

Implementación de una metodología basado en IA para el análisis de la gestión pública: un enfoque innovador para combatir la pobreza en la ciudad de San José de Cúcuta

Nombres y apellidos

Jorge Enrique Rodríguez Guevara

Código estudiantil:

2022255946273

Tesis Doctoral presentada como requisito para optar el título de:
Doctor en Gestión de la Tecnología y la Innovación

Tutor(es):

Silvia Carolina Moreno Trillos

RESUMEN

El proyecto propone una metodología innovadora basada en inteligencia artificial (IA) para analizar los comportamientos socioeconómicos de la población de San José de Cúcuta, con el fin de fortalecer la toma de decisiones públicas orientadas a la reducción de la pobreza extrema. La ciudad enfrenta un escenario complejo marcado por migración masiva, violencia territorial, desigualdad estructural y limitaciones institucionales que han profundizado las brechas socioeconómicas. Las metodologías tradicionales de análisis, centradas principalmente en indicadores monetarios o categorizaciones simplificadas, han demostrado insuficiencia para comprender la diversidad y la multidimensionalidad de la pobreza en territorios fronterizos. Por ello, el estudio plantea la necesidad de incorporar técnicas avanzadas de IA como una alternativa estratégica, rigurosa y replicable para segmentar a la población e identificar con mayor precisión los núcleos de vulnerabilidad.

El marco conceptual se fundamenta en la pobreza multidimensional, particularmente en la metodología Alkire-Foster, en los enfoques de desarrollo humano de Amartya Sen y en modelos contemporáneos de análisis causal. La investigación reconoce la pobreza como un fenómeno estructural que abarca múltiples dimensiones: educación, salud, condiciones de vivienda, empleo, acceso a servicios básicos y participación social. Desde la perspectiva tecnológica, se integran principios de machine learning, aprendizaje no supervisado, inteligencia artificial explicable (XAI) e interoperabilidad de datos, relevando su utilidad para la gestión pública cuando se aplican de forma ética, transparente y con criterios robustos de calidad de datos.

Metodológicamente, la investigación se desarrolló en tres fases principales. La primera consistió en la recolección, depuración y normalización de datos provenientes de fuentes oficiales, censos, registros administrativos y encuestas. Se aplicaron procesos de limpieza, eliminación de duplicados, imputación de valores faltantes y codificación categórica, generando un dataset confiable y representativo. La segunda fase correspondió al análisis exploratorio y segmentación mediante algoritmos de aprendizaje no supervisado. Se emplearon K-Means++, DBSCAN y métricas como el coeficiente de Silhouette, correlaciones de Pearson y validaciones piloto para asegurar la consistencia de los hallazgos. La tercera fase integró los resultados del clustering con análisis cualitativo a través de un grupo focal, logrando triangulación metodológica y una interpretación más profunda de los patrones detectados.

Los resultados del modelado identificaron tres clústeres principales con características diferenciadas. El Clúster 0 agrupó a la población con educación superior completa y acceso estable a servicios públicos; el Clúster 1 representó un segmento en transición educativa (primaria completa, secundaria incompleta); y el Clúster 2 correspondió a la población con mayores necesidades básicas insatisfechas, alto analfabetismo y precariedad en el acceso a servicios esenciales. La selección de tres conglomerados fue respaldada por el método del codo y por un coeficiente de Silhouette promedio de 0.60, indicador de una separación adecuada entre grupos. Los análisis revelaron correlaciones significativas entre variables educativas, acceso a servicios

públicos, condiciones de vivienda y cohesión interna de los clústeres, confirmando la influencia estructural de la educación en la calidad de vida y el bienestar territorial.

La validación cualitativa mediante grupo focal permitió interpretar los patrones cuantitativos desde la experiencia de actores locales, destacando la relación entre educación y desarrollo comunitario, la importancia de los servicios públicos como motor de inclusión social y las brechas persistentes en territorios periféricos. Esta triangulación fortaleció la comprensión de las dinámicas sociales detectadas y aportó solidez a la metodología propuesta.

Desde la perspectiva de gestión pública, el estudio demuestra que la IA ofrece un potencial significativo para mejorar la focalización de programas sociales, optimizar la asignación de recursos y reducir la improvisación institucional. La segmentación permite identificar subgrupos con características específicas, facilitando políticas públicas diferenciadas según niveles de privación. Asimismo, la metodología fomenta la transparencia y la trazabilidad de las decisiones, gracias al uso de algoritmos explicables (XAI), que permiten comprender las razones detrás de las clasificaciones generadas por los modelos. Esto contribuye a fortalecer la confianza de la ciudadanía en los procesos públicos y a promover una cultura de toma de decisiones basada en evidencia verificable.

El proyecto aporta, además, una contribución relevante al campo académico y científico al integrar enfoques teóricos, análisis causal e inteligencia artificial para estudiar la pobreza en contextos urbanos y fronterizos. Se posiciona a San José de Cúcuta como una ciudad pionera en América Latina en la implementación de tecnologías emergentes para el análisis social, ofreciendo un modelo replicable para otros municipios con dinámicas similares.

Entre las conclusiones más importantes, la investigación evidencia que el uso de IA no solo mejora la eficiencia y precisión del análisis socioeconómico, sino que también permite identificar patrones latentes invisibles para metodologías tradicionales. La segmentación poblacional ofrece una mirada más completa y justa de las necesidades comunitarias, permitiendo intervenciones más focalizadas y efectivas. Finalmente, el estudio reafirma la necesidad de invertir en

interoperabilidad de datos, infraestructura tecnológica y capacidades analíticas en el sector público, para garantizar la sostenibilidad y escalabilidad de estas herramientas.

La metodología propuesta constituye un instrumento robusto, replicable y ético para el análisis de la pobreza y la gestión pública, capaz de transformar la forma en que las instituciones comprenden, atienden y previenen las desigualdades sociales. Su implementación podría marcar un avance significativo hacia un modelo de gobernanza más eficiente, transparente y orientado a resultados, alineado con los retos actuales de territorios con alta vulnerabilidad social como San José de Cúcuta.

Palabras clave: Inteligencia Artificial, Pobreza Multidimensional, Gestión Pública, Segmentación Socioeconómica y Machine Learning

ABSTRACT

The project proposes an innovative methodology based on artificial intelligence (AI) to analyze the socioeconomic behaviors of the population of San José de Cúcuta, with the aim of strengthening public decision-making focused on reducing extreme poverty. The city faces a complex scenario marked by mass migration, territorial violence, structural inequality, and institutional limitations that have deepened socioeconomic gaps. Traditional analytical methodologies, primarily focused on monetary indicators or simplified categorizations, have proven insufficient to understand the diversity and multidimensionality of poverty in border regions. Therefore, the study argues for the need to incorporate advanced AI techniques as a strategic, rigorous, and replicable alternative for segmenting the population and more accurately identifying areas of vulnerability.

The conceptual framework is based on multidimensional poverty, particularly the Alkire-Foster methodology, Amartya Sen's human development approaches, and contemporary models of causal analysis. The research recognizes poverty as a structural phenomenon encompassing multiple dimensions: education, health, housing conditions, employment, access to basic services, and social participation. From a technological perspective, it integrates principles of machine learning, unsupervised learning, explainable artificial intelligence (XAI), and data interoperability, highlighting their usefulness for public administration when applied ethically, transparently, and with robust data quality criteria.

Methodologically, the research was developed in three main phases. The first consisted of collecting, cleaning, and normalizing data from official sources, censuses, administrative records, and surveys. Processes such as cleaning, duplicate removal, imputation of missing values, and categorical coding were applied, generating a reliable and representative dataset. The second phase involved exploratory analysis and segmentation using unsupervised learning algorithms. K-Means++, DBSCAN, and metrics such as the Silhouette coefficient, Pearson correlations, and pilot validations were used to ensure the consistency of the findings. The third phase integrated the

clustering results with qualitative analysis through a focus group, achieving methodological triangulation and a deeper interpretation of the detected patterns.

The modeling results identified three main clusters with distinct characteristics. Cluster 0 grouped the population with completed higher education and stable access to public services; Cluster 1 represented a segment in educational transition (completed primary education, incomplete secondary education); and Cluster 2 corresponded to the population with the greatest unmet basic needs, high illiteracy rates, and precarious access to essential services. The selection of three clusters was supported by the elbow method and an average Silhouette coefficient of 0.60, indicating adequate separation between groups. The analyses revealed significant correlations between educational variables, access to public services, housing conditions, and internal cluster cohesion, confirming the structural influence of education on quality of life and territorial well-being.

Qualitative validation through focus groups allowed for the interpretation of quantitative patterns from the perspective of local stakeholders, highlighting the relationship between education and community development, the importance of public services as drivers of social inclusion, and the persistent gaps in peripheral areas. This triangulation strengthened the understanding of the identified social dynamics and lent rigor to the proposed methodology.

From a public management perspective, the study demonstrates that AI offers significant potential for improving the targeting of social programs, optimizing resource allocation, and reducing institutional improvisation. Segmentation allows for the identification of subgroups with specific characteristics, facilitating differentiated public policies based on levels of deprivation. Furthermore, the methodology fosters transparency and traceability of decisions, thanks to the use of explainable algorithms (XAI), which allow for understanding the reasons behind the classifications generated by the models. This contributes to strengthening public trust in government processes and promoting a culture of decision-making based on verifiable evidence.

The project also makes a significant contribution to the academic and scientific field by integrating theoretical approaches, causal analysis, and artificial intelligence to study poverty in urban and

border contexts. It positions San José de Cúcuta stands out as a pioneering city in Latin America in the implementation of emerging technologies for social analysis, offering a replicable model for other municipalities with similar dynamics.

Among the most important findings, the research demonstrates that the use of AI not only improves the efficiency and accuracy of socioeconomic analysis but also allows for the identification of latent patterns invisible to traditional methodologies. Population segmentation provides a more comprehensive and equitable view of community needs, enabling more targeted and effective interventions. Finally, the study reaffirms the need to invest in data interoperability, technological infrastructure, and analytical capabilities in the public sector to ensure the sustainability and scalability of these tools.

The proposed methodology constitutes a robust, replicable, and ethical instrument for poverty analysis and public management, capable of transforming how institutions understand, address, and prevent social inequalities. Its implementation could mark a significant step towards a more efficient, transparent, and results-oriented governance model, aligned with the current challenges of territories with high social vulnerability, such as San José de Cúcuta. **Keywords:** Artificial intelligence, multidimensional poverty, public management, socioeconomic segmentation, and machine learning.

Keywords: Artificial Intelligence, Multidimensional Poverty, Public Management, Socioeconomic Segmentation, and Machine Learning

REFERENCIAS BIBLIOGRÁFICAS

1. Alcaldía de Cúcuta. (2022). *¿Cómo influye el Sisbén en el mejoramiento de los indicadores de pobreza del DANE en Cúcuta?* <https://cucuta.gov.co/como-influye-el-sisben-en-el-mejoramiento-de-los-indicadores-de-pobreza-del-dane-en-cucuta/>
2. Alvarado Rojas, M. E. (2015). Una mirada a la inteligencia artificial. *Revista Ingeniería, Matemáticas y Ciencias de la Información*, 2(3). <https://ojs.urepublicana.edu.co/index.php/ingenieria/article/view/234>
3. Amaya, A., Huerta, F., & Flores, C. O. (2020). Big Data, una estrategia para evitar la deserción escolar en las IES. *Revista de Investigación Educativa*, 31(1), 712-730. <https://doi.org/10.22201/iisue.20072872e.2020.31.712>
4. Amaya et al. (2020). Estrategia contra la deserción en Instituciones de Educación Superior (IES): El proceso de tomar la decisión de usar Big Data. *Revista Latinoamericana de Educación*, 15(2), 45–67.
5. Arroyo, D. (2006). *Tecnología y reducción de la pobreza en América Latina*. CEPAL.
6. Bangsawn, S. (2024). *AI for poverty reduction: Global case studies*. MIT Press.
7. Bautista Martínez, N. D., Gómez Calderón, K. M., Lizarazo Rodríguez, A. F., Valencia Angarita, E. C., & Zambrano Carrillo, M. F. (2020). *Adaptación de las familias frente a situaciones de pobreza en la comuna n.º 1: Barrio El Contento y El Páramo de la ciudad de San José de Cúcuta*. Universidad Simón Bolívar. <https://bonga.unisimon.edu.co/server/api/core/bitstreams/b7c99e2c-50f5-48d8-9fbf-b6b92f3f1e3c/content>
8. Bhattacharjee, A. (2020). *Social science research: Principles, methods, and practices* (3rd ed.). University of South Florida Scholar Commons.
9. Bellary, S., Kumar, P., & Jain, R. (2019). Validation metrics in machine learning: A comparative analysis. *Journal of Data Science*, 17(3), 210–225.
10. Bertsimas, D., & Dunn, J. (2017). *Machine learning under a modern optimization lens*. Dynamic Ideas.
11. Burt, M. (2015). Semáforo de Eliminación de Pobreza: Una herramienta para empoderar a las familias y mejorar las políticas. *World Economic Forum*. <https://www.weforum.org/agenda/2015/09/how-technology-can-help-us-eliminate-not-alleviate-poverty/>
12. Cámara de Comercio de Cúcuta. (2022). *Informe de calidad de vida 2022*. <https://cucutacomovamos.com/wp-content/uploads/2023/02/Informe-de-Calidad-de-Vida-2022.pdf>
13. Campos Acuña, M. (2019). Inteligencia artificial e innovación en la administración pública: (In)necesarias regulaciones para la garantía del servicio público. *Revista Vasca de Gestión Pública*, 74, 74–91. https://www.ivap.euskadi.eus/contenidos/informacion/rvvp_ultimo_numero/es_def/Campos%2074_91.pdf
14. Cano Molina, A. E., & Diaz Penagos, G. M. (2023). *Retos de la inteligencia artificial en Colombia: Un diagnóstico de los principales desafíos*. ESAP. <https://repositoriocdim.esap.edu.co/bitstream/handle/123456789/27210/ANDRES%20EDU ARDO%20CANO%20MOLINA.pdf>
15. Castro Contreras, J. D. (2019). Aplicación de la IA para la sostenibilidad en las organizaciones. *Universidad El Bosque*. <https://repositorio.unbosque.edu.co/server/api/core/bitstreams/b2119622-f892-49d8-b00c-67d83fe4d0b2/content>
16. Charris, L., Henríquez, C., Hernández, S., Jimeno, L., Guillén, O., & Moreno, S. (2018). Análisis comparativo de algoritmos de árboles de decisión en el procesamiento de datos biológicos. *Investigación y Desarrollo en TIC*, 9(1), 26–34.
17. Chkoniya, V. (Ed.). (2021). *Handbook of research on applied data science and artificial intelligence in business and industry*. IGI Global. <https://doi.org/10.4018/978-1-7998-6985-6>

18. DANE. (2021). *La información del DANE en la toma de decisiones regionales* (Vol. 219). <https://www.dane.gov.co/files/investigaciones/planes-departamentos-ciudades/210319-InfoDane-Cucuta-Norte-de-Santander.pdf>
19. Delgado Vargas, D. (2023). *Caracterización de los municipios de Colombia según el potencial de crecimiento del ingreso por impuesto predial utilizando Machine Learning* [Tesis de maestría, Universidad de La Sabana]. https://intellectum.unisabana.edu.co/bitstream/handle/10818/59338/Proyecto_Grado_John_Daniel_Delgado_Biblioteca.pdf
20. Departamento Nacional de Planeación (DNP). (2021). *Mercado laboral urbano – Resultados 2021: Cúcuta*. <https://colaboracion.dnp.gov.co/CDT/Estudios%20Economicos/6%20Informe%20C3%BAcuta%202021%20-%20r.pdf>
21. Durán, J., & Condori, A. (2017). Desarrollo y estimación de un índice de vulnerabilidad socio energética para departamentos de Argentina. *Universidad Nacional de La Plata*. https://sedici.unlp.edu.ar/bitstream/handle/10915/140360/Documento_completo.pdf
22. Easton, G. (2010). Critical realism in case study research. *Industrial Marketing Management*, 39(1), 118–128.
23. Fletcher, A. J. (2017). Applying critical realism in qualitative research: Methodology meets method. *International Journal of Social Research Methodology*, 20(2), 181–194.
24. Gallego, J. (2021). Impacto del programa Ingreso Solidario en Colombia. *Economía & Desarrollo*, 25(1), 89–104.
25. Gilabert López, J. (2025). *K-Means clustering: Principles and applications*. CRC Press.
26. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
27. Granados, F., & Jackeline, M. (2023). El impacto de la inteligencia artificial en los trabajadores despedidos por automatización de servicios. *Revista de Economía Laboral*, 14, 1–15.
28. Hall, O., Ohlsson, M., & Rögnvaldsson, T. (2022). A review of explainable AI in the satellite data, deep machine learning, and human poverty domain. *Patterns*, 3(10), 100600. <https://doi.org/10.1016/j.patter.2022.100600>
29. Herrera, T. J. F., Maturana, A. G. B., & Zambrano, F. M. (2024). Violent deaths during and after the pandemic in Colombia: An artificial intelligence analysis. *Revista Científica General José María Córdova*, 22(45), 175–197. <https://doi.org/10.21830/19006586.1271>
30. Holguín Ontiveros, E. P., & Menendez Mora, R. E. (2018). Aplicación del aprendizaje automático en la clasificación de textos cortos: Un caso de estudio en el conflicto armado colombiano. *Universidad Católica de Colombia*. <https://repository.ucatolica.edu.co/server/api/core/bitstreams/3bdfb760-c23a-41ef-84e5-dc3bc3694d89/content>
31. Instituto de Estadística de la UNESCO (UIS). (2023). *Datos sobre educación terciaria en América Latina*. <http://uis.unesco.org/es/tema/educacion-terciaria>
32. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer. <https://doi.org/10.1007/978-1-0716-1418-1>
33. López, R., & Márquez, P. (2019). Ética e inteligencia artificial: Retos para la política pública. *Revista Latinoamericana de Políticas Públicas*, 11(2), 98–115.
34. Márquez, C., & Rincón, H. (2020). Inteligencia artificial y brechas sociales: Una aproximación crítica desde América Latina. *Revista de Estudios Sociales*, 71, 55–70. <https://doi.org/10.7440/res71.2020.04>
35. Martínez, S., & Pérez, M. (2021). Minería de datos y políticas públicas: Aplicaciones en salud y educación. *Revista de Políticas Públicas*, 17(2), 203–220.
36. Martínez-Vargas, J. R., Rodríguez, E. G., Rojas, J. A., & Rivera, F. G. (2022). Aplicación de técnicas de aprendizaje automático para la predicción de abandono escolar en Colombia. *Revista Colombiana de Computación*, 23(1), 45–59.

37. Maxwell, J. A. (2021). *A realist approach for qualitative research* (2nd ed.). SAGE Publications.
38. Ministerio de Ciencia, Tecnología e Innovación. (2022). *Misión de Sabios: Inteligencia artificial y tecnologías convergentes*. <https://minciencias.gov.co/mision-sabios/informes-finales>
39. Ministerio de Educación Nacional. (2021). *Informe de avances del Plan Nacional Decenal de Educación*. <https://www.mineducacion.gov.co/plan-nacional>
40. Moreno, A. I., & Porras, A. (2023). Modelos predictivos para focalización del gasto social en Colombia: Una evaluación empírica. *Revista de Economía del Sector Público*, 15(1), 113–136.
41. Ocampo, J. A., & Vallejo, J. (2020). Políticas públicas y pobreza en Colombia: Una revisión crítica. *Revista CEPAL*, 131, 43–60.
42. OECD. (2021). *AI in the public sector: Risks and opportunities*. <https://www.oecd.org/gov/ai-in-the-public-sector.htm>
43. PNUD. (2022). *Índice de desarrollo humano y pobreza multidimensional en América Latina*. <https://www.undp.org/es/latin-america>
44. Rangel, C. A., & Méndez, L. A. (2018). Análisis de conglomerados para segmentación de hogares pobres en zonas urbanas. *Revista de Estadística Aplicada*, 12(2), 97–112.
45. Riquelme, F., & González, R. (2016). Machine learning for poverty prediction: A comparative study using household surveys. *Journal of Data Science for Policy*, 2(1), 1–18.
46. Ritchie, J., Lewis, J., Nicholls, C. M., & Ormston, R. (Eds.). (2013). *Qualitative research practice: A guide for social science students and researchers*. SAGE.
47. Rueda, A. (2023). Analítica de datos y decisiones públicas: El caso de la focalización de subsidios en Colombia. *Revista Estudios de Administración Pública*, 94, 73–94.
48. Salinas, D., & Ríos, M. (2021). Inteligencia artificial en la administración pública: Usos, desafíos y recomendaciones. *Revista Chilena de Administración Pública*, 17(1), 25–48.
49. Sánchez-Torres, J. A., & Ovalle, D. A. (2021). Aplicación de modelos de aprendizaje automático para la predicción de pobreza multidimensional en Colombia. *Revista Ingenierías Universidad de Medellín*, 20(38), 161–179. <https://doi.org/10.22395/rium.v20n38a9>
50. Sen, A. (1999). *Development as freedom*. Oxford University Press.
51. Trujillo, A., & Prada, S. (2017). Big Data y sistemas de salud en América Latina: Potencialidades y desafíos. *Revista Panamericana de Salud Pública*, 41, e17. <https://doi.org/10.26633/RPSP.2017.17>
52. UNESCO. (2021). *La inteligencia artificial en América Latina: Panorama y desafíos*. https://unesdoc.unesco.org/ark:/48223/pf0000375700_spa
53. Urueña, R., & Guzmán, D. (2022). Inteligencia artificial, derechos humanos y políticas públicas: Un análisis desde el contexto colombiano. *Revista Derecho del Estado*, 53, 55–84.
54. UNESCO. (2021). *Global education monitoring report 2021/2: Non-state actors in education*. United Nations Educational, Scientific and Cultural Organization.
55. UNICEF & OMS. (2023). *Impacto de la nutrición en la mortalidad infantil*. Fondo de las Naciones Unidas para la Infancia y Organización Mundial de la Salud.
56. Unidad para las Víctimas. (2022). *Informe de desplazamiento forzado 2020 – 2021*. <https://www.unidadvictimas.gov.co/>
57. Valencia, M., & Suárez, L. (2020). Aplicaciones de minería de datos en programas de subsidios sociales: Experiencia del SISBÉN IV. *Revista de Políticas Sociales*, 19(2), 90–110.
58. Valle-Cruz, D., Sandoval-Almazán, R., & Gil-García, J. R. (2023). Predictive models for budget optimization in Mexican public administration. *Government Information Quarterly*, 40(1), 103–120.

59. Valladares, C. (2018). IA y transparencia en programas alimentarios: el caso del vaso de leche. *Gestión Pública y Ética*.
60. Vanegas, A., Rojas, L., & Cifuentes, M. (2024). Sesgos y opacidad en algoritmos públicos: desafíos éticos de la IA en Colombia. *Cuadernos de Gobierno Digital*.
61. Vélez Herrera, A. (2007). *Algoritmos genéticos y sistemas expertos en la gestión del conocimiento*. Universidad Nacional de Colombia.
62. Villanueva, R. J., & Pérez, A. M. (2023). Técnicas de clustering aplicadas al análisis de pobreza multidimensional: Un estudio empírico en Colombia. *Revista Colombiana de Estadística*, 46(1), 33–54. <https://doi.org/10.15446/rce.v46n1.95817>
63. World Bank. (2022). *Digital government and data-driven policy in Latin America*. World Bank Group.
64. Zhang, W., Liu, Y., & Chen, H. (2022). Explainable artificial intelligence in public policy: Challenges and frameworks. *AI & Society*, 37(3), 621–638. <https://doi.org/10.1007/s00146-021-01237-w>
65. Zhang, X., Wang, L., & Liu, Y. (2022). Explainable AI in policy design and public accountability. *AI & Society*, 37(3), 599–618.