamiento y disposición del expediente documental.

ML dentro del Componente Económico

Las técnicas de ML han mostrado tener una gran aplicabilidad en el manejo de variables asociadas al componente económico, la valoración de inmuebles mediante el uso de minería de

datos (Y. Yu et al., 2021), redes neuronales difusas (Guan et al., 2014), clustering multivariante espacialmente restringido (Sisman & Aydinoglu, 2022b) y random forest (Carranza et al., 2019), permiten generar predicciones valuatorias de manera efectiva con un amplio conjunto de datos, esto es sumamente importante dentro de procesos de formación y actualización catastral ya que en estas etapas se configura de manera inicial el marco económico de la región objeto de implementación y a su vez permite considerar múltiples variables que representen de la manera más amplia posible el valor final, tanto m² para la definición de zonas homogéneas geoeconómicas como para la asignación de avalúos puntuales predio a predio.

En el marco del componente económico del catastro, la aplicación de técnicas avanzadas de ML y análisis de datos geospaciales se ha destacado como un factor crucial para mejorar la precisión y eficiencia de la valoración masiva de bienes inmuebles. Diversos estudios han demostrado la eficacia de integrar sistemas de información geográfica (SIG) con métodos de ML para desarrollar modelos de valuación más precisos. Por ejemplo, Mo & Yomralioglu T. (2022) han desarrollado una metodología híbrida que combina SIG y ML para la valuación automática de predios, mostrando una mejora significativa en la precisión al incluir factores de ubicación en los modelos de precios. También se encontraron buenos resultado aplicando un preprocesamiento exhaustivo de los datos, en conjunto con técnicas de clustering espacial, mostrando mejores agrupaciones, teniendo en cuenta el componente de georreferenciación de cada una de las instancias (Sisman & Aydinoglu, 2022).

El uso de nuevos modelos ha demostrado ser una alternativa más precisa a los modelos tradicionales, Guan et al. (2014) aplican un modelo neuro-fuzzy adaptativo, combinado con técnicas de suavizado exponencial cuadrático para la estimación de precios, los resultados evidenciaron una muy eficiente adaptabilidad del modelo a las múltiples variaciones del mercado

inmobiliario. Que el modelo pueda identificar y manejar esta variabilidad es sumamente útil para el ámbito catastral, ya que permite mantener una predicción efectiva, garantizando que la valuación refleje la realidad del mercado.

Estudios adicionales han explorado el uso de aprendizaje profundo y regresiones ponderadas geográficamente para ajustar las evaluaciones catastrales a las características específicas locales, lo que resulta en valuaciones más precisas y con contextos más enriquecedores (Sisman & Aydinoglu, 2022a). La incorporación de datos tridimensionales y el uso de técnicas de análisis de big data son ejemplos de cómo la tecnología de ML puede ofrecer una visión más detallada y completa, crucial para la valoración en el contexto del catastro multipropósito (Kara et al., 2020).

De manera general las técnicas de Machine Learning son ampliamente usadas en la valoración de la tierra, estas aplicaciones son claramente aplicables al ámbito catastral, cada una aporta de manera diferenciada ventajas y desventajas, ofrecen diferentes rendimientos según la disponibilidad y calidad de los datos, Carranza et al. (2019) pone en evidencia la capacidad de ofrecer resultados masivos lo cual sería muy útil para procesos de actualización, el uso de técnicas de big data también pueden ser opciones óptimas para realizar procesos masivos (Sisman & Aydinoglu, 2022b).

Esta revisión busca tener una vista clara de cómo se puede aplicar de manera efectiva las diferentes herramientas o técnicas de ML para cada uno de los componentes del catastro, estos aportes individuales a cada uno de estos se agruparán acorde a las etapas de formación, actualización y conservación catastral, buscando que la revisión sirva como una hoja de ruta de cómo se pueden aplicar estas nuevas tecnologías dependiendo de la etapa en la que se encuentre un proyecto de aplicación.

El ML en la Operatividad del Catastro

Dentro de la gestión catastral multipropósito podemos relacionar 3 procesos fundamentales que permiten consolidar una base de datos catastral, para los procesos de formación, actualización y conservación catastral, todas estas etapas requieren de manera general una recopilación de información, un procesamiento de esta y una toma de decisiones basada en las 2 etapas anteriores. En ese sentido, en este apartado se abordarán actividades que habitualmente se realizan en los 3 procesos de la gestión catastral multipropósito, quizás haciendo más énfasis en la formación y actualización catastral, debido a la masividad de estas y la puntualidad que requiere la conservación catastral.

Dentro de los procesos de formación y actualización catastral se deben realizar barridos prediales que recopilen todas las variables necesarias para consolidar y actualizar una base de datos catastral, ahora con el enfoque multipropósito este levantamiento de información sugiere ser más preciso y por tanto más desafiante si se ve desde la operatividad, dicha operatividad tiene dos grandes etapas preoperativa, operativa y postoperativa. En la etapa preoperativa se deben recopilar y procesar diferentes fuentes de datos que permiten focalizar de mejor manera las labores en territorio. La operativa concierne netamente al desarrollo de labores planificadas en la etapa anterior y en la postoperativa de la consolidación de la información, presentación de resultados y demás mecanismos que permitan hacer interoperable los productos del proceso.

Esto sugiere un desarrollo de múltiples actividades que van desde la caracterización inicial del territorio a censar, un diagnóstico de lo esperado en el territorio a partir de los insumos disponibles, la socialización de este diagnóstico y los requerimientos de información que sugiere el mismo, la definición de unas unidades de intervención que dividan el territorio global en áreas con características similares o que por operatividad sugieran trabajarse en conjunto, la puesta en

marcha de esta recopilación a partir de los diversos métodos, la consolidación de esta información que a posteriori surta el componente económico, resultando en el valor catastral de los predios y adicionalmente todas las instancias de socialización de avances parciales y finalización del proyecto, la Figura 7 muestra un flujo de las actividades dentro de las diferentes etapas operativas.

Figura 7

Actividades de las Etapas de Formación y Actualización Catastral Multipropósito.



Nota. Procesos y actividades en proyectos de gestión catastral multipropósito (DNP, 2020)

Etapa Preoperativa

Esta etapa es crucial para procesos de formación y actualización catastral, pues permite establecer una línea base de las características del territorio en materia general y un diagnóstico del estado inmobiliario del área geográfica de implementación, generalmente esta etapa permite consolidar la estrategia de recopilación de información y los demás mecanismos que permitirán la adecuada conformación de una base de datos catastral actualizada. Para procesos de

conservación esta etapa no aplica pues como ya se mencionó estos obedecen a actividades más puntuales.

Se pueden establecer de manera general 3 productos principales de esta etapa, 1. La caracterización territorial, 2. El diagnóstico catastral y 3. La definición del plan de intervención. La caracterización territorial permite establecer las características geográficas, socioeconómicas, demográficas, de seguridad y demás componentes relevantes para garantizar la efectiva intervención catastral. El Diagnóstico catastral hace referencia al estado de la base de datos catastral vigente, para los casos que requieren formación el diagnóstico está enfocado en la dinámica inmobiliaria no consolidada de manera catastral, esto permite establecer un marco de referencia de la cantidad de predios a identificar y las diferentes condiciones de estos.

Finalmente, el plan de intervención toma como insumo la caracterización y el diagnóstico, ponderando los hallazgos de estos dos instrumentos se define por operatividad, características inmobiliarias, logística y seguridad, para definir unidades de intervención que dividen el territorio global en zonas que preservan características similares que permiten ser abordadas de manera conjunta.

Caracterización Territorial

Este producto consiste en describir las características del territorio en los componentes geográfico, socioeconómico, demográfico, de seguridad y otros que permiten establecer oportunidades y desafíos en el territorio, para conseguir esto es necesario recopilar una gran cantidad de información, procesarla e interpretarla dentro del contexto propio estableciendo las relaciones o implicaciones en materia catastral. En este sentido el ML presenta oportunidades para la gestión y procesamiento de todos estos insumos.

Para obtener este producto se debe analizar un sinnúmero de información y documentos, estos desde la vista de ML podrían procesarse mediante técnicas de procesamiento de lenguaje natural - PLN, la cual permite obtener conocimiento de manera ágil cuando se tiene volúmenes de información que exceden la capacidad humana, Wu et al. (2022) exponen las posibilidades de las técnicas de ML para extraer conocimiento valiosos de datos en formato de texto, muy aplicables a un contexto de construcción de un documento de caracterización territorial.

Usar modelos de lenguaje natural para analizar y resumir normativas legales, documentos históricos o estudios técnicos previos son los elementos promisorios para agilizar la consecución de este tipo de productos. Estos modelos pueden ayudar a extraer automáticamente la información más relevante para consolidar la caracterización y resaltar los aspectos críticos de la normativa que deben ser tenidos en cuenta.

Diagnostico Catastral

El diagnostico catastral este asociado a la evaluación de la base de datos catastral del municipio conforme a la normatividad catastral vigente, esto es, la adecuación de esta en función de cumplir los criterios de calidad y consistencia para los componentes físico, jurídico y económico. Dentro de esto también se contempla la necesidad, si es el caso, de implementar el nuevo modelo de catastro multipropósito.

En esta etapa se identifican los hallazgos del grado de desactualización de en materia catastral para el municipio objeto de implementación, para los casos de formación será la conformación de cero de la base de datos catastral, para los casos de actualización los elementos clave a identificar pueden ser, la cantidad de predios nuevos, predios con cambios en los componentes físico, jurídico y económico, cantidad de informalidades, predios en condición de

propiedad horizontal y claramente los predios que no presentan cambios y que per se no requieren una inspección puntual.

Para identificar estos posibles cambios, se hace uso de dos mecanismos de análisis, 1. El cruce de bases de datos, reportes, información geográfica, satelital y demás que permita clasificar desde oficina estos cambios, 2. Inspecciones en campo, estas tradicionalmente se denominan precenso o prereconocimiento predial y consiste en enviar operarios de campo a identificar directamente posibles cambios.

Considerando estos 2 mecanismos de identificación podemos resaltar las posibilidades del ML para potenciar estas actividades. El ML tiene amplias capacidades para procesos de predicción y clasificación, dentro de esta etapa esos son elementos cruciales, pues, la etapa de diagnóstico consiste y el documento en sí, es la base para la adecuada planificación y ejecución operativa del proyecto, en este sentido las herramientas de predicción como métodos de regresión, y de clasificación como clúster, arboles de decisión y máquinas de soporte vectorial tienen una gran cabida, a continuación se detallan posibles aplicaciones con estos métodos.

En materia catastral las proyecciones y predicciones de cantidad de predios, zonas de expansión, valores de terreno y construcción, e incluso población pueden permitir focalizar zonas que sugieran un mayor impacto de actualización, en este sentido técnicas de regresión lineal, arboles de decisión o modelo ARIMA pueden ser herramientas bastante útiles. los datos históricos de actualizaciones catastrales, los censos de población y otros indicadores socioeconómicos, se puede construir modelos de regresión que predigan cómo crecerán ciertas áreas en términos de desarrollo inmobiliario, infraestructura y cambios en la base catastral.

En cuanto a los modelos de clasificación como Random Forest o máquinas de soporte vectorial, pueden ser de mucha utilidad si se los insuma con la gran cantidad de datos recopilada

para esta etapa. A partir de variables como el estado de la información catastral, densidad poblacional, calidad del acceso a servicios, y cambios recientes en la infraestructura, los modelos de clasificación pueden identificar cuáles zonas requieren una actualización inmediata y cuáles pueden esperar. Esto puede mejorar la planificación de recursos y enfocar los esfuerzos donde son más necesarios.

Uno de los análisis esenciales de esta etapa es la identificación de zonas que preserven características similares, en este contexto las técnicas de agrupamiento o clustering pueden sugerir una alternativa ideal, pues técnicas como K-means o clustering jerárquico permitirían a partir del uso de variables como características físicas del territorio (topografía, clima), datos demográficos, infraestructura, y estado de la información catastral, agrupar zonas con características similares. Esto permite crear planes de actualización más eficientes y ajustar las estrategias a nivel local, optimizando los recursos de acuerdo con las necesidades particulares de cada grupo.

Como se ha mencionado, la identificación de cambios en los predios es quizás el producto bandera del documento de diagnóstico, en este sentido los datos de observación terrestre, como son los provenientes de satélites y aeronaves tripuladas y no tripuladas, permiten aplicar técnicas de procesamiento de imágenes con redes neuronales convolucionales (CNN), estas permiten analizar imágenes satelitales para detectar cambios en el uso del suelo, edificaciones nuevas, o cambios en la infraestructura. Con esto, se puede generar una serie de marcas que permiten identificar zonas con mayor cantidad de cambios, que pudieran priorizarse para inspecciones más rigurosas. Dentro de esto también se puede considerar técnicas de geoestadística, como el kriging espacial, que permite interpolar zonas no cubiertas o carentes de información espacial.

Adicionalmente dentro del campo de machine learning existen métodos para identificar valores atípicos en grandes conjuntos de datos, estos son sumamente importante tanto en la etapa de diagnóstico como en la consolidación de la información actualizada. A menudo, los datos físicos, jurídicos y económicos contienen errores o inconsistencias que pueden pasar desapercibidos. Los algoritmos de detección de anomalías como Isolation Forest pueden ayudar a identificar estos problemas analizando patrones en los datos y señalando registros que difieren significativamente de lo esperado, como valores inusualmente altos o bajos en las diferentes variables catastrales. Esto facilita la corrección o validación de los datos antes de su consolidación.

Finalmente una alternativa quizás un poco más utópica, sugeriría el uso de técnicas de generación de lenguaje natural, las cuales después de analizar los datos y generar los resultados mediante las técnicas mencionadas, permitan generar automáticamente un informe con los hallazgos más importantes del diagnóstico. Esto facilita la creación de un documento de diagnóstico final, ahorrando tiempo en la redacción.

Plan de Intervención

El plan de intervención es el cómo se abordarán y recopilarán los diferentes cambios identificados en el diagnóstico, como todos los hallazgos de este se podrán consolidar en la base de datos catastral a partir de procedimientos técnicos definidos, en este particular las técnicas de ML cobran importancia en 2 elementos esenciales del plan de intervención, 1. La definición de unidades de intervención y 2. la operatividad logística para recopilar la información identificada en los hallazgos para cada una de estas unidades de intervención.

Para la definición de las unidades de intervención se puede tomar como insumo las variables compiladas y analizadas en el diagnóstico catastral, claramente aquí tienen una

relevancia sumamente importante las técnicas de agrupamiento y clasificación, estas pueden tomar una gran cantidad de variables que permitan definir porciones geográficas que puedan ser intervenidas de manera homogénea y ágil. Como variables asociadas se puede considerar cantidad de predios, topografía, densidad poblacional, condiciones de seguridad, dinámica inmobiliaria y otros, estos algoritmos pueden usar esta gran cantidad de información y generar grupos bien segmentados y representativos, lo cual desde un esfuerzo humano puede resultar en un ejercicio bastante demandante.

En cuanto al desarrollo operativo y logístico podemos considerar de manera general 2 elementos cruciales, la optimización de rutas de barrido predial y el seguimiento al desarrollo de estas actividades. En cuanto a la optimización de rutas o logística del barrido predial Algoritmos de optimización como los algoritmos genéticos o la programación lineal pueden resultar útiles, Estos pueden ser usados para diseñar rutas eficientes para la recolección de información física en campo, minimizando los tiempos y costos operativos. También pueden ayudar a asignar recursos humanos y tecnológicos en las zonas de mayor necesidad. En cuanto al seguimiento de las actividades con la integración de dispositivos móviles y herramientas de seguimiento GPS, Algoritmos de tracking y geolocalización basados en ML pueden monitorear en tiempo real la ubicación de los equipos de campo, la cobertura geográfica de la recolección de datos y el progreso del levantamiento catastral. Esto permite ajustar la intervención de manera más eficiente, redirigiendo los recursos hacia zonas que aún no han sido completamente cubiertas o que enfrentan retrasos.

Etapas Operativas

La etapa operativa es la puesta en marcha de lo definido en el plan de intervención, esta enfocada en recopilar la mayor cantidad de cambios o modificaciones que actualmente la base de

datos catastral no tiene incorporado, en esta se puede recurrir a los métodos directos, indirectos y colaborativos/declarativos. De manera general los métodos directos son los mas usados para la identificación física y económica, mientras que el componente jurídico tradicionalmente se recopila mediante métodos indirectos, esto debido a su naturaleza que requiere de soportes oficiales que validen la relación de tenencia de las personas naturales o jurídicas con los predios. Adicionalmente con la puesta en marcha del enfoque de CM, los métodos declarativos están ganando cabida en el ejercicio operativo. Dentro de esta etapa también se desarrolla de manera paralela el componente social, el cual permite un adecuado relacionamiento con la comunidad y permite la efectividad y agilidad del ejercicio técnico.

Se puede desglosar la etapa operativa en 4 actividades clave que permiten recopilar toda la información identificada en el diagnostico. 1. Reconocimiento Predial, 2. Barrido Jurídico, 3. Ajuste cartográfico y 4. Consolidación económica. Estas pueden desarrollarse de manera singular para procesos de formación y actualización, en algunos casos con diferencias entendiéndose que la formación es una recopilación más ardua por la consolidación desde cero de la base de datos catastral.

Reconocimiento Predial y Ajuste Cartográfico

Una de las actividades que mayor personal y tiempo requiere es la de reconocimiento predial, esta generalmente consiste en enviar un operario a campo a inspeccionar los predios, tomando información de terreno, construcciones, y demás información jurídica que permita establecer las relaciones de personas naturales o jurídicas con los predios, las condiciones de formalidad o informalidad del predio, y demás variables que pueden resumirse a la tabla 1. Esta labor tiene como productos principales la calificación o puntaje de las construcciones

identificadas en el predio, un bosquejo con las medidas de estas y los usos o actividades económicas desarrolladas en cada una de ellas.

Como se mencionó, habitualmente el levantamiento de esta información se realiza por métodos directos, pero no es la única alternativa, actualmente la normativa colombiana contempla la aplicación de métodos indirectos y los denominados declarativos y colaborativos. Estos permiten hacer uso de insumos secundarios que no considera per se el reconocimiento predial, como imágenes satelitales, documentos secundarios, aplicación de modelos, adicionalmente información compartida por los usuarios de manera participativa.

Las labores de reconocimiento predial suelen ser arduas, cada operario debe identificar las diversas características en los predios y posteriormente disponerlas en una estructura y formato que permita consolidarlo, en este caso acorde al modelo LADM de levantamiento catastral vigente, esta etapa de consolidación de la información suele ser crítica, pues mucha de la información se recopila en medios análogos o físicos (Figura 8), sugiriendo una posterior digitalización de la misma, que a su vez debe adecuarse al modelo de datos mencionado.

productos que se pueden implementar como son el mapeo de huellas de construcciones de Google o Microsoft, los cuales son de acceso libre y fueron obtenidos mediante técnicas de ML.

Figura 9

Detección de Construcciones con Técnicas de ML.



Nota. Detección automática de construcciones haciendo uso de técnicas de ML en el procesamiento de imágenes satelitales de BingMaps (Microsoft, 2024)

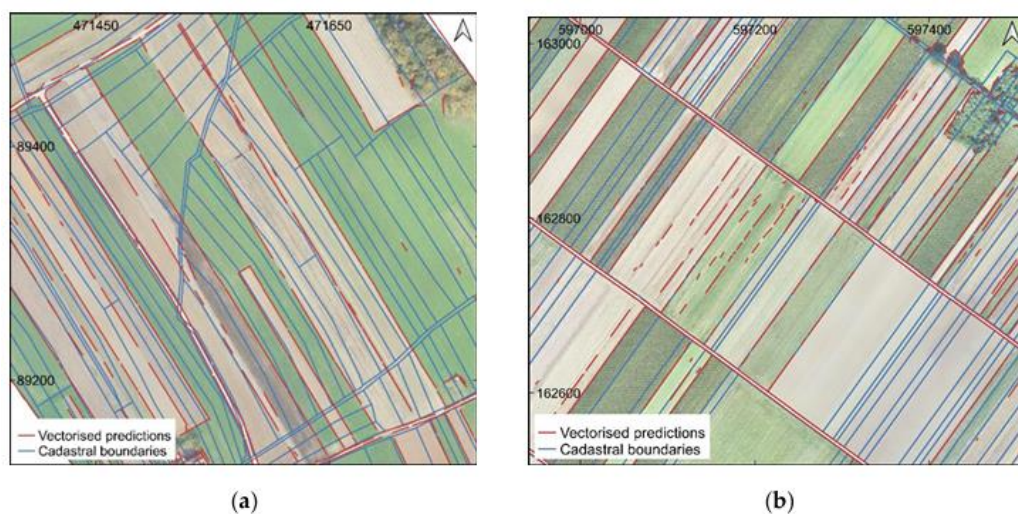
Adicionalmente existen algoritmos como el SAM que permite aplicar técnicas similares a imágenes satelitales o aéreas propias, lo cual puede ser más aplicable para este caso por la temporalidad y temas de resolución especial y la necesidad de productos como ortofotos que deben preservar unos criterios de calidad acorde a la normativa catastral, este tipo de técnicas

también sería eficiente en la extracción de linderos y generación de cartografía vectorial a partir de este tipo de insumos.

Estas técnicas de extracción automática de vectores son útiles para procesos de formación y actualización catastral, en formación para generar vectores base que permitan consolidar la representación de terrenos, construcciones, manzanas, entre otros (Figura 11). En la actualización para identificar linderos que se encuentran desplazados respecto a la realidad física identificable mediante imágenes provenientes de sensores remotos (Figura 10) y para la incorporación de nuevas construcciones (Figura 9).

Figura 10

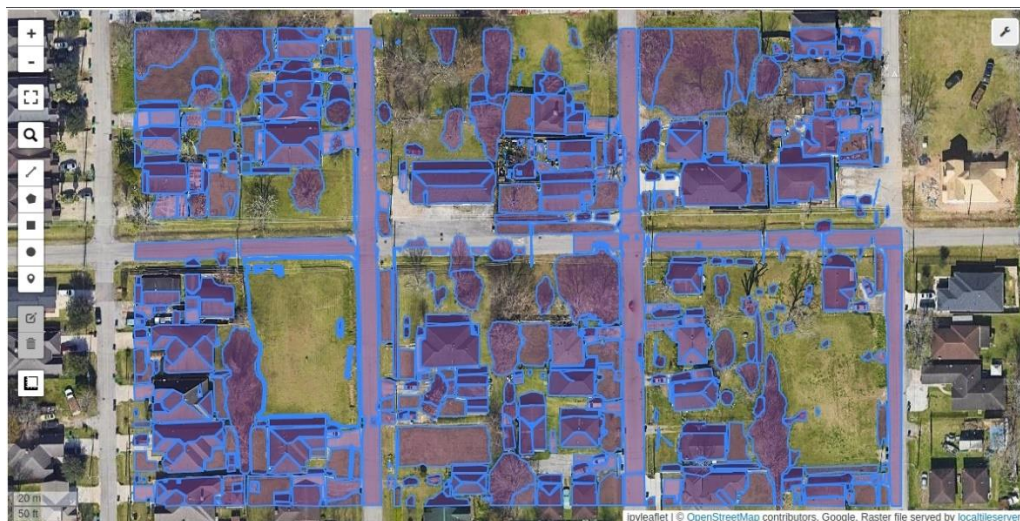
Comparación de Cartografía o Vectores Actuales y Redes Neuronales Convolucionales.



Nota. Actualización de linderos de parcelas haciendo uso de redes neuronales a partir de imágenes satelitales (Fetai et al., 2022)

Figura 11

Aplicación de Método SAM para Generación de Vectores.



Nota. Segmentación de imágenes satelitales haciendo uso de modelos de Machine Learning para procesamiento de imágenes de satélite (Q. Wu & Prado Osco, 2023)

Ya se mencionaron las posibilidades del ML en cuanto al componente cartográfico que es esencial para un buen reconocimiento predial. Ahora si nos centramos específicamente en la labor a realizar in situ por los operados podemos identificar que la calificación de construcciones, identificación de usos o actividades económicas desarrolladas en las mismas, edades de construcción y demás variables identificables de manera visual, pueden verse potenciadas significativamente por herramientas de ML enfocadas en el aprendizaje basado en imágenes.

Como se mencionó anteriormente el aprendizaje automático basado en imágenes es el campo macro que puede permitir extraer información catastral valiosa a un menor coste, con la fotografía de las construcciones es posible aplicar diversidad de técnicas que pueden ser refinadas para identificar características específicas, como daños o estados de conservación de

las construcciones, clasificaciones supervisadas basados en parámetros preestablecido para generar calificaciones de construcciones. Extracción de información como posibles usos o actividades económicas mediante aprendizaje profundo, (Jia & Li (2023) muestran una revisión de técnicas para identificar características de las construcciones, dentro de esta se encuentra el procesamiento de imágenes.

Uno de los puntos en los que el machine learning puede contribuir es en la geolocalización de direcciones que se encuentran dispuestas en documentos como escrituras, compraventas, licencias de construcción, certificados de tradición, recibos de servicios públicos y demás, pues técnica de ML como el XGB con ajuste de hiperparámetros ya han demostrado ser eficientes en materia de geocodificación de direcciones, incluso previendo errores humanos de digitación u ortográficos (K. Lee et al., 2020).

Finalmente la diversidad de productos generados por los procesos operativos del catastro requiere normativamente unas precisiones y criterios de calidad específicos, en cuanto al componente geográfico, estos deben estar soportados bajo un alto grado de exactitud posicional, en este sentido el ML también puede sugerir una alternativa, Siemuri et al. (2022) muestran las posibilidades del ML en las tecnologías GNSS que permiten el adecuado posicionamiento de elementos geográficos. Para este caso la precisión es crucial pues estos vectores son el insumo que permite una adecuada localización de predios de manera puntual, tanto en la instancia operativa como por ser insumo en instancias posteriores para diversas entidades.

Barrido Jurídico

En el componente jurídico de los procesos de formación catastral y regularización de tierras, es fundamental identificar a todas las personas naturales y jurídicas que mantienen relaciones con los predios, tanto de manera formal como informal. La identificación de titulares

formales se realiza mediante la consulta del certificado de tradición y otros documentos que evidencian la adquisición de derechos sobre el bien. En cambio, la identificación informal está relacionada con el reconocimiento in situ de las condiciones de posesión y ocupación de los predios. Esto implica un análisis exhaustivo de numerosos documentos legales y administrativos.

En este contexto, el machine learning (ML) puede ser un aliado poderoso, particularmente con el apoyo de técnicas de procesamiento de lenguaje natural (NLP) y reconocimiento de imágenes. Estas tecnologías permiten automatizar la extracción de información de documentos complejos, facilitando el trabajo de los analistas y reduciendo significativamente el tiempo necesario para procesar grandes volúmenes de datos.

El ML, apoyado en NLP, puede acelerar la revisión de documentos como certificados de tradición, contratos de compraventa y escrituras notariales. Estos modelos se entrenan para identificar patrones textuales y extraer información clave, como nombres de propietarios, fechas, referencias a otros documentos, y descripciones de propiedades. Esta capacidad de extracción automática mejora la eficiencia y disminuye la probabilidad de error humano, ya que los documentos suelen ser extensos y con vocabulario especializado.

El reconocimiento óptico de caracteres (OCR), potenciado por algoritmos de machine learning, es esencial cuando se trabaja con documentos antiguos o en mal estado. Mediante redes neuronales convolucionales, los sistemas OCR pueden mejorar la legibilidad de documentos escaneados y convertirlos en texto editable. Esta función es particularmente útil en el levantamiento de información histórica de predios, donde muchas veces los registros están deteriorados o son poco legibles debido al paso del tiempo.

Figura 12

Ejemplo Información Catastral Histórica.

REPUBLICA DE COLOMBIA
INSTITUTO GEOGRÁFICO
"AGUSTÍN GODAZZI"
SECCIONAL DE CALDAS
CATASTRO NACIONAL

Ficha predial
Municipio: GUINCHINA
Tipo: PREDIO RURAL
Vereda: BASSO CHUSCAL
Sección de Predio: 000-000

PROPIETARIOS (O POSEEDORES) SUCESIVOS

CLAVE DEL TÍTULO	NOMBRES (Personas naturales o jurídicas)	Matrícula de Clase No. Año	Estatus civil, Profesión, Actividad	Objeto o Tarjeta No.	Percepciones del terreno de propiedad	REPRESENTANTE LEGAL Nombre, calidad y domicilio
1	LUIS CALABREZ JOSUIS HINJA	FC- 01970	SAS-NIQUÉ-CHIMENNA	1273394	CC	6102
2	ASOCIATA BARRIO JOSE MARINO	001/ 011		1387046	CC	
3-4	Herencia de Simón María Jarchel	415 95	Res. 029/95		P1	

905-IV-C-4

JUSTIFICACIÓN DEL DERECHO DE PROPIEDAD (o de la posesión) DEL PREDIO

Medio de adquisición	Propietario anterior	CLAVE DEL TÍTULO	Base	Título	Fecha	Clase	Extensión	Valor	Fecha de inscripción
Compraventa	Valencia de Valencia Obesilla	2	NOT. GUINCHINA 412	25 ABRIL 1975	MATRICAL	19 28 451 709	7	mayo 1975	22/ 70
Compraventa	Donato	3	Padungala	01/ 1975		118 26 215	11	abril 1975	22/ 70
Compraventa	Bonnie-Jordano Bamardo	4	1ª Chichena	16 01 95		1408 7085	08/02	95	1

000-000

415 95

905-IV-C-4

08/02/95

Número predial tachado en inventario se rescata con estado de REGISTRO INACTIVO.

Número predial Vigente

Número de predio fiscal o catastro anterior

Número de matricula

Nota. Ejemplo del estado en el que se puede encontrar la información histórica para analizar dentro de los distintos procesos del catastro (IGAC, 2017)

Una vez digitalizados, los documentos pueden ser clasificados e indexados automáticamente según su tipo y contenido. Utilizando algoritmos de clasificación de texto, es posible categorizar documentos en función de criterios específicos, como los tipos de documentos ya mencionados propios de los barridos jurídicos catastrales. Esta organización permite una búsqueda más eficiente y una gestión documental más ordenada, lo cual es crucial en procesos catastrales donde el acceso eficiente a la información es fundamental.

Otra aplicación relevante es la integración de información documental con datos geoespaciales. Mediante técnicas de ML, es posible correlacionar información contenida en documentos con datos de ubicación obtenidos de sistemas de información geográfica (SIG). Esto

permite una visión integral del contexto del predio, facilitando el análisis de las condiciones de ocupación y el estado de formalización.

En resumen, el uso de ML y procesamiento de lenguaje natural en el ámbito jurídico para la gestión de predios representa una herramienta innovadora para optimizar el proceso de análisis documental, mejorar la precisión de los datos y reducir significativamente el esfuerzo manual. La combinación de estas tecnologías permite no solo acelerar los procesos, sino también garantizar una mayor transparencia y trazabilidad en la gestión de tierras.

Consolidación Económica

En cuanto al componente económico, el ML ha demostrado un gran potencial, especialmente en la clasificación y valoración masiva de inmuebles, obteniendo resultados prometedores en el ámbito nacional. Las técnicas de ML permiten la clasificación de zonas homogéneas físicas y económicas a partir de múltiples variables, así como la valoración de inmuebles utilizando información de ofertas, variables geográficas y socioeconómicas, lo cual resulta en estimaciones de valor más precisas y cercanas a la realidad inmobiliaria.

El ML se utiliza para la valoración masiva de predios mediante modelos de regresión y técnicas de aprendizaje supervisado, como Random Forests o Gradient Boosting. Estos modelos consideran una amplia variedad de variables, como características del inmueble, ubicación, datos socioeconómicos y comparables de mercado, para generar estimaciones de valor precisas y objetivas.

Otra actividad relevante es la extracción de información de páginas web que contienen ofertas de venta y renta de inmuebles. Utilizando técnicas de web scraping, junto con NLP, se puede recolectar y organizar información valiosa sobre precios, características de las propiedades

y ubicación. Esta información se integra en los modelos de valoración para mantener actualizados los datos de mercado.

El Machine Learning también se aplica para la generación de zonas homogéneas geoeconómicas, utilizando técnicas de clustering no supervisado, como K-means o DBSCAN. Estas técnicas permiten agrupar predios con características similares en términos de valor, ubicación y contexto socioeconómico, facilitando la planificación urbana y la gestión catastral.

Una vez realizada la valoración masiva, el ML puede ayudar a identificar valores atípicos que sugieran la necesidad de ajustes puntuales. Algoritmos de detección de anomalías, como Isolation Forest o Local Outlier Factor, se utilizan para detectar predios cuyo valor estimado difiere significativamente del valor esperado, permitiendo a los evaluadores realizar revisiones adicionales y asegurar la consistencia de la valoración.

De manera general es posible decir que el uso de Machine Learning y procesamiento de lenguaje natural en el componente económico de la gestión catastral representa una herramienta innovadora para optimizar la valoración de inmuebles. Estas tecnologías permiten mejorar la precisión de los datos, reducir el esfuerzo manual y garantizar una mayor transparencia y trazabilidad en la gestión de tierras y la valoración inmobiliaria.

Etapas Postoperativa

En la etapa post-operativa de un proyecto de formación o actualización catastral, el ML puede ser de gran ayuda para garantizar la consistencia y la calidad de los datos catastrales antes de cerrar el proyecto y dejar en vigencia las modificaciones del proceso. Durante esta etapa, es fundamental verificar que todas las mutaciones y trámites en las diferentes unidades de intervención hayan sido resueltos adecuadamente durante la etapa operativa, además de validar

los valores catastrales y los registros correspondientes. A continuación, se presentan algunas aplicaciones del ML en esta etapa.

Al tener consolidada la información física jurídica y económica, es posible utilizar técnicas de ML para verificar la consistencia de todas las variables de la base catastral generadas durante la etapa operativa. Haciendo uso de técnicas de detección de anomalías, como Isolation Forest, es posible identificar predios con valores atípicos o inconsistentes que requieran una revisión adicional. Esto permite garantizar que toda la información sea precisa y cumplan con la normatividad vigente.

De manera general al finalizar este tipo de procesos es notorio que el enfoque se encuentre en los avalúos, en este sentido los algoritmos de clasificación supervisada pueden facilitar la validación automática de los avalúos catastrales y las mutaciones registradas durante la etapa operativa. Los modelos pueden entrenarse con datos históricos para identificar inconsistencias o errores comunes en los registros, lo cual permite una revisión más rápida y precisa de estos, antes de la entrada en vigor fiscal de los mismos.

Retomando esta etapa de sensibilización de los valores obtenidos durante el proceso, es común la etapa de preliquidación, donde es posible emplear técnicas de ML para validar la información desde el componente físico y económico que en teoría son los que tienen un impacto directo en los valores catastrales finales usados para el cobro del impuesto predial unificado. Modelos de regresión y clustering pueden utilizarse para analizar y validar la coherencia de los valores catastrales generados, asegurando que los valores sean consistentes con las características físicas y socioeconómicas de los predios, conforme a la disponibilidad de estas según la definición de variables a levantar en el reconocimiento predial.

Posterior a todo el proceso de sensibilización, preliquidación y demás instancias que permitan subsanar de manera previa las posibles inconsistencias de la base catastral, se debe proceder con el cargue de información al SNC, tal como dicta la normatividad catastral vigente, este debe ser preciso y libre de inconsistencias. Tomando como referencia esta actividad crucial, podemos decir que el ML puede ayudar a identificar múltiples inconsistencias en esta instancia, tomando como referencia los validadores requeridos por el SNC y el modelo LADM vigente, haciendo uso de algoritmos de clustering y comparación de similitudes. Por ejemplo, estos modelos permiten detectar registros que puedan corresponder al mismo predio, evitando duplicidades y asegurando la integridad de la base de datos.

Finalmente a efectos de reportes adicionales a entidades del orden municipal, departamental y nacional, el ML puede facilitar la generación de productos estadísticos y reportes sobre el proceso de actualización catastral. Utilizando técnicas de análisis de datos, se pueden generar reportes automáticos sobre la distribución de avalúos, identificando patrones y tendencias en los valores catastrales. Estos reportes pueden ser de gran utilidad para la toma de decisiones y la publicación de resultados en los distintos canales oficiales y de orden común que permiten visibilizar el trabajo realizado y las ventajas en materia de ordenamiento del territorio que esto implica.

Finalmente una vez validados los datos catastrales, es importante socializar los resultados con todos los actores involucrados, desde autoridades locales hasta la comunidad en general. El ML puede facilitar la preparación de presentaciones y reportes personalizados para diferentes niveles de comunicación, utilizando técnicas de análisis de datos para resaltar la información relevante según el público objetivo. Además, herramientas de NLP pueden ser utilizadas para

generar resúmenes automáticos y adaptar el lenguaje técnico a un formato más comprensible para el público en general.

El Machine Learning puede ser una herramienta valiosa para mejorar la comunicación y socialización de los resultados, facilitando una segmentación adecuada del público objetivo. Utilizando técnicas de clustering y análisis de datos, es posible segmentar la comunidad según zonas específicas, características socioeconómicas y otros factores relevantes. Esto permite personalizar los mensajes y adecuar la información a cada segmento, logrando una comunicación más efectiva y una mayor aceptación de los resultados catastrales.

Además, el ML puede ayudar a identificar patrones de comportamiento y preferencias de los diferentes grupos de interés, lo cual permite adaptar la información y los canales de comunicación a cada grupo. Esto asegura una llegada más amplia y efectiva a la comunidad, mejorando la comprensión de los resultados del proceso de actualización catastral y garantizando que los diferentes actores involucrados reciban la información relevante de acuerdo con sus necesidades y nivel de conocimiento.

En conclusión, el uso de Machine Learning en la etapa post-operativa de un proyecto de actualización catastral permite garantizar la calidad y la consistencia de los datos catastrales, optimizar la validación de información y facilitar la generación de reportes y productos estadísticos. Estas aplicaciones contribuyen a una gestión catastral más eficiente y transparente, asegurando que los resultados sean precisos y estén alineados con la normatividad vigente.

Implementación del Machine Learning en la Gestión Catastral Multipropósito

Para la efectiva implementación del ML en todos los procesos inmersos de la gestión catastral multipropósito podemos considerar 6 fase que se acoplan a los diferentes modelos de ML y permiten implementarlos. 1. Evaluación de las necesidades, 2. Adquisición y preparación de datos, 3. Selección y entrenamiento de modelos de ML, 4. Validación e integración, 5. Capacitación del personal y 6. Evaluación de riesgos y desafíos éticos.

Evaluación de las Necesidades

El primer paso para implementar ML en la gestión catastral multipropósito es evaluar las necesidades específicas del proyecto. Esta fase implica identificar problemas actuales en el proceso de formación, actualización y conservación catastral, así como el potencial de aplicación del ML para abordar estos problemas. La evaluación se centra en determinar qué aspectos de los datos y de los procesos de gestión se beneficiarían de automatización, predicción y optimización mediante modelos de ML.

En esta fase, también se deben definir los objetivos de aplicación del ML, como mejorar la precisión en la identificación de propiedades, acelerar los procesos de formación y actualización catastral o generar estimaciones más precisas del valor de los predios. El conocimiento claro de las necesidades ayuda a definir los modelos y datos requeridos en las fases siguientes.

Adquisición y Preparación de Datos

La base del éxito en cualquier aplicación de ML es la calidad de los datos utilizados. La fase de adquisición y preparación de datos incluye la recopilación de información relevante, como datos geoespaciales, registros alfanuméricos, datos históricos de transacciones

inmobiliarias y datos socioeconómicos. Además, es crucial asegurar que los datos sean de alta calidad, consistentes y con la menor cantidad de errores posibles.

El proceso de preparación de datos involucra actividades como limpieza de datos, eliminación de valores atípicos, normalización y transformación de datos en un formato adecuado para el entrenamiento de modelos. En el contexto de la gestión catastral, es común tener que combinar información de diferentes fuentes, por lo que también es necesario asegurar la interoperabilidad de los datos y su correcta alineación en un modelo de datos estándar, tal como el modelo LADM.

Selección y Entrenamiento de Modelos de Machine Learning

En la fase de selección y entrenamiento de modelos, es importante identificar qué algoritmos de ML son más adecuados para resolver los problemas identificados en la fase de evaluación. Por ejemplo, los modelos de clasificación, como los árboles de decisión o las máquinas de soporte vectorial (SVM), pueden ser útiles para categorizar los predios según sus características constructivas, localización, características sociales y demás variables distintivas del territorio en términos del ordenamiento; mientras que los modelos de regresión pueden ayudar a predecir el valor de los inmuebles, calificaciones o categorías constructivas a partir de variables asociadas al componente físico.

El entrenamiento de los modelos requiere dividir los datos en conjuntos de entrenamiento y prueba, y utilizar métricas de evaluación para medir el rendimiento de los modelos seleccionados. Durante esta fase, también se pueden aplicar técnicas de ajuste de hiperparámetros para optimizar el desempeño del modelo y garantizar que sea adecuado para las condiciones particulares del contexto catastral, es fundamental direccionar de manera clara el propósito de la aplicación del modelo para obtener resultados acordes a las necesidades

territoriales, en este sentido dicho planteamiento se sugiere realizar de manera conjunta con expertos del contexto catastral y personal formado en la aplicación de modelos de ML.

Validación e Integración

Una vez entrenados los modelos, es fundamental realizar una validación rigurosa para asegurar que los resultados generados sean precisos y confiables. La validación de los modelos se puede realizar utilizando técnicas como validación cruzada, que permite evaluar el rendimiento del modelo utilizando diferentes subconjuntos de la totalidad de datos.

Adicionalmente se considera relevante una inspección muestral por parte de personal experto en materia catastral, asociar los resultados a las posibilidades ofrecidas por métodos tradicionales, en aras de garantizar la adecuada aplicación considerando los múltiples desafíos de la realidad territorial que tiene el catastro.

Luego de la validación, se debe proceder con la integración del modelo en los procesos actuales de gestión catastral. Esto implica la implementación del modelo en una plataforma operativa y su integración con las herramientas utilizadas por los gestores catastrales, asegurando que los resultados obtenidos sean útiles puedan ser interpretados fácilmente por los actores involucrados. En este proceso de integración es fundamental robustecer el componente de presentación del producto, tanto para las instancias que interpretaran la información producida, como para las instancias de producción y de adecuación de la misma, haciendo que todos los actores involucrados en la integración del ML en los procesos catastrales tengan las herramientas necesarias para su fácil adopción.

Capacitación del Personal

La capacitación del personal es un elemento clave para garantizar la adopción efectiva de las nuevas tecnologías basadas en ML. Es esencial que el personal técnico encargado de la

gestión catastral conozca los conceptos básicos del aprendizaje automático, así como el funcionamiento de los modelos implementados y cómo utilizar los resultados para mejorar los procesos de toma de decisiones.

La capacitación también debe incluir el uso de nuevas herramientas tecnológicas, como software de procesamiento de datos y herramientas de visualización, que permitan al personal analizar la información de manera eficiente. De igual manera, se deben realizar sesiones de sensibilización para que el personal entienda el impacto positivo de la implementación de ML en su trabajo diario.

Adicionalmente es importante considerar la etapa quizás mas influyente en la adopción de técnicas de aprendizaje automática, que es la producción de la información que van a alimentar a los diversos modelos de ML a implementar, esto implica que se debe garantizar una adecuada captura y adecuación de la información recopilada en campo y oficina para los diferentes componentes, deben ser acorde a las normativas técnicas catastrales vigentes y preservar la estructura semántica, que actualmente esta regida por los modelos de aplicación LADM.

En este sentido surge la posibilidad de no solo considerar el ML como una herramienta de procesamiento y automatización de procesos, sino también como un instrumento que contribuya a la adecuada capacitación del personal, como elementos que actualmente ya se encuentran implementados en otros sectores, el uso de interfaces tipo chat impulsado por IA, generación de contenido conciso para consulta efectiva y demás herramientas del ML que permita al personal consultar todos los requerimientos necesarios para garantizar la calidad de la información. También generando menor dedicación de personal a estas labores, facilitando su enfoque y experiencia en otras actividades también cruciales del proceso.

Evaluación de Riesgos y Desafíos Éticos

La implementación de modelos de ML conlleva una serie de riesgos y desafíos éticos que deben ser considerados cuidadosamente. Uno de los principales riesgos está relacionado con la privacidad de los datos, ya que la gestión catastral maneja información sensible sobre propiedades y propietarios, y dentro del contexto colombiano el acceso a tierras siempre ha sido un elemento de conflicto, tal como se enuncia en los acuerdos de paz de la Habana. En este sentido es fundamental asegurar la protección de los datos mediante políticas estrictas de seguridad y privacidad, las cuales garanticen que la implementación del ML en materia catastral no sea un instrumento para la adición de conflictividades al territorio.

Otro elemento que se debe considerar son los posibles sesgos en los datos y modelos utilizados. Si los datos contienen prejuicios inherentes, los modelos podrían reproducir y maximizar dichos sesgos, lo cual resultaría en categorizaciones quizás discriminatorias o generadoras de conflictividad. Por lo tanto, es esencial implementar métricas de equidad y garantizar la transparencia en el desarrollo e implementación de los modelos, en este particular es esencial la inclusión de actores multisectoriales y culturales para garantizar el sesgo en la menor medida posible.

Finalmente el elemento quizás con mayor relevancia actual es la implicación de la automatización de procesos en la disminución de empleos en diferentes campos, sobre este particular se considera que en materia catastral la adopción del ML no implica una eliminación de empleos sino más bien una transformación, claramente hay empleos que se enfocan en tareas rutinarias las cuales están más propensas a automatización, pero en vista de las implicaciones que tiene continuar con un catastro desactualizado y la eficiencia que debe mantener el sector público para la adecuada gestión territorial, es necesario considerar la transformación del sector,

buscando de manera conjunta la transformación de los empleos con habilidades más tecnológicas que a su vez garanticen su competitividad.

Conclusiones

La implementación de técnicas de Machine Learning (ML) en la gestión catastral multipropósito representa una oportunidad significativa para acelerar los procesos de actualización catastral en Colombia, especialmente considerando el bajo porcentaje actual de actualización (< 13%) y la meta del 70% establecida en el Plan Nacional de Desarrollo 2022-2026.

Las técnicas de ML pueden contribuir a optimizar los siguientes aspectos clave, como automatización en la identificación de cambios físicos mediante el procesamiento de imágenes fotográficas, satelitales y aerofotografías, clasificación y extracción automática de información de documentos jurídicos, valoración masiva de predios mediante modelos predictivos y detección de inconsistencias en las bases de datos catastrales.

La aplicación de ML en los diferentes componentes del catastro (físico, jurídico y económico) debe realizarse de manera diferenciada, considerando las particularidades de cada uno. En el componente físico, las técnicas de visión por computador y redes neuronales convolucionales han demostrado ser efectivas para la identificación automática de construcciones y cambios en el territorio, para el componente jurídico, el procesamiento de lenguaje natural (NLP) y reconocimiento óptico de caracteres (OCR) facilitan el análisis masivo de documentos legales, en el componente económico, los modelos de regresión y técnicas de aprendizaje supervisado permiten realizar valoraciones masivas que sugieren ser más precisas y objetivas.

La implementación efectiva de ML en el catastro multipropósito requiere una estrategia integral que contemple, una evaluación detallada de necesidades específicas por componente y sus subprocesos, adquisición y preparación rigurosa de los datos a usar en el entrenamiento,

selección y validación minuciosa de modelos apropiados para cada proceso o actividad, capacitación continua del personal en nuevas tecnologías y consideración de aspectos éticos y de privacidad en el manejo de datos y los productos de implementación del ML.

El impacto laboral de la implementación de ML en el sector catastral debe entenderse como una transformación más que como una eliminación de empleos, se requiere desarrollar nuevas competencias en el personal actual, la posibilidad de nuevos roles relacionados con la supervisión y validación de resultados del ML, adicionalmente la complejidad del territorio colombiano seguirá requiriendo criterio humano experto, finalmente esta implementación permitirá redirigir recursos humanos hacia tareas más estratégicas.

Los desafíos principales para la implementación exitosa de ML en el catastro pueden ser amplios, algunos de los más cruciales pueden ser, garantizar la calidad y consistencia de los datos de entrenamiento, asegurar la interoperabilidad con el modelo LADM colombiano, mantener la privacidad y seguridad de la información catastral, evitar sesgos en los modelos que puedan afectar la equidad en las diferentes predicciones o clasificaciones y también gestionar el cambio cultural y organizacional necesario para su oportuna adopción.

Entendiendo que la implementación es una desafío mayor se sugieren una recomendaciones para cumplir dicho fin, como el desarrollo de proyectos piloto en municipios con diferentes características para validar la efectividad de los modelos, establecer métricas claras de evaluación de resultados, crear equipos multidisciplinarios que combinen experiencia catastral y en ML, implementar un sistema de mejora continua basado en retroalimentación multisectorial, desarrollar guías metodológicas para la implementación estandarizada en lo basto del territorio nacional, con las particularidades a que diera lugar.

La implementación de ML en el catastro multipropósito debe verse como una herramienta complementaria que potencia las capacidades humanas, no como un reemplazo completo o tajante de los procesos actuales. Su adopción gradual, planificada, basada en el principio de progresividad puede contribuir significativamente a alcanzar las metas de actualización catastral del país, mejorando la eficiencia y precisión de los procesos mientras se mantiene la calidad y confiabilidad de la información catastral.

Es fundamental mantener un enfoque equilibrado que combine la innovación tecnológica con el conocimiento experto del territorio y sus particularidades sociales y culturales, especialmente considerando la diversidad y complejidad del contexto colombiano, que a su vez sugiere ser un instrumento sustancial para el establecimiento del desarrollo de las comunidades en función de los recursos y la eliminación de las conflictividades históricas fundamentalmente en el ámbito rural.

Referencias

- Bastidas Méndez, K. T., & Bonilla Bautista, A. V. (2018). *Uso de inteligencia artificial aplicando árboles de decisión ejecutados en r para la determinación de zonas homogéneas físicas*. Universidad Distrital Francisco Jose de Caldas.
- bin Abdullah, M. R., & Iqbal, K. (2022). A Review of Intelligent Document Processing Applications Across Diverse Industries. *Journal of Artificial Intelligence and Machine Learning in Management*, 6(2), 29-42.
- Caltech. (2024, mayo 14). *What are the main algorithms used in ML?* California Institute of Technology.
- Carranza, J. P., Piumetto, M. A., Salomón, M. J., Monzani, F., Montenegro, M. G., & Córdoba, M. A. (2019). Valuación Masiva de la Tierra Urbana Mediante Inteligencia Artificial. *Revista Vivienda y Ciudad*, 6, 90-112.
- Catastro Bogota. (2020). *Catastro Multipropósito*.
- Decreto 148 de 2020 (2020).
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. <http://arxiv.org/abs/1810.04805>
- DNP. (2020). Actualización catastral con enfoque multipropósito. En *Proyectos Tipo - Actualización catastral con enfoque multipropósito* .
- DNP. (2023, septiembre 10). *Con Inteligencia Artificial, el DNP apoya el proceso de actualización del Catastro Multipropósito*. xxxxx.
- Edwards, A. S., Kaplan, B., & Jie, T. (2021). *A primer on machine learning. Transplantation*. 105(4), 699-703.

- Fetai, B., Grigillo, D., & Lisec, A. (2022). Revising Cadastral Data on Land Boundaries Using Deep Learning in Image-Based Mapping. *ISPRS International Journal of Geo-Information*, *11*(5). <https://doi.org/10.3390/ijgi11050298>
- García-Retuerta, D., Bondía, R. A., Prieto-Tejedor, J., & Corchado, J. M. (2018). Inteligencia artificial para la asignación automática de categorías constructivas. *CT catastro*, *94*, 111-121.
- Gobierno Nacional, & FARC-EP. (2016). *Acuerdo Final Gobierno de Colombia y FARC-EP para la terminación del conflicto y la construcción de una paz estable y duradera*.
- Guan, J., Shi, D., Zurada, J. M., & Levitan, A. S. (2014). Analyzing Massive Data Sets: An Adaptive Fuzzy Neural Approach for Prediction, with a Real Estate Illustration. *Journal of Organizational Computing and Electronic Commerce*, *24*(1), 94-112.
<https://doi.org/10.1080/10919392.2014.866505>
- Gustavo Marulanda. (2023, mayo 2). *La meta es llegar al 70 % de actualización catastral en este cuatrienio*. <https://www.larepublica.co/economia/la-actualizacion-del-catastro-va-en-9-4-no-en-40-con-10-7-millones-de-hectareas-3469843>
- Hajiheidari, A., Delavar, M. R., & Rajabifard, A. (2024). Smart Urban Cadastral Map Enrichment—A Machine Learning Method. *ISPRS International Journal of Geo-Information*, *13*(3). <https://doi.org/10.3390/ijgi13030080>
- Hester, T., Vecerik, M., Pietquin, O., Lanctot, M., Schaul, T., Piot, B., & Gruslys. (2018). Deep q-learning from demonstrations. *Proceedings of the AAAI conference on artificial intelligence*.

- ICDE. (2023). Metodología Para La Gestión Del Modelo Núcleo Y Modelos Extendidos LADM (Land Administration Domain Model) En Colombia. En *Modelo de la Administración del Territorio - Land Administration Domain Mode*.
- ICDE. (2024). ICDE. Nuevo repositorio de modelo núcleo y modelos extendidos LADM_COL.
- IGAC. (2017). Organización Documental De Expedientes De Ficha Predial. En *Manual De Procedimiento Organización Documental De Expedientes De Ficha Predial*.
- IGAC, & SNR. (2024). *Resolución Conjunta SNR 05346 IGAC 662*.
- Ignjatić, J., Nikolić, B., Rikalović, A., & Čulibrk, D. (2018). Deep learning for historical cadastral maps digitization: Overview, challenges and potential. *Computer Science Research Notes*, 2803(May), 42-47. <https://doi.org/10.24132/CSRN.2018.2803.6>
- Jang, B., Kim, M., Harerimana, G., & Kim, J. W. (2019). Q-learning algorithms: A comprehensive classification and applications. *IEEE access*, 7, 133653-133667.
- Jia, J., & Li, Y. (2023). Deep learning for structural health monitoring: Data, algorithms, applications, challenges, and trends. *Sensors*, 23(21), 8824.
- Jorge Gonzalez. (2023, septiembre 6). *Es el catastro multipropósito la verdadera revolución para la reforma rural integral*.
- Kara, A., van Oosterom, P., Çağdaş, V., Işıkdag, Ü., & Lemmen, C. (2020). 3 Dimensional data research for property valuation in the context of the LADM Valuation Information Model. *Land Use Policy*, 98. <https://doi.org/10.1016/j.landusepol.2019.104179>
- Kozan, M. (2021). *Supervised and Unsupervised Learning (an Intuitive Approach)*.
- Lee, C., & de Vries, W. T. (2020). Bridging the semantic gap between land tenure and EO data: Conceptual and methodological underpinnings for a geospatially informed analysis. *Remote Sensing*, 12(2). <https://doi.org/10.3390/rs12020255>

- Lee, K., Claridades, A. R. C., & Lee, J. (2020). Improving a street-based geocoding algorithm using machine learning techniques. *Applied Sciences*, *10*(16), 5628.
- Ley 223 de 1995 (1995).
- Mango, J., Wang, M., Mu, S., Zhang, D., Ngondo, J., Valerian-Peter, R., Claramunt, C., & Li, X. (2023). Transform paper-based cadastral data into digital systems using GIS and end-to-end deep learning techniques. *International Journal of Geographical Information Science*, *37*(5), 1099-1127. <https://doi.org/10.1080/13658816.2023.2178002>
- Microsoft. (2024). Microsoft Global Building Footprints. En *Microsoft Global Building Footprints GitHub*.
- MIT. (2020). *Introduction to machine learning*. MIT - Massachusetts Institute of Technology.
- Mo, M., & Yomralioglu T. (2022). Mass Valuation of Real Estate Using GIS-based Nominal Valuation and Machine Learning Methods. *28th YSAual Conference of the European Real Estate Society (ERES)*, 22-25.
- Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python: a guide for data scientists*. O'Reilly Media, Inc.
- Oliveira, S. A., Kaplan, F., & Di Lenardo, I. (2017). Machine Vision algorithms on cadaster plans. *Premiere Annual Conference of the International Alliance of Digital Humanities Organizations*, 1-6. <https://github.com/dhlab-epfl/cadasters>.
- Oliveira, S. A., Lenardo, I. Di, Tourenc, B., & Kaplan, F. (2019). A deep learning approach to Cadastral Computing. *Digital Humanities Conference*, 1-7.
- Omurca, S. İ., Ekinici, E., Sevim, S., Edinç, E. B., Eken, S., & Sayar, A. (2023). A document image classification system fusing deep and machine learning models. *Applied Intelligence*, *53*(12), 15295-15310. <https://doi.org/10.1007/s10489-022-04306-5>

- Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3), 128-138.
- Park, J. S., & Park, J. H. (2020). Enhanced machine learning algorithms: deep learning, reinforcement learning, and q-learning. *Journal of Information Processing Systems*, 16(5), 1001-1007.
- Park, Y., & Guldmann, J. M. (2019). Creating 3D city models with building footprints and LIDAR point cloud classification: A machine learning approach. *Computers, Environment and Urban Systems*, 75, 76-89. <https://doi.org/10.1016/j.compenvurbsys.2019.01.004>
- Plan Nacional de Desarrollo 2018-2020 - Pacto por Colombia Pacto por la Equidad, Pub. L. No. Ley 1955 (2018).
- Plan nacional de desarrollo 2022-2026 - Colombia potencia mundial de la vida, Pub. L. No. Ley 2294 (2023).
- Potsiou, C., Doulamis, N., Bakalos, N., Gkeli, M., & Ioannidis, C. (2020). Indoor Localization for 3D Mobile Cadastral Mapping Using Machine Learning Techniques. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 6(4/W1), 159-166. <https://doi.org/10.5194/isprs-annals-VI-4-W1-2020-159-2020>
- Resolución 1040 del 2023, Pub. L. No. 1040 (2023).
- Siemuri, A., Selvan, K., Kuusniemi, H., Valisuo, P., & Elmusrati, M. S. (2022). A systematic review of machine learning techniques for GNSS use cases. *IEEE Transactions on Aerospace and Electronic Systems*, 58(6), 5043-5077.

- Sisman, S., & Aydinoglu, A. C. (2022a). A modelling approach with geographically weighted regression methods for determining geographic variation and influencing factors in housing price: A case in Istanbul. *Land Use Policy*, 119.
<https://doi.org/10.1016/j.landusepol.2022.106183>
- Sisman, S., & Aydinoglu, A. C. (2022b). Improving performance of mass real estate valuation through application of the dataset optimization and Spatially Constrained Multivariate Clustering Analysis. *Land Use Policy*, 119.
<https://doi.org/10.1016/j.landusepol.2022.106167>
- Sutton, R. S., & Barto, A. G. (2015). *Reinforcement Learning: An Introduction* (1.^a ed., Vol. 1).
- Wu, C., Li, X., Guo, Y., Wang, J., Ren, Z., Wang, M., & Yang, Z. (2022). Natural language processing for smart construction: Current status and future directions. *Automation in Construction*, 134.
- Wu, Q., & Prado Osco, L. (2023). samgeo: A Python package for segmenting geospatial data with the Segment Anything Model (SAM). *Journal of Open Source Software*.
- Yu, W., Lu, N., Qi, X., Gong, P., & Xiao, R. (2020). PICK: Processing Key Information Extraction from Documents using Improved Graph Learning-Convolutional Networks. *International Conference on Pattern Recognition (ICPR)*, 4363-4370. <http://arxiv.org/abs/2004.07464>
- Yu, Y., Lu, J., Shen, D., & Chen, B. (2021). Research on real estate pricing methods based on data mining and machine learning. *Neural Computing and Applications*, 33(9), 3925-3937.
<https://doi.org/10.1007/s00521-020-05469-3>
- Yuan, X., Shi, J., & Gu, L. (2021). A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Systems with Applications*, 169.
<https://doi.org/10.1016/j.eswa.2020.114417>

Zhao, X., Niu, E., Wu, Z., & Wang, X. (2019). *CUTIE: Learning to Understand Documents with Convolutional Universal Text Information Extractor*. <http://arxiv.org/abs/1903.12363>