

DISEÑO DE UN MODELO QUE INCORPORE DESAGREGADAMENTE CONDICIONES SOCIOECONÓMICAS ESPECÍFICAS PARA LA VALORACIÓN PREDIAL EN LA CIUDAD DE BARRANQUILLA

Nombres y apellidos

Jair José Barreto Leguía

Código estudiantil: 2022120243680

Trabajo de Investigación presentado como requisito para optar el título de:

Magister en Ingeniería de Sistemas y Computación

Tutor(es):

Diana Judith Heredia vizcaíno

Juan Manuel Rúa Ascar

RESUMEN

La influencia significativa del sector inmobiliario en la economía de los países y la importancia de la gestión catastral destaca la relevancia de la exactitud en la valoración predial; circunscrito en ese acotamiento se presenta esta investigación de enfoque mixto con tendencia cualitativa para el diseño de una metodología que incorpora de manera más detallada condiciones socioeconómicas específicas para la valoración predial en Barranquilla. Éste enriquece y complementa los métodos existentes para obtener avalúos congruentes con la realidad de su entorno socioeconómico.

El proyecto fija en calidad de derrotero la revisión de antecedentes bibliográficos, la caracterización de condiciones socioeconómicas y la estructuración de un modelo de análisis que relacione lo anterior con la valuación de los predios como instrumento promisorio que apunte criterios a los tomadores de decisión en el ámbito financiero general y en la administración pública. Entre los aspectos destacados se encuentra el método de revisión documental basado en las

especificaciones PRISMA, con búsqueda en recursos académicos como Scopus, Scielo y otras base de datos en el orden de postgrados; también las estrategias de acopio de data que se concentraron transversalmente en los procesos de carácter tecnológicas de la Gerencia de Gestión catastral de la ciudad de Barranquilla y el enriquecimiento de estos con información estructurada de oficinas de titularidad externa, asegurando la legalidad y protección de datos; así se da paso a la operacionalización de las variables en las que cobra preponderancia el avalúo predial como variable dependiente, las condiciones socioeconómicas que permean la favorabilidad de los predios como variables independientes y la transformación de estas ideas abstractas a términos concretos, observables y medibles con la definición de las dimensiones e indicadores.

Una vez identificadas las CSE para la exploración de datos, se entabla la "Depuración de datos", eliminando y corrigiendo registros incompletos e inconsistentes, y estandarizando, discretizando y normalizando los datos. En lo referente al componente geográfico, se utiliza *buffering* e intersecados para la identificación de predios afectados por las CSE. Para automatizar tareas se emplea procesos ETL y ELT con Pentaho y Oracle Express.

La antesala descrita conduce a la estructuración del modelo, que imperativamente requiere actividades de preprocesamiento, como también, análisis exploratorio, destacándose la estandarización LADMCOL de direcciones con una efectividad del 92% en la muestra utilizada, asimismo un proceso de extracción de datos mediante técnicas de raspado web, seguido del geoprocésamiento de las variables independientes y otras actividades amanuenses que finalmente aterrizan estos flujos heterogéneos de información en una consolidación de datos provistos de caracterización y validación de resultados mediante métodos no supervisados de inteligencia artificial como la descomposición de componentes principales y la clusterización con Python. Esto muestra información relevante con relación a algunos periodos de avalúos, entre estos, la correlación con variables como "AREA_TOTAL" y "TOTAL_UNIDADES", como también, se destaca el hecho que, las CSE analizadas no muestran correlación evidente. Asimismo, se revela la presencia de outliers y la evaluación de *clustering*, abordado desde el conocimiento del contexto y la situación problémica de la investigación.

Este enfoque integral argumenta la hipótesis inicial sobre la suficiencia en el nivel de desagregación de las variables físicas y geoeconómicas de los modelos de valoración predial en la ciudad como situación problémica de la presente investigación, denota además el grado de incidencia de las condiciones en los avalúos cuando presentan cercanía a los predios, en lo que se prevé como resultado esperado específicamente que la presencia de condiciones socioeconómicas adversas desvaloriza los predios y viceversa. También genera conocimiento adicional y enriquece las metodologías existentes para una valoración predial más precisa y ajustada a la realidad socioeconómica de Barranquilla

Palabras clave: : predial, avalúo, Python, Clusterización, Geoprocésamiento, raspado, PLSQL, LADMCOL, inmobiliario, Prisma, Scopus, Scielo

ABSTRACT

The significant influence of the real estate sector on the economy of countries and the importance of cadastral management highlight the relevance of accuracy in property valuation. Within this argument, this mixed-method research with a qualitative tendency aims to design a methodology that incorporates specific socioeconomic conditions in greater detail for property valuation in Barranquilla. This approach enriches and complements existing methods to obtain valuations consistent with the socioeconomic reality of the area.

The project has as a guideline the review of bibliographic references, the characterization of socioeconomic conditions, and the structuring of an analysis model that relates these factors to property valuation. This model is envisioned as a promising tool to support decision-makers in the financial sector and public administration. Among the highlights is the document review method based on PRISMA specifications, utilizing academic resources such as Scopus, Scielo, and other postgraduate databases. Additionally, data collection strategies focused on technological processes from the Gerencia de gestión catastral de Barranquilla office and the improvement of this data with structured information from external offices, ensuring data legality and protection. This process leads to the operationalization of variables, with property appraisal as the dependent variable, socioeconomic conditions influencing property favorability as independent variables, and the transformation of these abstract ideas into concrete, observable, and measurable terms through the definition of dimensions and indicators.

Once the socioeconomic conditions (CSE) for data exploration are identified, data cleansing is carried out, eliminating and correcting incomplete and inconsistent records, and standardizing, discretizing, and normalizing the data. For the geographic component, buffering and intersecting techniques are used to identify properties affected by the CSE. ETL and ELT processes with Pentaho and Oracle Express are employed to automate tasks.

This context leads to the model's structuring, which imperatively requires preprocessing activities, as well as exploratory analysis. This includes the LADMCO standardization of addresses, achieving a 92% effectiveness in the sample used, a data extraction process using web scraping techniques, followed by geoprocessing of the independent variables, and other manual activities. These activities culminate in consolidating the heterogeneous information flows, characterized and validated through unsupervised artificial intelligence methods such as principal component analysis and clustering with Python. This reveals relevant information related to valuation periods, including correlations with variables like "TOTAL_AREA_PROPERTY" and "TOTAL_UNITS_PROPERTY," as well as the notable absence of evident correlation with the analyzed CSE. Additionally, the presence of outliers and clustering evaluation is highlighted, approached from the research context and problematic situation.

This comprehensive approach supports the initial hypothesis regarding the adequacy of the disaggregation level of physical and geoeconomic variables in property valuation models in the city, reflecting the research's problematic situation. It also denotes the degree of impact of conditions on valuations when in proximity to

properties, anticipating the expected result that adverse socioeconomic conditions depreciate properties and otherwise. Furthermore, it generates additional knowledge and enriches existing methodologies for a more precise and socioeconomically adjusted property valuation in Barranquilla.

Key Words: predial, Valuation, Python, Clustering, Georeferencing, Scraping, Plsql, Ladmcol

REFERENCIAS BIBLIOGRÁFICAS

1. Cano, J, El desafío de los algoritmos: riesgos y amenazas invisibles en el contexto digital. 2021.
<https://www.linkedin.com/pulse/el-desaf%C3%ADo-de-los-algoritmos-riesgos-y-amenazas-en-cano-ph-d-cfe/?originalSubdomain=es>
2. Cifuentes Medina, J., & Torres Ortiz, J. Reflexiones en y para la enseñanza de la historia de la ética. 2019
<https://doi.org/10.15332/s1794-3841.2019.0031.07>
3. Departamento Administrativo de la Presidencia de la República. Marco ético para la inteligencia artificial en Colombia. 2021
<https://minciencias.gov.co/sites/default/files/marco-etico-ia-colombia-2021.pdf>
4. Hao, K. MIT technology review. 2022.
<https://www.technologyreview.es/s/14144/la-inteligencia-artificial-esta-creando-un-nuevo-orden-mundial-colonial>
5. Kant, I.). fundamentos de la metafísica de las costumbres. El cid editor. 2003.
https://www.cervantesvirtual.com/obra-visor/fundamentacion-de-la-metafisica-de-las-costumbres--0/html/dcb0941a-2dc6-11e2-b417-000475f5bda5_3.html.
6. MATILLION. User Guide To Machine Learning and data transformation in the cloud. 2019.
<https://www.matillion.com/learn/blog/data-transformation>.
7. REAL ACADÉMIA ESPAÑOLA. 23.^a ed., [versión 23.6 en línea]. Diccionario de la lengua Española. <<https://dle.rae.es>> [15/02/2023].
<https://dle.rae.es>
8. R. O. Badiang, B. D. Gerardo and R. P. Medina, Relocating Local Outliers Produced by K-means and K-medoids Using Local Outlier Rectifier V.2.0, 2019 International Conference on Advanced Computer Science and information Systems (ICACSIS), Bali, Indonesia, 2019, pp. 89-94.
<https://doi.org/10.1109/ICACSIS47736.2019.8979741>
9. Alcaldía Distrital de Barranquilla. Historia de nuestra creación.2023.
<https://www.barranquilla.gov.co/hacienda/catastro/historia-de-nuestra-creacion>.
10. Amazon. ¿Qué es extracción, transformación y carga (ETL)? 2023.
<https://aws.amazon.com/es/what-is/etl/>
11. Avendaño Arias, J. A. Información catastral: dimensiones contemporáneas para la gestión de las ciudades en el siglo XXI. 2022.

- http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S2462-91032021000300115&lng=en&tlng=es
12. Banco de la república de Colombia. (s.f.). Producto interno bruto. 2023
<https://www.banrep.gov.co/es/glosario/producto-interno-bruto-pib>
 13. Bermúdez Ayala, M., Ortiz, J. C., & Fuentes López, H. J. Deficit habitacional en los municipios del litoral Pacífico. 2022.
<https://doi.org/10.15446/bitacora.v32n3.98342>
 14. Bosch Rué, A. Casas Roma, & J. & Lozano Bagén. Deep learning: principios y fundamentos. UOC. 2019.
<https://elibro.net/es/ereader/unisimon/126167?page=19>
 15. Cambridge. (1 de mayo de 2024). cambridge Dictionary.
<https://dictionary.cambridge.org/es/diccionario/ingles/outlier>
 16. Campos Medina, L. A., & Sandoval Pavez, J. L. Resistencia a la basura y dinámicas de territorialización a través del uso de la escritura expuesta. 2020.
<http://dx.doi.org/10.22320/07183607.2020.23.42.04>
 17. Carranza, J. P., & Lucca, C. M. Mass appraisal as affordable public policy: Open data and machine learning for mapping urban land values. Córdoba, Argentina. 2022.
<https://doi.org/10.1016/j.landusepol.2022.106211>
 18. Catastro Bogotá. Unidad Administrativa Especial de Catastro Bogotá
 19. Codazzi-IGAC, I. G. Experience Arcgis. 2022
<https://beta.igac.gov.co/el-igac/areas-estrategicas/direccion-de-investigacion-prospectiva/observatorio-inmobiliario-catastral>
 20. Colado Velázquez, L. A., Roca Cladera, J., & Humarán Nahed, I. Inequidad en el impuesto inmobiliario: análisis multicriterio (CRITIC) del valor catastral de vivienda en Mazatlán. 2021.
<https://doi.org/10.22136/est20201593>
 21. Colón Llamas, L. C., & Picón Rodríguez, A. Digitalizar, georreferenciar, visualizar y narrar la ciudad. Usos del suelo y orden urbano: Bogotá, 1894. 2022.
<https://doi.org/10.12804/revistas.urosario.edu.co/territorios/a.11785>
 22. Congreso de la República de Colombia. LEY 44, Por la cual se dictan normas sobre catastro e impuestos sobre la propiedad raíz, se dictan otras disposiciones de carácter tributario, y se conceden unas facultades extraordinarias. 1990.
<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=28>
 23. Departamento Administrativo de estadísticas (DANE), D. a. Boletín Técnico - Indicadores económicos alrededor de la construcción (IEAC). 2022.
<https://www.dane.gov.co/index.php/estadisticas-por-tema/construccion/indicadores-economicos-alrededor-de-la-construccion/historicos-indicadores-economicos-alrededor-de-construccion>
 24. Datos Abiertos. ACCIDENTALIDAD EN BARRANQUILLA. 2023.
<https://www.datos.gov.co/Transporte/Accidentalidad-en-Barranquilla/yb9r-2dsi>
 25. Departamento administrativo de la Función pública. MIPG-Modulo introductorio. 2021.
<https://www1.funcionpublica.gov.co/web/eva/curso-mipg>
 26. Departamento administrativo de la función pública. (s.f.). Departamento administrativo de la función pública. (s.f.). 2023.

- <https://www.catastromultiproposito.gov.co/entes-territoriales/Paginas/habilitacion-de-gestores-catastrales.aspx>
27. Departamento nacional de planeación. (s.f.). Catastro multiproposito.2023
<https://www.catastromultiproposito.gov.co/entes-territoriales/Paginas/habilitacion-de-gestores-catastrales.aspx>
28. Departamento Administrativo Nacional de Estadística. Metodología Índice de Valoración Predial. 2009.
<https://www.dane.gov.co/index.php/estadisticas-por-tema/precios-y-costos/indice-de-valoracion-predial-ivp>
29. Esri. Arcgis Resource, Introducción a ArcGIS. 2023.
<https://resources.arcgis.com/es/help/getting-started/articles/026n00000014000000.htm>
30. Esri - Environmental Systems Research Institute. ArcMap-Estandarizar direcciones.2023
<https://doc.arcgis.com/es/pretrained-models/latest/text/introduction-to-address-standardization.htm>
31. Esri - Environmental Systems Research Institute Buffer arcgis pro. 2023.
<https://pro.arcgis.com/es/pro-app/latest/tool-reference/analysis/buffer.htm>
32. Esri. Arcgis Resource, Introducción a ArcGIS. Cómo funciona Intersecar ArcGIS Pro 3.2. 2023
<https://pro.arcgis.com/es/pro-app/latest/tool-reference/analysis/how-intersect-analysis-works.htm>
33. Esri. Arcgis Resource, Introducción a ArcGIS. Zona de influencia desktop arcgis. 2023
<https://desktop.arcgis.com/es/arcmap/latest/tools/coverage-toolbox/buffer.htm>
34. George M. L., B. D. Lean six sigma en la era de la inteligencia artificial. 2019
<https://elibro.net/es/ereader/unisimon/189588?prev=fs>
35. Hair, J. F. FUNDAMENTOS DE ANALITICA DE MARKETING. McGraw-Hill Interamericana. 2022
<https://ebooks7-24.com/?il=20750>
36. Heraldo, E. Parques, un motor de transformación urbana y social en Barranquilla. El Heraldo. 5 de febrero de 2023.
<https://www.elheraldo.co/barranquilla/parques-en-barranquilla-las-claves-detras-del-premio-como-el-proyecto-mas-innovador-del>
37. HITACHI. Productos. 14 de noviembre de 2023.
https://help.hitachivantara.com/Documentation/Pentaho/Data_Integration_and_Analytics/9.5/Products
38. IBM. ¿Qué es el aprendizaje no supervisado? 2024.
<https://www.ibm.com/es-es/topics/unsupervised-learning>
39. IGAC-Instituto Geográfico Agustín Codazzi. Resolución 070, artículo 52. 2011.
<https://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=41638>
40. IGAC-Instituto Geográfico Agustín Codazzi. Resolución 070 de 2011, artículo 51. 2011.
<https://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=41638>
41. IGAC-Instituto Geográfico Agustín Codazzi. ¿Qué hacemos? 2022.
<http://www.igac.gov.co/es/que-hacemos>

42. IGAC-Instituto Geográfico Agustín Codazzi. LADM_COL. 2022
<https://igac.gov.co/es/catastro-multiproposito/ladm-col>
43. IGAC-Instituto Geográfico Agustín Codazzi. IGAC 80 años. 2022.
<https://igac.gov.co/es/contenido/igac-80-anos>
44. Congreso de la República de Colombia. LEY 44 DE 1990- Por la cual se dictan normas sobre catastro e impuestos sobre la propiedad raíz, se dictan otras disposiciones de carácter tributario, y se conceden unas facultades extraordinarias. 18 de diciembre de 1990.
<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=283>
45. Lin, T.-Y., Hung-Lin Lin, & Chih-Wei Hou. Research on the production of 3D image cadastral map. 2018
<https://doi.org/10.1109/ICASI.2018.8394582>
46. Lisseth Jami Aymacaña, E. Catalina Vallejo-Coral, Luis Godoy, Francis David Vásquez Aza, & José Luis Condo León. Methodology of association of cadastral and electrical information using GIS and SQL tools: Case study in Quito Ecuador. Revista Técnica energía, 19(1), 158-168. 2022.
<https://doi.org/10.37116/revistaenergia.v19.n1.2022.532>
47. Maria Pilar Moreno, Sandra Martinez Cuevas, Xiomara Tamayo, & Teresa Iturrioz. Visualisation of Cadastral Lineage. 2017
<https://doi.org/10.1109/TLA.2017.8070433>
48. Universidad Del Norte - Observatorio de Seguridad Ciudadana. Informe 003. Barranquilla. 2020.
<https://www.uninorte.edu.co/web/observatorio-de-seguridad-ciudadana>
49. OpenAI. ChatGPT version 1 de mayo. 2023.
<https://chat.openai.com/chat>
50. Oracle. LinkedIn. 20 de noviembre de 2023.
<https://co.linkedin.com/company/oracle>
51. Oracl. PL/SQL hereda la robustez, la seguridad y la portabilidad de la base de datos. 2023.
<https://www.oracle.com/co/database/technologies/appdev/plsql.html>
52. Osorio González, W., & Rozo Patiño, A. Predicción del avalúo catastral de los predios urbanos en la ciudad de Pereira mediante aprendizaje de máquina. 2017.
<https://repositorio.utp.edu.co/items/29332b11-3898-4a2d-93f6-00dba5e6dad6>
53. Otarola, Y. La descripción del inmueble y la discordancia registral. 2022.
<http://dx.doi.org/10.7770/rchdcp-v13n1-art2863>
54. Pydata. seaborn. mayo de 2024.
<https://seaborn.pydata.org/generated/seaborn.boxplot.html>
55. Rosso Mateús, A. E., Montilla Montilla, Y. M. , & Garzón Martínez, S. C. Metodología para obtención y análisis de datos inmobiliarios usando fuentes alternativas: estudio de caso en tres ciudades intermedias de Colombia. 2022.
<https://doi.org/10.14483/23448393.17952>
56. Song, W., & Xiaoming Yang. A spatio-temporal cadastral data model based on space-time composite model. 2013.
<https://doi.org/10.1109/Geoinformatics.2013.6626047>
57. Subirats Maté, L., & Calvo González, M. Web scraping. Cataluña. 2021.

- https://openaccess.uoc.edu/bitstream/10609/147437/1/WebScraping_Modulo1_WebScraping.pdf
58. TÉLLEZ BUITRAGO, V. J., & MARTÍNEZ SÁNCHEZ, D. N. Método automático para la predicción del avalúo comercial de un inmueble en la ciudad de Bogotá. 2021.
<https://repository.ucatolica.edu.co/entities/publication/1c7a29cf-e871-413a-9317-d2835137c2b6>
59. Tic, G. Implementación de Modelos de Aprendizaje Automático. 2021.
<https://aulas.generaciontic.gov.co/my/courses.php>
60. Triple A S.A. E.S.P. Quiénes somos. 2023
<https://www.aaa.com.co/quienes-somos/>
61. Unda Gutiérrez, M. Los límites de la recaudación predial en los municipios urbanos de México: un estudio de casos. 2018.
<https://doi.org/10.24201/edu.v33i3.1741>
62. United Nations Committee of Experts on Global Geospatial Information Management. Framework for Effective Land Administration. 2020.
https://ggim.un.org/meetings/GGIM-committee/10th-Session/documents/E-C.20-2020-29-Add_2-Framework-for-Effective-Land-Administration.pdf