

FICHA RESUMEN

PROYECTO APLICADO – MAESTRÍA EN CIENCIA DE DATOS

TÍTULO: Impacto de la Expansión Urbana en la Pobreza: Análisis Geoespacial y Socioeconómico en Bogotá D.C.

TÍTULO: Impacto de la Expansión Urbana en la Pobreza: Análisis Geoespacial y Socioeconómico en Bogotá D.C.

1. **ÁREA DE TRABAJO:** Percepción remota.
2. **TIPO DE PROYECTO:** Aplicado.
3. **ESTUDIANTE(S):** Julián Felipe Castiblanco Hernández, Christian David Vera Mendivelso, Andrew Rafael Arnedo Pertuz.
4. **CORREO ELECTRÓNICO:** juliancastiblanco@javerianacali.edu.co, andrewarnedo@javerianacali.edu.co, christianvera@javerianacali.edu.co.
5. **DIRECCIÓN Y TELÉFONO:**
6. **DIRECTOR:** Yady Tatiana Solano Correa.
7. **VINCULACIÓN DEL DIRECTOR:** Pontificia Universidad Javeriana.
8. **CORREO ELECTRÓNICO DEL DIRECTOR:** tatiana.solano@javerianacali.edu.co.
9. **CO-DIRECTOR :** Mario Milver Patiño Velasco.
10. **GRUPO O EMPRESA QUE LO AVALA :** Universidad del Cauca.
11. **OTROS GRUPOS O EMPRESAS:** No aplica.
12. **PALABRAS CLAVE:** Percepción remota, pobreza, expansión urbana, series de tiempo, machine learning, análisis geoespacial.
13. **FECHA DE INICIO:** Noviembre de 2024.
14. **FECHA DE FINALIZACIÓN:** Noviembre de 2025.
15. **RESUMEN:** El proyecto se centra en analizar la influencia de la expansión urbana en los índices de pobreza en la ciudad de Bogotá D.C., abordando un tema crucial para la planificación urbana y el desarrollo social. En particular, resulta esencial comprender cómo el crecimiento descontrolado de la ciudad afecta la distribución de la pobreza y la desigualdad.

Este fenómeno tiene un impacto significativo en las áreas periféricas, que suelen experimentar un desarrollo informal y enfrentan serias limitaciones en el acceso a servicios básicos, perpetuando ciclos de vulnerabilidad y exclusión social. La investigación parte del reconocimiento de la creciente brecha socioeconómica en la ciudad, evidenciada por la expansión urbana desregulada, agravando los niveles de pobreza. Para abordar este problema, se integraron datos geoespaciales usando imágenes satelitales, junto con el índice de pobreza multidimensional, como indicador socioeconómico, para identificar las dinámicas de expansión urbana en cada una de las UPZ de la ciudad, en la última década (2014–2024) y analizar cómo estas influyen en los patrones de pobreza. Se analizó y explicó por qué la densidad poblacional, el porcentaje de urbanización, entre otras métricas, impactan de manera distinta al norte, el centro y el sur de la ciudad. Finalmente, se desarrollaron visualizaciones interactivas que facilitan la interpretación de los resultados, contribuyendo así a la generación de conocimiento útil para la toma de decisiones en políticas públicas y planificación urbana.



**IMPACTO DE LA EXPANSIÓN URBANA EN LA POBREZA:
ANÁLISIS GEOESPACIAL Y SOCIOECONÓMICO EN BOGOTÁ
D.C.**

Autores:

Julián Felipe Castiblanco Hernández – 9014955

Christian David Vera Mendivelso – 9015043

Andrew Rafael Arnedo Pertuz – 9015410

Proyecto Aplicado para optar al título de
Magíster en Ciencia de Datos

Directora

Yady Tatiana Solano Correa

Codirector

Mario Milver Patiño Velasco

**FACULTAD DE INGENIERÍA Y CIENCIAS
MAESTRÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI, DICIEMBRE 1 DE 2025**

TABLA DE CONTENIDO

1. Definición del problema	3
1.1. Planteamiento del problema	3
1.2. Formulación del problema	4
2. Objetivos del proyecto	5
2.1. Objetivo general	5
2.2. Objetivos específicos	5
3. Marco Teórico	6
3.1. Expansión Urbana	6
3.2. Métodos de medición de la expansión urbana	6
3.3. Espacial	7
3.3.1. Imágenes satelitales	7
3.3.2. Percepción Remota	7
3.3.3. Landsat	8
3.3.4. Otros satélites	10
3.3.5. Técnicas de preprocesamiento de imágenes	10
3.3.6. Índices espectrales	11
3.3.7. Detección de cambios de imágenes	12
3.3.8. Métodos de detección de cambios en imágenes satelitales	12
3.3.9. Separabilidad espectral entre clases	12
3.4. Métodos de clasificación	14
3.4.1. Clasificación mediante métodos de umbralización (no supervisados)	14
3.4.2. Clasificación mediante métodos supervisados	16
3.4.3. Métricas de Evaluación	18
3.5. Demográfica	19
3.5.1. Densidad urbana	20
3.5.2. Ordenamiento territorial	20
3.5.3. Migración	20
3.5.4. Gentrificación	20
3.6. Pobreza	21
3.6.1. Métodos de medición de la pobreza	21
3.6.2. Pobreza monetaria	21
3.6.3. Índice de pobreza multidimensional	22
3.6.4. Estimación por coincidencia de cuantiles (Quantile Matching) (MQE)	23
3.6.5. Encuesta Multipropósito	23
3.7. Relación entre expansión urbana con la desigualdad	24

3.8.	Técnicas para relacionar la expansión urbana y la pobreza	24
3.9.	Métodos de correlación	25
3.9.1.	Índice de Morans	25
3.9.2.	Mapas LISA (Local Indicators of Spatial Association)	26
3.9.3.	Modelo de Regresión Ponderada Geográficamente (GWR)	26
3.10.	Visualización interactiva de datos espaciales	27
3.11.	Antecedentes	28
4.	Construcción de base de datos de imágenes satelitales e indicadores socio-económicos de pobreza	32
4.1.	Procesamiento de imágenes	32
4.1.1.	Descripción de las imágenes utilizadas	33
4.1.2.	Área de estudio	33
4.1.3.	Preprocesamiento de las Imágenes	34
4.1.4.	Selección del área de estudio y filtrado de la colección de Landsat-8	34
4.1.5.	Enmascaramientos de calidad	35
4.1.6.	Construcción de composiciones anuales	35
4.2.	Procesamiento de bases de datos pobreza	38
4.2.1.	Extracción de los datos de la EM	38
4.2.2.	Cálculo del IPM	40
4.2.3.	Aplicación de modelo de estimación	41
5.	Identificación de dinámicas de expansión urbana usando técnicas de aprendizaje de máquina	43
5.1.	Clasificación supervisada	43
5.1.1.	Generación y preparación de muestras de entrenamiento	44
5.1.2.	Evaluación de la separabilidad de las clases extraídas	46
5.1.3.	Desarrollo de los modelos supervisados	46
5.1.4.	Entrenamiento de los modelos	47
5.1.5.	Clasificación mediante Random Forest	48
5.1.6.	Clasificación mediante SVM	50
5.1.7.	Clasificación mediante regresión logística	51
5.1.8.	Clasificación mediante XGBoost	53
5.1.9.	Comparativa entre métodos supervisados	54
5.2.	Métodos no supervisados de umbralización	56
5.3.	Reclasificación de coberturas	56
6.	Evaluación el modelo escogido y caracterización de las variables relacionadas	58

6.1. Caracterizaciones	61
6.1.1. Expansión	61
6.1.2. Urbanización	62
6.1.3. Población	63
6.1.4. Pobreza	63
6.1.5. Densidad	65
6.2. Métodos de análisis estadístico	65
6.2.1. Análisis de correlación	65
6.3. Índices de Morán	67
6.4. Mapas LISA	68
6.4.1. IPM	68
6.4.2. Densidad	69
6.4.3. Urbanización	71
6.5. Modelo de regresión ponderada geográficamente (GWR)	72
6.5.1. Proporción de expansión	73
6.5.2. Densidad	74
6.5.3. Urbanización	75
6.6. Dinámicas socioespaciales y económicas	76
7. Desarrollo de visualizaciones interactivas	79
7.1. Selección del entorno	79
7.2. Desarrollo de visualizaciones	79
7.2.1. Clasificación del suelo urbano	79
7.2.2. Población en condición de pobreza 2014, 2017 y 2021	81
7.2.3. Expansión Urbana	83
7.2.4. Relación Expansión Urbana y Pobreza	84
8. Conclusiones y trabajos futuros	85
8.1. Conclusiones	85
8.2. Trabajos futuros	87
8.2.1. Integración de imágenes satelitales de mayor resolución espacial y temporal	87
8.2.2. Ampliación del análisis a nivel regional o nacional	87
8.2.3. Incorporación de variables ambientales y de sostenibilidad	87
8.2.4. Desarrollo de modelos predictivos y escenarios de crecimiento urbano	88
9. Referencias	98

LISTA DE TABLAS

3.1. Características de los sensores OLI y TIRS de Landsat-8	9
3.2. Dimensiones y variables del IPM para Colombia (Adaptado de [46]).	22
4.1. Variables con valores faltantes de las encuestas multipropósito	40
4.2. Resumen de porcentajes de personas en condición de pobreza por año	42
5.1. Métricas de Separabilidad Espectral entre las Clases	46
5.2. Espacio de búsqueda de hiperparámetros usados en Random Forest	49
5.3. Mejor configuración de hiperparámetros en Random Forest por tipo de entrada .	49
5.4. Comparativa de desempeño de Random Forest según tipo de variables de entrada	50
5.5. Espacio de búsqueda de hiperparámetros usados en SVM	50
5.6. Mejor configuración de hiperparámetros en SVM por tipo de entrada	51
5.7. Comparativa de desempeño de SVM según tipo de variables de entrada	51
5.8. Espacio de búsqueda de hiperparámetros usados en regresión logística	52
5.9. Mejor configuración de hiperparámetros en regresión logística por tipo de entrada	52
5.10. Comparativa de desempeño de Regresión Logística según tipo de variables de entrada	52
5.11. Espacio de búsqueda de hiperparámetros usados en XGBoost	53
5.12. Mejor configuración de hiperparámetros en XGBoost por tipo de entrada	54
5.13. Comparativa de desempeño de XGBoost según tipo de variables de entrada	54
5.14. Comparativa de mejores modelos según su F1 ponderado	55
5.15. Comparativa de métodos de umbralización vs Random Forest	56
6.1. Impacto del post-procesamiento en la estimación del área urbana	58
6.2. Dinámica de Urbanización y Crecimiento Urbano Anual (2014–2024)	59
6.3. Evolución del índice de Moran global	67

LISTA DE FIGURAS

4.1. Diagrama de flujo preprocesamiento de imágenes Landsat 8	32
4.2. Área de estudio seleccionada	34
4.3. Visualización de Índices Espectrales – 2014	37
4.4. Resultados del preprocesamiento de las imágenes Landsat 8 en el periodo 2013-2024	38
4.5. Diagrama de flujo metodológico para el cálculo del IPM, para los años 2014, 2017 y 2021	39
5.1. Diagrama de flujo metodológico para la clasificación de cobertura urbana en la ciudad de Bogotá D.C.	43
5.2. Importancia relativa de las variables predictoras en el modelo Random Forest (entrada: bandas+índices).	55
6.1. Cambios en el área urbana 2014-2024.	60
6.2. Diagrama de flujo metodológico para evaluar la caracterización de las relaciones entre las variables.	61
6.3. Diagrama de caja de proporción de expansión por UPZ en Bogotá D.C. (2014 – 2024) (Con 2014 como año base).	62
6.4. Diagrama de caja de % de urbanización por UPZ en Bogotá D.C. (2014 – 2024).	62
6.5. Diagrama de caja de la población total en Bogotá D.C. (2014 – 2024).	63
6.6. Diagrama de caja del IPM por año.	64
6.7. Distribución del IPM para los años 2014, 2017 y 2021.	64
6.8. Diagrama de caja de densidad poblacional	65
6.9. Matriz de correlación de variables seleccionadas.	66
6.10. Evolución Espaciotemporal de los Clústeres LISA del IPM (2014-2021).	68
6.11. Evolución Espaciotemporal de los Clústeres LISA de la densidad poblacional (2014-2021).	70
6.12. Evolución espaciotemporal de los clústeres LISA del porcentaje de urbanización (2014-2021).	71
6.13. Mapa GWR de la proporción de expansión (2017 – 2021).	73
6.14. Mapa GWR de la densidad poblacional (2014 – 2021).	74
6.15. Mapa GWR del porcentaje de urbanización (2014 – 2021).	75
7.1. Visualización de clasificación del suelo urbano por año.	80
7.2. Visualización gráfica de barras de área urbana y no urbana para ejemplo en el año 2024.	81
7.3. Visualización población en condición de pobreza (2014,2017 y 2021).	82
7.4. Visualización población en condición de pobreza para el año 2021.	82
7.5. Visualización de la evolución del crecimiento urbano por UPZ.	83
7.6. Visualización del mapa LISA para la densidad poblacional.	84

Introducción

La expansión urbana es un fenómeno inherente al crecimiento poblacional y económico de las ciudades, pero su desarrollo descontrolado puede generar profundas desigualdades sociales y económicas. En Bogotá D.C., la urbanización acelerada, caracterizada por la proliferación de asentamientos informales en las periferias, ha incrementado las brechas socioeconómicas y ha intensificado los niveles de pobreza en zonas vulnerables. Este crecimiento, impulsado por factores como la migración interna y la falta de planificación territorial adecuada, ha resultado en la exclusión de amplios sectores de la población de los beneficios del desarrollo urbano formal. La concentración de recursos y servicios en las zonas centrales de la ciudad contrasta con las carencias que enfrentan las áreas periféricas, donde las familias viven en condiciones de precariedad, sin acceso adecuado a servicios básicos como agua, alcantarillado, educación y salud.

Un modo de explorar las consecuencias de la expansión urbana es mediante el análisis de los índices de pobreza, los cuales reflejan las condiciones de vida de la población y su acceso a recursos básicos. En este contexto, resulta imperativo examinar la relación entre la expansión urbana y dichos índices, no sólo para comprender las dinámicas que perpetúan la desigualdad, sino también para diseñar intervenciones que promuevan un desarrollo urbano más equitativo y sostenible. Este proyecto busca abordar esta problemática integrando herramientas de análisis geoespacial y datos socioeconómicos, lo que permitirá una exploración precisa y detallada del impacto de la expansión de la mancha urbana en la distribución de la pobreza en Bogotá D.C.

El uso de imágenes satelitales y modelos de ciencia de datos ofrece una oportunidad única para observar patrones espaciales de urbanización y pobreza con un alto nivel de detalle. Las técnicas de clasificación, detección de cambios y análisis de correlación no solo permiten identificar las áreas más afectadas, sino también evaluar las dinámicas de crecimiento urbano y sus consecuencias sociales. Estos hallazgos, presentados a través de visualizaciones interactivas, facilitarán la comunicación de resultados a las personas interesadas (analistas, gobierno, ciudadanía, entre otros), fomentando una evaluación y toma de decisiones conscientes, basadas en datos.

En un contexto global donde las ciudades enfrentan retos cada vez mayores para garantizar la sostenibilidad y la inclusión social, este proyecto ofrece un marco analítico y práctico para enfrentar las consecuencias de la urbanización descontrolada. Al aplicar un enfoque interdisciplinario que combina ciencias de datos, análisis geoespacial y estudios socioeconómicos, se pretende no sólo generar conocimiento relevante, sino también aportar soluciones concretas para promover un desarrollo urbano más justo y equitativo en la ciudad.

Por tanto, este proyecto tiene como objetivo principal analizar la influencia de la expansión

urbana en los índices de pobreza en Bogotá D.C. durante una década. Para ello, se propone consolidar bases de datos geospaciales y socioeconómicas, evaluar modelos analíticos que integren esta información y generar visualizaciones que permitan interpretar las dinámicas encontradas. A través de este enfoque, se espera contribuir al desarrollo de herramientas que orienten la planificación urbana y la focalización de inversiones en infraestructura y servicios básicos, reduciendo así la desigualdad en las zonas más vulnerables de la ciudad.

1. Definición del problema

1.1. Planteamiento del problema

La expansión urbana descontrolada en Bogotá D.C. ha generado un crecimiento urbano desigual que impacta profundamente en las áreas periféricas, particularmente en los asentamientos informales. Estos sectores enfrentan condiciones de vida deterioradas debido a la falta de planificación urbana efectiva y la ausencia de políticas de desarrollo territorial inclusivas. Este fenómeno, impulsado por el rápido aumento poblacional y la migración hacia el centro urbano, ha dado lugar a una serie de problemáticas, que incluyen el acceso limitado a servicios básicos. Las condiciones de informalidad en las periferias de Bogotá D.C. se traducen en altos niveles de pobreza y exclusión social, afectando directamente la calidad de vida de sus habitantes y perpetuando un ciclo de desigualdad en el que la pobreza se concentra en los bordes de la ciudad, aumentando así la vulnerabilidad social.

La centralización de recursos, servicios y oportunidades en las zonas centrales contribuye a la marginación de las áreas periféricas, incrementando la brecha socioeconómica entre el centro y los bordes de la ciudad. Este patrón de desarrollo no solo fomenta una segregación espacial en Bogotá D.C., sino que también obstaculiza la sostenibilidad urbana, ya que una parte importante de la población queda excluida de los beneficios del desarrollo formal, debilitando la capacidad de la ciudad para construir un futuro inclusivo y equitativo [1].

Esta problemática representa una oportunidad para aplicar herramientas avanzadas de análisis geoespacial junto con datos socioeconómicos, que permitan una comprensión detallada de la relación entre la expansión urbana y la concentración de la pobreza. La integración de datos satelitales con indicadores, como el índice de pobreza multidimensional, podría ofrecer la posibilidad de observar patrones espaciales de pobreza con un alto nivel de precisión, ayudando a identificar y caracterizar las áreas más afectadas y sus dinámicas de exclusión. Mediante este enfoque, es posible sentar las bases para formular políticas públicas informadas y basadas en datos, que orienten la integración de las zonas periféricas en el desarrollo urbano formal y promuevan una distribución equitativa de los recursos y las oportunidades. Este enfoque es especialmente relevante en una ciudad como Bogotá D.C., donde el crecimiento urbano sigue una tendencia acelerada y en gran medida informal, lo que hace urgente la intervención estratégica y planificada para evitar una profundización de la desigualdad social y permitir que la expansión urbana no represente una amenaza a la sostenibilidad y equidad de la ciudad [2].

Esta investigación busca abordar una problemática actual y urgente en Bogotá D.C., donde el uso de la ciencia de datos puede contribuir de manera significativa a la comprensión y medición de la pobreza en su dimensión espacial.

1.2. Formulación del problema

En esta investigación, la pregunta que guía el análisis es: ¿Cómo ha influido la expansión urbana en los índices de pobreza de la ciudad de Bogotá D.C.? Esta interrogante permite explorar cómo el crecimiento urbano afecta la distribución de la pobreza, abriendo la puerta a un análisis profundo de las implicaciones sociales y económicas en la ciudad.

Para abordar esta cuestión central, se plantean las siguientes preguntas de sistematización:

- ¿Qué datos geoespaciales y socioeconómicos son necesarios recopilar y cómo deben estructurarse en una base de datos para analizar la influencia de la expansión urbana en los índices de pobreza de la ciudad de Bogotá D.C.?
- ¿Qué tipo de modelo de ciencia de datos es el más adecuado para analizar las dinámicas de expansión urbana en la ciudad de Bogotá D.C. durante la última década?
- ¿Qué tan efectivo y preciso es el uso de modelos estadísticos y de ciencia de datos para identificar cómo la expansión urbana se ha relacionado con los índices de pobreza durante la última década?
- ¿Qué tipo de interacción se observa entre la expansión urbana y la distribución de la pobreza en Bogotá D.C.?

2. Objetivos del proyecto

2.1. Objetivo general

Analizar la influencia de la expansión urbana en los índices de pobreza de Bogotá D.C., integrando información geoespacial y datos socioeconómicos para evaluar su distribución y sus dinámicas a través del uso de herramientas de ciencia de datos

2.2. Objetivos específicos

- Construir una base de datos que consolide imágenes satelitales e indicadores socioeconómicos de pobreza en Bogotá D.C. sobre la última década.
- Identificar las dinámicas de expansión urbana en la ciudad de Bogotá D.C. en la última década haciendo uso de técnicas de aprendizaje de máquina y metodologías de procesamiento de imágenes geoespaciales.
- Evaluar el modelo escogido para el monitoreo de la expansión urbana, con el fin de integrar los datos socioeconómicos y caracterizar la relación entre la expansión urbana y los niveles de pobreza en la ciudad de Bogotá D.C.
- Desarrollar visualizaciones interactivas que permitan interpretar de manera clara la relación entre la expansión urbana y la distribución de la pobreza en Bogotá D.C.

3. Marco Teórico

A continuación, se presentan los conceptos base para poder entender la problemática a trabajar en este proyecto.

3.1. Expansión Urbana

Es un proceso de ordenamiento espacio temporal impulsado por factores tanto humanos como naturales. Entre los factores humanos destacan el aumento poblacional, el desarrollo económico y la infraestructura de transporte urbano, mientras que los factores naturales incluyen la topografía y las características del suelo. En conjunto, estos elementos transforman las estructuras del paisaje, afectando el uso y la cobertura del suelo y generando desafíos sociales, ambientales y económicos para las ciudades [3]. Astorkiza y Ferrero también coinciden con ello, al definir la expansión como una síntesis de elementos urbanos y rurales, que determinan un proceso de ocupación del espacio, acompañado de un crecimiento demográfico, transitorio o permanente, donde la expansión residencial es causa y consecuencia. Expansión que muchas veces ha sido más bien resultado de una demanda urgente, que de un proceso razonado y previsto de descentralización, por lo que no siempre se ha logrado responder con las infraestructuras necesarias al ritmo requerido [4]. En la actualidad, para medir y analizar la expansión urbana se utilizan diversas tecnologías, entre las que se destacan el procesamiento remoto de imágenes satelitales y modelos de monitoreo.

3.2. Métodos de medición de la expansión urbana

La expansión urbana se puede medir de varias maneras, utilizando métodos que abarcan enfoques demográficos, económicos, cartográficos y tecnológicos. Entre estos, destacan las tasas de crecimiento de la población, los cambios en el uso del suelo, el análisis de mapas históricos y censos, así como la creación de indicadores de densidad o dispersión. En los últimos años, el uso de imágenes satelitales, fotografías aéreas y Sistemas de Información Geográfica (SIG) ha proporcionado una visión más detallada y precisa de los procesos de urbanización. Dentro de estos enfoques, los métodos de medición espacial juegan un papel fundamental, ya que permiten identificar patrones de expansión, concentración o fragmentación del territorio urbano.

3.3. Espacial

El método de medición del espacio urbano se basa en herramientas de análisis geográfico, como la teledetección y los Sistemas de Información Geográfica (SIG). A través de imágenes satelitales, fotografías aéreas y técnicas de detección de cambios a lo largo del tiempo, se puede ver cómo se transforma el territorio, localizando áreas que están creciendo, densificándose o fragmentándose.

3.3.1. Imágenes satelitales

Son capturas de la superficie terrestre obtenidas desde satélites equipados con sensores remotos. Estas son utilizadas ampliamente en diversas áreas, como la agricultura, el monitoreo ambiental, o la planificación urbana, debido a su capacidad para proporcionar información detallada y en tiempo real de grandes extensiones de terreno. Sin embargo, antes de analizar las imágenes, es crucial realizar un preprocesamiento que corrija las distorsiones causadas por factores como la atmósfera, la topografía y la posición del satélite [5].

Las imágenes de sensores remotos, disponibles gratuitamente en fuentes como Landsat, Sentinel, USGS y NASA Earthdata, capturan energía electromagnética reflejada o emitida desde la superficie terrestre en varias longitudes de onda. A diferencia de las fotografías tradicionales, que solo registran el espectro visible, estas imágenes también incluyen el infrarrojo y otras bandas espectrales, permitiendo analizar fenómenos invisibles al ojo humano, como la salud de la vegetación o la humedad del suelo. Los sensores remotos emplean detectores electrónicos avanzados para generar estas imágenes [6].

3.3.2. Percepción Remota

Según el Servicio Geológico de los Estados Unidos (USGS), la percepción remota consiste en la detección y monitoreo de las características físicas de un área mediante la medición de la radiación que esta emite o refleja, realizada a cierta distancia, generalmente desde satélites o aeronaves [7]. Se forma a partir de 7 componentes. Una fuente de iluminación, que proporciona la energía electromagnética al área de estudio. La atmósfera, que constituye el medio a través del cual esta energía se desplaza hacia el objeto y regresa al sensor. La interacción con el objeto, que depende de sus propiedades y de la radiación incidente. Un sensor remoto, que registra los niveles de energía y capta sus valores electromagnéticos. Los datos finales, que son transmitidos, recibidos y procesados para transformarse en una imagen interpretable, mediante la cual se lleva el análisis e interpretación para extraer información relevante del área estudiada [8].

Para desarrollar dicho proceso existen distintos tipos de satélites capaces de ello. Uno de ellos es el satélite con órbita helio sincrónica, también conocido como de órbita solar, el cual orbita la tierra pasando por los polos, aproximadamente por el mismo lugar de la tierra a la misma hora cada día. Este permite obtener imágenes con condiciones de iluminación similares, los cuales son útiles para analizar cambios en la superficie terrestre. Un ejemplo es el satélite Landsat, el cual se utilizó para el análisis del presente proyecto.

3.3.3. Landsat

El programa estadounidense Landsat es un programa conjunto entre la NASA y el Servicio Geológico de los Estados Unidos (USGS) [9]. La idea de un satélite para la observación de recursos terrestres surgió en la década de 1960, inspirada en las imágenes de la Tierra obtenidas durante las misiones Apolo. En 1965, el director del USGS, William Pecora, propuso la idea de un programa de satélites de teledetección para recopilar datos sobre los recursos naturales de nuestro planeta [10].

Entre los eventos más relevantes de la historia del satélite, se encuentra que Landsat 1 se lanzó en 1972. Posteriormente, en 1978 se lanzó la versión 3, que incorporaba mejoras en los sensores para aumentar la calidad de las imágenes, con una resolución terrestre de 38 metros y el uso de cámaras con generación de imágenes de banda espectral amplia [11] [11]. La versión 4, que se lanzó en 1982, introdujo el sensor Thematic Mapper (TM), con mayor resolución y más bandas espectrales. Landsat 6 se intentó poner en funcionamiento en 1993, pero no logró alcanzar la órbita debido a una falla en el lanzamiento. Más adelante, Landsat 7 se puso en órbita en 1999, equipado con el Enhanced Thematic Mapper Plus (ETM+) y estuvo activo hasta 2024 [12].

Más tarde, Landsat 8 se lanzó el 11 de febrero de 2013, desde la base aérea de Vandenberg, California. La carga del satélite consta de dos instrumentos científicos: el generador de imágenes terrestres operacionales (OLI) y el sensor infrarrojo térmico (TIRS). OLI es un sensor de barrido transversal con un telescopio de 4 espejos y cuantificación de 12 bits. Recopila datos para las bandas espectrales visibles, infrarroja cercana e infrarroja de onda corta, así como para una banda pancromática, teniendo una vida útil de 5 años [13].

OLI provee dos bandas espectrales, una diseñada específicamente para la detección de cirros y la otra para observaciones en zonas costeras. El sensor TIRS proporciona datos para dos bandas espectrales más estrechas en la región térmica, que permiten una mejor resolución y detección de temperaturas en comparación con la banda térmica ancha de los satélites Landsat 4–7. Estos sensores proporcionan cobertura con una resolución espacial de 30 metros (visible, NIR, SWIR), 100 metros (térmico); y 15 metros (pancromático). Este satélite se

desarrolló conjuntamente entre la NASA y el USGS. La NASA dirigió las fases de diseño, construcción, lanzamiento y calibración, y en el 2013, la USGS asume las operaciones de funcionamiento [13].

En la Tabla 3.1 se detallan las características de los sensores OLI y TIRS, incluyendo su resolución, número de la banda, nombre y algunas aplicaciones.

Tabla 3.1: Características de los sensores OLI y TIRS de Landsat-8
 (Adaptado de [14])

Banda	Nombre	Resolución (m)	Aplicaciones
Banda 1	Coastal/Aerosol	30	Estudios costeros y de aerosol.
Banda 2	Azul	30	Distingue el suelo de la vegetación y la vegetación caducifolia de la conífera.
Banda 3	Verde	30	Enfatiza la vegetación en su punto máximo, lo cual es útil para evaluar el vigor de las plantas.
Banda 4	Rojo	30	Discrimina las pendientes de la vegetación.
Banda 5	Infrarrojo cercano (NIR)	30	Enfatiza el contenido de biomasa y las líneas de costa.
Banda 6	Infrarrojo de onda corta (SWIR) 1	30	Discrimina el contenido de humedad del suelo y la vegetación.
Banda 7	Infrarrojo de onda corta (SWIR) 2	30	Mejora el contenido de humedad del suelo y la vegetación y la penetración de nubes delgadas.
Banda 8	Pancromática	15	Definición de imagen más nítida.
Banda 9	Cirros	30	Detección mejorada de la contaminación por cirros.
Banda 10	Infrarrojo de onda larga (SWIR) 1	100*	Mapeo térmico y estimación de la humedad del suelo.
Banda 11	Infrarrojo de onda larga (SWIR) 2	100*	Mapeo térmico mejorado y estimación de la humedad del suelo.

* Las bandas TIRS tienen una resolución de 100 m, pero se remuestrean a 30 m en el producto de datos entregado.

3.3.4. Otros satélites

Sentinel-2

Lanzada por la Agencia Espacial Europea, entrega imágenes ópticas de alta resolución (10 metros en sus bandas principales) y presenta una frecuencia de revisita mejorada debido a la operación conjunta de los satélites 2A y 2B. Desafortunadamente, su recopilación de datos solo comenzó en junio de 2015, dejando un vacío para el año 2014 y restringiendo su aplicación para investigaciones que requieran una cronología anterior a su año de inicio [15].

MODIS

Proporciona series históricas continuas desde principios de los años 2000 y ofrece una cobertura casi diaria, lo que lo convierte en una herramienta valiosa para el monitoreo de fenómenos a nivel regional o continental. Sin embargo, en el caso de Bogotá D.C. donde barrios, equipamientos, zonas verdes y áreas en transformación se distribuyen en espacios menores a 250 metros, su resolución espacial mínima no es suficiente. Un solo píxel de MODIS puede abarcar múltiples usos del suelo, lo que diluye las diferencias reales en la superficie y limita la capacidad de identificar procesos urbanos sutiles, como la expansión en las periferias, la densificación o la pérdida de cobertura vegetal a nivel local [16].

3.3.5. Técnicas de preprocesamiento de imágenes

Se usan para corregir distorsiones y errores presentes en los datos originales, con el fin de obtener representaciones más precisas y consistentes de la superficie terrestre. Esto permite realizar análisis comparativos, cuantitativos y espaciales confiables, asegurando que las variaciones observadas en las imágenes se deban a cambios reales en el terreno y no a factores externos como la atmósfera, el ángulo de adquisición o las características del sensor [17].

Entre las principales técnicas de preprocesamiento se destacan las siguientes [18]:

- Correcciones geométricas: Corrigen distorsiones causadas por la geometría terrestre, el movimiento del sensor y la altitud de la plataforma. Factores como la curvatura de la Tierra, los cambios en la altitud y el movimiento del satélite generan errores en la posición de los píxeles. Estas correcciones permiten alinear las imágenes con coordenadas geográficas reales, lo que es esencial para el análisis espacial en estudios multitemporales.
- Correcciones radiométricas: Ajustan variaciones en la energía reflejada que llega al sensor, influenciadas por el ángulo solar, la atmósfera y el tipo de terreno, asegurando que las imágenes reflejen fielmente la reflectancia real de la superficie.

- Correcciones atmosféricas: Eliminan las distorsiones causadas por la absorción y dispersión de la radiación solar al atravesar la atmósfera, mejorando la representación de las características de la superficie terrestre.

3.3.6. Índices espectrales

Es una fórmula que une los valores de píxeles de dos o más bandas espectrales en una imagen multiespectral a través de diferentes algoritmos, que se enfocan sobre todo en la relación entre las bandas o en el escalado de características (como algoritmos normalizados o estandarizados). Se calculan con el fin de poner en relieve los píxeles de una imagen que, además de mostrar la abundancia relativa de una cobertura terrestre en particular, destacan también una función del ecosistema [19].

Uno de los índices es el Índice de Vegetación de Diferencia Normalizada (NDVI), que mide la salud y la densidad de la vegetación a partir de la reflectancia en las bandas del rojo y el infrarrojo cercano [20]. El NDVI se calcula mediante la siguiente expresión:

$$\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}} \quad (1)$$

Por otra parte, el Índice de Agua de Diferencia Normalizada (NDWI), que permite identificar la presencia de agua en la superficie, utiliza la radiación infrarroja cercana reflejada y la luz verde visible para realzar la presencia de dichas características, a la vez que elimina la presencia de características del suelo y la vegetación terrestre [21]. Se expresa de la siguiente manera:

$$\text{NDWI} = \frac{\text{Green} - \text{NIR}}{\text{Green} + \text{NIR}} \quad (2)$$

Finalmente, el Índice de Área Construida de Diferencia Normalizada (NDBI) es una herramienta clave para detectar áreas urbanas y superficies construidas, ya que se basa en la reflectancia en las bandas del infrarrojo cercano y el infrarrojo de onda corta [20] [20]. El cálculo se realiza de la siguiente manera:

$$\text{NDBI} = \frac{\text{SWIR} - \text{NIR}}{\text{SWIR} + \text{NIR}} \quad (3)$$

Estos índices no solo simplifican la clasificación de la cobertura del suelo, sino que también proporcionan métricas cuantitativas que pueden ser utilizadas para el análisis temporal de cambios en el territorio. Por ejemplo, el monitoreo de la deforestación, la expansión urbana

o la degradación de suelos se basa en gran medida en la comparación de estos índices a lo largo del tiempo.

3.3.7. Detección de cambios de imágenes

La detección de cambios espaciales mediante imágenes satelitales multitemporales es esencial para analizar la expansión urbana, ya que permite observar y cuantificar transformaciones en el uso del suelo y la estructura de la ciudad a lo largo del tiempo. Al comparar imágenes de diferentes momentos, es posible identificar patrones de urbanización, como la extensión de áreas residenciales, comerciales e industriales sobre terrenos previamente no urbanizados

3.3.8. Métodos de detección de cambios en imágenes satelitales

Para abordar la detección de cambios se pueden utilizar distintos enfoques, dentro de los cuales se destacan [22]:

- Métodos basados en álgebra: Se basan en la aplicación de operaciones matemáticas entre los valores de píxeles de una misma área, mediante técnicas como la diferenciación de imágenes y el Análisis del Vector de Cambio.
- Métodos basados en transformaciones: Se transforman las imágenes a un nuevo espacio que facilite la detección de cambios. Un método usado es el análisis de componentes principales para condensar la información en nuevas bandas.
- Métodos basados en clasificación: Estos métodos involucran procesos de post-clasificación de imágenes, donde primero se clasifican las imágenes y luego se comparan para identificar cambios.
- Métodos avanzados: Usan modelos complejos y Sistemas de Información Geográfica (SIG) que integran datos de distintas fuentes, como, por ejemplo, el tipo de suelo y las redes viales.

3.3.9. Separabilidad espectral entre clases

Antes de aplicar métodos de clasificación supervisada, es fundamental evaluar si las clases de interés son distinguibles en el espacio espectral. Para ello, se utilizan métricas de separabilidad, que cuantifican la distancia entre las distribuciones espectrales de las clases. Entre las más empleadas se encuentran las siguientes:

Distancia Euclidiana

Mide la separación directa entre los centroides espectrales de las clases, sin considerar la varianza ni la covarianza entre las bandas. Es una métrica sencilla e intuitiva, aunque puede resultar limitada cuando las distribuciones espectrales no presentan forma esférica [23]. Se expresa de la siguiente manera:

$$d_E(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

Donde:

- x, y son vectores espectrales representativos de cada clase.
- n es el número de bandas consideradas.

Distancia de Mahalanobis

Incorpora la dispersión y correlación entre las variables espectrales mediante la matriz de covarianza, lo que la hace más robusta frente a clases con diferentes formas o varianzas. A diferencia de la Euclidiana, esta métrica ajusta la distancia en función de la estructura estadística de los datos [24]. Se define de la siguiente manera:

$$d_M(x, y) = \sqrt{(x - y)^T S^{-1} (x - y)} \quad (5)$$

Donde:

- S es la matriz de covarianza de las clases.
- $(x - y)^T$ representa el vector transpuesto de diferencias espectrales entre los centroides de las clases.

Distancia de Jeffries–Matusita

Es una métrica derivada de la divergencia de Bhattacharyya, utilizada para evaluar la separabilidad espectral entre dos clases. Sus valores están acotados entre 0 y 2, donde los valores cercanos a 2 indican una alta separabilidad y una mínima superposición entre las distribuciones espectrales [25]. Se define como:

$$J_{ij} = 2(1 - e^{-B_{ij}}) \quad (6)$$

Donde B_{ij} corresponde a la divergencia de Bhattacharyya, definida como:

$$B_{ij} = \frac{1}{8}(\mu_i - \mu_j)^T \left(\frac{\Sigma_i + \Sigma_j}{2} \right)^{-1} (\mu_i - \mu_j) + \frac{1}{2} \ln \left(\frac{|\Sigma_i + \Sigma_j|}{\sqrt{|\Sigma_i| |\Sigma_j|}} \right) \quad (7)$$

Donde:

- μ_i y μ_j son los vectores de medias espectrales de las clases i y j .
- S_i y S_j son las matrices de covarianza correspondientes.
- $|\cdot|$ indica el determinante de la matriz.

3.4. Métodos de clasificación

Los métodos de clasificación se dividen en dos grandes categorías: supervisados y no supervisados. Los métodos supervisados requieren un conjunto de datos de entrenamiento previamente etiquetado, es decir, ejemplos conocidos de cada clase (por ejemplo, cobertura urbana, vegetación, agua, entre otros). En cambio, los métodos no supervisados no dependen de información previa sobre las clases, sino que identifican patrones inherentes en los datos, como agrupamientos naturales o umbrales espectrales, para realizar la clasificación.

3.4.1. Clasificación mediante métodos de umbralización (no supervisados)

Los métodos de clasificación basados en umbrales constituyen una estrategia no supervisada ampliamente utilizada en la clasificación de la cobertura del suelo, ya que permiten segmentar o separar distintas categorías del terreno mediante la identificación de valores óptimos en los histogramas de reflectancia o de índices espectrales. Estos métodos establecen un valor umbral T que divide el conjunto de píxeles en dos o más clases sin requerir entrenamiento previo ni información auxiliar [26].

Este proceso consiste en particionar una imagen directamente en regiones basadas en su intensidad mediante la selección de umbral que separe los modos dominantes del histograma [26] Si $f(x, y)$ representa el valor de intensidad o un valor espectral de la imagen segmentada se define como:

$$g(x, y) = \begin{cases} 1, & \text{si } f(x, y) > T \\ 0, & \text{si } f(x, y) \leq T \end{cases} \quad (8)$$

Método de Otsu

El método de Otsu [27] selecciona el umbral T^* que maximiza la varianza entre clases o de manera equivalente minimiza la varianza dentro de las clases $\sigma_b^2(T)$, asumiendo que el histograma de una imagen puede representarse mediante $\omega_0(T)$, $\omega_1(T)$, y medias $\mu_0(T)$, $\mu_1(T)$. La varianza entre clases se define de la siguiente manera:

$$\sigma_b^2(T) = \omega_0(T)\omega_1(T)[\mu_0(T) - \mu_1(T)]^2 \quad (9)$$

Y el umbral óptimo se obtiene mediante:

$$\sigma_B^2(T^*) = \max_{1 \leq T < L} \sigma_B^2(T) \quad (10)$$

Método de Li y Lee

El método de Li y Lee [28] se basa en la minimización de la entropía cruzada entre la imagen original y la imagen umbralizada, buscando un valor T^* , haciéndolo más robusto frente a distribuciones espectrales complejas.

La entropía cruzada $\eta(T)$ se calcula para cada posible umbral T usando el histograma de la imagen (h_j):

$$\eta(T) = \sum_{j=1}^{T-1} j h_j \log \left(\frac{j}{\mu_1(T)} \right) + \sum_{j=T}^L j h_j \log \left(\frac{j}{\mu_2(T)} \right) \quad (11)$$

El umbral óptimo T^* se selecciona como el valor T que minimiza la función de entropía cruzada:

$$T^* = \min_{1 \leq T < L} [\eta(T)] \quad (12)$$

Método de Yen, Chang y Chang

Este método se basa en una función de costo $C(k)$ que busca determinar de manera automática el número de clases k y sus umbrales asociados, buscando equilibrar la discrepancia ($Dis(k)$) entre la imagen original y la segmentada, y el número de bits requeridos para representarla [29].

El número óptimo de clases K^* es aquel que minimiza la función de costo:

$$k^* = \min_{1 \leq k < L} [C(k)] \quad (13)$$

Mientras que la función de costo se define de la siguiente manera:

$$C(k) = \rho \sqrt{Dis(k)} + (\log_2 k)^2 \quad (14)$$

3.4.2. Clasificación mediante métodos supervisados

Máquinas de Vector de Soporte (SVM)

Las Máquinas de Vectores de Soporte (SVM por sus siglas en inglés), propuestas por Vapnik [30], son un método supervisado de clasificación binaria. Su principio se basa en encontrar una línea o hiperplano que separe las clases en el espacio de características, maximizando la distancia (margen) entre dichas clases y el hiperplano. Dado un conjunto de entrenamiento $\{(x_i, y_i)\}_{i=1}^n$, donde $x_i \in R^d$ representa el vector de características y $y_i \in \{-1, +1\}$ la etiqueta de clase, el problema de optimización puede expresarse de la siguiente manera:

$$\min_{\{w, b, \xi\}} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (15)$$

Sujeto a

$$(y_i(w^\top \phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i) \quad (16)$$

Donde w representa el vector normal al hiperplano, b el término de sesgo, $C > 0$ un parámetro que controla la penalización de errores en clasificación, ξ_i las variables de holgura y $\phi(\cdot)$ una función de mapeo a un espacio de mayor dimensionalidad.

Para llevar a cabo este mapeo, se utilizan funciones kernel, como el kernel de base radial (RBF), que permiten separar de manera efectiva datos no linealmente separables, como los provenientes de imágenes satelitales multiespectrales. Sin embargo, el rendimiento de la SVM puede verse afectado cuando la dimensionalidad del espacio de características es muy alta en comparación con el número de muestras disponibles, aspecto conocido como la “maldición de la dimensionalidad”. La utilización de kernels no lineales, especialmente el RBF, proyecta los datos a un espacio donde las clases son más fácilmente separables, manteniendo un equilibrio entre precisión y complejidad computacional [31].

Clasificación basada en Random Forest

Los árboles de decisión son un procedimiento de clasificación que divide recursivamente un

conjunto de datos en subconjuntos más pequeños mediante decisiones basadas en pruebas definidas en cada nodo. Estos árboles consisten en un nodo raíz (que abarca todos los datos), nodos internos (donde se realizan las divisiones) y nodos terminales (u hojas) que contienen las clases finales [32].

Una técnica derivada y ampliamente utilizada de los árboles de decisión es el algoritmo Random Forest, que se basa en la construcción de múltiples árboles independientes a partir de subconjuntos aleatorios del conjunto de entrenamiento y de las variables disponibles. Cada árbol del bosque genera una predicción individual, y la clase final se determina mediante un esquema de votación mayoritaria entre todos los árboles. Esta estrategia permite reducir el sobreajuste que puede presentarse en árboles individuales y mejora la precisión general del modelo. Este procedimiento se puede representar de la siguiente manera:

$$\hat{y} = \text{mode}\{f_1(x), f_2(x), \dots, f_T(x)\} \quad (17)$$

Donde $f_t(x)$ es la predicción del árbol t para la entrada x , mientras que T es el número total de árboles y \hat{y} representa la clase final obtenida por votación mayoritaria.

Además de su precisión, una de sus principales ventajas es la capacidad de proporcionar medidas sobre la importancia de las variables, lo cual facilita la interpretación del modelo y la identificación de las características más relevantes para la clasificación [33].

Clasificación mediante regresión Logística

La regresión logística es un método estadístico de clasificación que modela la probabilidad de que una observación pertenezca a una clase determinada en función de un conjunto de variables predictoras. El modelo se basa en la función sigmoide, que relaciona las variables predictoras $x_i \in R^d$ con la probabilidad de que la observación pertenezca a la clase $y_i \in \{0, 1\}$.

Esta probabilidad se expresa como:

$$P(y_i = 1 | x_i) = \frac{1}{1 + e^{-(w^\top x_i + b)}} \quad (18)$$

Donde w es el vector de coeficientes que pondera cada variable predictora, b es el término de sesgo o intercepto, y el objetivo es encontrar los parámetros que maximicen la verosimilitud del modelo para los datos de entrenamiento. La función de verosimilitud permite estimar los parámetros del modelo de manera que las probabilidades predichas se ajusten lo mejor posible a los datos de entrenamiento.

$$\mathcal{L}(w, b) = \prod_{i=1}^n P(y_i | x_i) \quad (19)$$

Una de las principales ventajas de la regresión logística es que produce resultados fácilmente interpretables y genera probabilidades como salida, que es útil en escenarios donde estos aspectos son relevantes. Además, al no depender de supuestos estrictos sobre la distribución de los datos, ofrece mayor flexibilidad frente a otros métodos paramétricos. Sin embargo, su desempeño puede verse limitado cuando las relaciones entre las variables predictoras y las clases no son lineales o cuando es necesario capturar patrones espaciales complejos [34].

Clasificación mediante Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) es un algoritmo de ensemble basado en árboles de decisión que aplica la técnica de gradient boosting, en la que cada nuevo árbol se construye para corregir los errores de los árboles anteriores. A diferencia de un árbol individual, XGBoost incluye mecanismos de regularización que contribuyen a reducir el sobreajuste y mejorar la capacidad de generalización del modelo [35].

El objetivo de XGboost es minimizar la función de pérdida regularizada:

$$L(\phi) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (20)$$

Donde $l(y_i, \hat{y}_i)$ es la función de pérdida entre la predicción \hat{y}_i y el valor real y_i , K es el número de árboles, f_k representa cada árbol, y $\Omega(f_k)$ es un término de regularización que penaliza la complejidad de los árboles.

Además, puede procesar grandes volúmenes de datos y manejar clases desbalanceadas, al tiempo que presenta medidas de importancia de las variables, lo que facilita la interpretación de los resultados. Su principal limitación es la necesidad de un ajuste de hiperparámetros, lo cual puede aumentar los costos computacionales [35].

3.4.3. Métricas de Evaluación

En el contexto de la detección de áreas urbanas mediante modelos supervisados, la evaluación del desempeño es fundamental para garantizar la precisión y confiabilidad de los resultados. Entre las métricas más utilizadas en problemas de clasificación se encuentran la exactitud, la precisión, la sensibilidad o recall y la medida F1, las cuales son ampliamente empleadas para evaluar el desempeño de modelos supervisados, especialmente en escenarios con desbalance entre clases [36].

La exactitud representa la proporción de predicciones correctas respecto al total de observaciones y se define como:

$$Accuracy = \frac{VP + VN}{VP + VN + FP + FN} \quad (21)$$

Donde VP corresponde a los verdaderos positivos, VN a los verdaderos negativos, FP a los falsos positivos y FN a los falsos negativos.

La precisión mide la proporción de instancias correctamente clasificadas como positivas entre todas las predicciones positivas realizadas por el modelo, mientras que la sensibilidad o recall evalúa la capacidad del modelo para identificar correctamente las instancias positivas. Estas métricas se expresan como:

$$Precision = \frac{VP}{VP + FP} \quad (22)$$

$$Recall = \frac{VP}{VP + FN} \quad (23)$$

Para integrar ambas métricas en una única medida, se utiliza el puntaje F1, definido como el promedio armónico entre precisión y recall, el cual resulta especialmente útil cuando existe un desbalance significativo entre las clases:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (24)$$

Asimismo, al seleccionar un modelo supervisado, es importante considerar los tiempos de cómputo asociados a su entrenamiento y ejecución, especialmente en análisis multitemporales o de gran extensión espacial. Un modelo con alto desempeño, pero elevado costo computacional puede resultar poco eficiente en la práctica, por lo que debe buscarse un equilibrio entre precisión, generalización y eficiencia computacional.

3.5. Demográfica

Es el estudio de las características y dinámicas cuantitativas de la población. En un entorno urbano, ayuda a comprender como se distribuyen y se transforman los grupos humanos en el espacio, lo cual es para la planificación de servicios, infraestructura y políticas públicas en las ciudades [37].

3.5.1. Densidad urbana

Se refiere al número de personas que habitan o viven en un área que se considera urbana; particularmente, es un indicador específico del urbanismo que mide la concentración de personas que viven exclusivamente en un área urbanizada o construida. Este indicador, resulta importante puesto que permite identificar como se distribuye la población a medida que una ciudad crece [38].

3.5.2. Ordenamiento territorial

En Bogotá D.C., el territorio urbano se organiza en tres niveles de subdivisión espacial con fines de gestión y planeación. Las localidades constituyen divisiones administrativas de primer nivel, con autonomía política y estadística, pero su gran extensión y diversidad interna, que combina zonas consolidadas con sectores en expansión, limitan su utilidad para análisis detallados. Las Unidades de Planeación Local (UPL) agrupan varias zonas con cierta afinidad funcional y sirven como base para planes de desarrollo intermedios, aunque su escala sigue siendo amplia y tiende a diluir dinámicas locales, especialmente en áreas de crecimiento reciente. Por su parte, las Unidades de Planeación Zonal (UPZ) se definen como subdivisiones más finas, basadas en criterios de cohesión social, histórica, funcional y morfológica, lo que las hace representativas de realidades urbanas más homogéneas.

3.5.3. Migración

La migración se refiere al movimiento de personas entre diferentes territorios, ya sea de manera temporal o permanente. En el contexto del crecimiento urbano, la migración, especialmente del campo a la ciudad o entre regiones, se convierte en uno de los principales motores de la expansión urbana, ya que aumenta la demanda de vivienda y servicios en las ciudades que reciben a estos migrantes. Cuando esta presión demográfica no se gestiona en áreas urbanas ya establecidas, a menudo resulta en la ocupación de terrenos periféricos, lo que contribuye a un crecimiento horizontal del tejido urbano. Así que, la migración no es solo un fenómeno demográfico, sino también un factor espacial fundamental en la forma y expansión de las ciudades [39].

3.5.4. Gentrificación

Hace referencia a procesos de renovación urbana de antiguas zonas ubicadas en el centro de la ciudad, producto de la llegada de personas de estratos medios y altos, y el desplazamiento

de las personas pobres que habitaban en estos lugares. Dicho proceso implica dinámicas de extracción de renta de suelo, y cómo se distribuyen las ganancias obtenidas en ello [40]. La gentrificación no solo afecta a la clase trabajadora, inmigrantes y personas sin hogar, sino que también termina afectando a pequeños comerciantes, artistas y profesionales de clase media, que no logran resistirse a los aumentos en los costos de vida de las zonas centrales de la ciudad [41].

3.6. Pobreza

La pobreza es un fenómeno multidimensional que no se limita a la falta de ingresos económicos, sino que también incluye la carencia de acceso a servicios básicos, educación, salud y oportunidades de desarrollo lo cual niega la posibilidad de elevar el nivel de vida [42]. El análisis de la cobertura del suelo y su relación con indicadores socioeconómicos permite identificar áreas donde la pobreza está asociada a factores como la degradación ambiental, la falta de infraestructura o la escasez de recursos naturales [2]. Por ejemplo, regiones con alta degradación de suelos o pérdida de vegetación suelen presentar mayores niveles de pobreza debido a la disminución de la productividad agrícola y la pérdida de medios de subsistencia.

3.6.1. Métodos de medición de la pobreza

En Colombia la pobreza se mide a partir de dos enfoques principales que se complementan: el enfoque monetario y el enfoque multidimensional.

3.6.2. Pobreza monetaria

Por un lado, la pobreza monetaria considera pobres a los hogares cuyo ingreso corriente per cápita de la unidad de gasto (y_i) está por debajo de un umbral denominado línea de pobreza (z). Y por otro lado se distingue la pobreza extrema, que se determina cuando el ingreso per cápita está por debajo de la línea de pobreza extrema (z_e) [43]. De acuerdo con la dirección de metodología y producción estadística, los pasos para construir las líneas de pobreza (LP) y pobreza extrema (LI) son: la construcción del gasto corriente per cápita a nivel de la unidad de gasto, la construcción del Deflactor Espacial de Precios (DEP), el ordenamiento de los hogares por percentil de gasto corriente per cápita deflactado, la selección de una población de referencia a priori, la construcción de la canasta básica de alimentos, el ajuste de cantidades de la canasta básica de alimentos para alcanzar los requerimientos calóricos, la valoración de la canasta de alimentos ajustada: LI, el paso de la LI a la LP a partir del coeficiente de Orshansky y la estimación de la incidencia de la pobreza [44].

3.6.3. Índice de pobreza multidimensional

Propuesto por Alkire y Foster [45], presenta una metodología alternativa a la medición tradicional basada únicamente en los ingresos de una persona. Esta se basa en el enfoque de “dual cutoff”, que evalúa la pobreza mediante la identificación de múltiples carencias simultáneas en diferentes dimensiones, como la educación, la salud, las condiciones de vivienda y el acceso a servicios básicos. Primero, se determina si una persona es privada en cada dimensión según los umbrales establecidos y después se define si es considerada pobre dependiendo del número total de privaciones que enfrenta.

En Colombia, la metodología oficial para el cálculo de este indicador se encuentra en el documento CONPES 150 de 2012. Se debe tener en cuenta que la unidad de análisis es el hogar, por lo cual todas las personas que lo conforman experimentan simultáneamente dichas privaciones en lugar de realizarse de manera individual [46]. A partir de ello, se considera que un hogar se encuentra en condición de pobreza si tiene privaciones de por lo menos el 33 % de los indicadores seleccionados, los cuales se muestran en la Tabla 3.2.

Tabla 3.2: Dimensiones y variables del IPM para Colombia (Adaptado de [46]).

Dimensión	Indicador	Ponderación
Condiciones educativas del hogar	Bajo logro educativo	0.1
	Analfabetismo	0.1
Condiciones de la niñez y juventud	Inasistencia escolar	0.05
	Rezago escolar	0.05
	Barreras de acceso a servicios para el cuidado de la primera infancia	0.05
	Trabajo infantil	0.05
Trabajo	Desempleo de larga duración	0.1
	Empleo informal	0.1
Servicios públicos domiciliarios y condiciones de la vivienda	Sin acceso a fuente de agua mejorada	0.04
	Inadecuada eliminación de excretas	0.04
	Pisos inadecuados	0.04
	Paredes exteriores inadecuadas	0.04
	Hacinamiento crítico	0.04

Luego de evaluar la nota metodológica de la Encuesta Multipropósito [47], se identificó que existen componentes del IPM que no se calculan a partir de las respuestas de la encuesta, sino que se extraen de otras fuentes externas. Por tanto, se evidencia la necesidad de usar

una técnica o modelo que permita estimar la diferencia, al momento de calcular el IPM para cada hogar y, posteriormente, el porcentaje de hogares en condición de pobreza. Para ello, se decide usar el MQE.

3.6.4. Estimación por coincidencia de cuantiles (Quantile Matching) (MQE)

Es un método estadístico que busca ajustar los parámetros de un modelo de forma que la función de cuantiles predicha por el modelo se alinee con los cuantiles observados empíricamente en los datos. A pesar de que comparte algunas similitudes con la estimación por mínimos cuadrados para el uso de modelos de regresión, la diferencia fundamental es que el MQE se usa para coincidir funciones de distribución, ya que admite expresiones explícitas. De acuerdo con Sgouropoulos, Yao y Yastremiz [48], Sea Y una variable aleatoria y $X = (X_1, X_2, \dots, X_p)$ una colección de p variables aleatorias, se busca encontrar una combinación lineal,

$$\beta'X = \beta_1X_1 + \dots + \beta_pX_p \quad (25)$$

tal que su distribución coincida con la distribución de Y . Se propone buscar un vector β tal que se minimice la diferencia cuadrática integrada entre las funciones de cuantiles:

$$\int_0^1 \{Q_Y(\alpha) - Q_{\beta'X}(\alpha)\}^2 d\alpha \quad (26)$$

Donde $Q_\xi(\alpha)$ denota el cuantil α -ésimo de la variable aleatoria ξ , es decir:

$$P\{\xi \leq Q_\xi(\alpha)\} = \alpha, \quad \text{para } \alpha \in [0, 1] \quad (27)$$

3.6.5. Encuesta Multipropósito

La Encuesta Multipropósito (EM) es una herramienta complementaria al censo, diseñada para recopilar información específica sobre diversos aspectos socioeconómicos a nivel regional en el área metropolitana de la ciudad de Bogotá D.C. A diferencia del censo, que se realiza cada diez años, las encuestas multipropósito pueden llevarse a cabo con mayor frecuencia, lo que permite obtener datos actualizados y más detallados. Está estructurada en 13 capítulos que abarcan aspectos fundamentales de las condiciones de vida en la ciudad de Bogotá D.C y algunos municipios cercanos. Incluye información sobre la identificación y ubicación de las viviendas, las características constructivas y los riesgos ambientales, las condiciones de tenencia y financiación de la vivienda, el acceso a servicios públicos y tecnologías de la

información, la composición demográfica y las dinámicas migratorias de los hogares, así como la situación de salud con énfasis en la seguridad social. Todo esto proporciona una visión integral que permite orientar el diseño de políticas públicas enfocadas en el mejoramiento del bienestar social [49].

3.7. Relación entre expansión urbana con la desigualdad

De acuerdo con Wei y Ewing [50], a medida que avanza la urbanización en las ciudades, la expansión y dispersión urbana se han identificado como un factor significativo en el aumento de la desigualdad económica, tanto en países desarrollados como en vías de desarrollo. Este crecimiento, junto con la concentración de pobreza en áreas urbanas, restringe aún más las oportunidades equitativas y amplían la desigualdad espacial. En 1968 se planteó la hipótesis del desajuste espacial, que sostenía que una mayor distancia de los centros laborales a los centros de vivienda puede incrementar las tasas de desempleo y, en consecuencia, agravar los niveles de pobreza en una región. Wheeler [51] identificó que existe una fuerte correlación negativa entre la variación en la distribución de ingresos de los hogares en un área metropolitana y su densidad general, de manera que, a medida que la población metropolitana migra a otras zonas, el ingreso promedio de los hogares disminuye en esa misma área. Wassmer y Baass, citados por Wei y Ewing [50], identificaron que metrópolis más centralizadas demostraron una menor mediana para los precios de las viviendas y una menor proporción de viviendas en la categoría de precios altos que en regiones dispersas. Y en otros estudios se reportaron ahorros en costos de infraestructura a medida que la densidad urbana aumenta, donde se destaca que precios más bajos en las viviendas en áreas dispersas se compensan con mayores costos de transporte, por lo que áreas más compactas parecían mostrar una ligera ventaja. Por último, de acuerdo con Zhao [52], en países en vías de desarrollo, la concentración de los servicios públicos y las infraestructuras de transporte en las zonas urbanas centrales resultan en una aglomeración de personas de altos ingresos en el centro de la urbe, mientras que las personas de bajos ingresos y los migrantes son confinados a la periferia.

3.8. Técnicas para relacionar la expansión urbana y la pobreza

Comprender cómo el crecimiento urbano influye en los niveles de pobreza, o viceversa, requiere combinar enfoques espaciales, económicos y sociales, capaces de capturar la complejidad de las interacciones entre desarrollo urbano, desigualdad y exclusión. En este contexto diversos métodos han sido desarrollados para medir dicha correlación, que van desde modelos econométricos y espaciales tradicionales hasta enfoques basados en imágenes satelitales y aprendizaje automático [4]. A continuación, se profundizará en los métodos de correlación y algunas herramientas estadísticas que permiten dicho análisis.

3.9. Métodos de correlación

El análisis de correlación examina la relación lineal entre dos variables, proporcionando una medida tanto de la intensidad como de la dirección de dicha relación, sin asumir causalidad entre ellas. Se considera únicamente relaciones lineales, de modo que variables con una relación no lineal fuerte pueden mostrar una correlación débil o nula. Para verificar los resultados, se recomienda el uso de diagramas de dispersión. La correlación se puede aplicar a variables tanto paramétricas como no paramétricas. El método paramétrico más común es la correlación de Pearson (producto-momento). En el caso de métodos no paramétricos, destacan la correlación de Spearman y los métodos de Kendall. Además, la correlación parcial permite evaluar la correlación entre dos variables mientras se controla el efecto de otras variables. Por otro lado, el análisis de regresión explora la relación entre una variable dependiente (observada) y una o más variables independientes (explicativas), asumiendo una relación causal. A diferencia de la correlación, permite inferir relaciones y predecir el valor de la variable dependiente a partir de las independientes, e incluye tanto modelos lineales como no lineales. Los modelos de regresión lineal pueden ser simples (una variable independiente) o múltiples (dos o más variables independientes). La regresión no lineal aborda relaciones más complejas, como las exponenciales o de potencia. La regresión logística extiende este análisis a casos donde la variable dependiente es dicotómica (binaria) o discreta (multinomial), en lugar de continua [53].

3.9.1. Índice de Morans

El Índice de Moran es una herramienta estadística que permite evaluar si la distribución espacial de una variable responde a un patrón aleatorio o, por el contrario, evidencia dependencia espacial. En términos concretos, mide si los valores altos o bajos tienden a concentrarse en zonas específicas (fenómeno conocido como agrupamiento) o si, por el contrario, se distribuyen de manera dispersa en el territorio. Para incorporar la dimensión geográfica, el índice se apoya en una matriz de pesos espaciales que define las relaciones de vecindad entre las unidades de análisis, ya sea por contigüidad, distancia o algún otro criterio funcional. Su interpretación se basa en una escala que va de -1 a +1: valores cercanos a +1 indican autocorrelación positiva, es decir, agrupamientos de valores similares, tales como áreas con altos niveles de pobreza ubicadas junto a otras con características parecidas; valores próximos a 0 sugieren ausencia de patrón espacial discernible, compatible con una distribución aleatoria; mientras que valores cercanos a -1 reflejan autocorrelación negativa, donde valores opuestos aparecen adyacentes por ejemplo, zonas de alta densidad poblacional colindantes con áreas de baja densidad [54]. Existen dos versiones del índice: la global, que resume el patrón general en todo el territorio, y la local, que identifica puntos o áreas específicas donde se generan

agrupamientos o contrastes significativos. Resulta especialmente útil en el análisis de fenómenos en los que la proximidad geográfica condiciona comportamientos o resultados, aunque su interpretación requiere cautela, pues los resultados pueden variar según la escala de análisis y la forma en que se defina la vecindad espacial.

3.9.2. Mapas LISA (Local Indicators of Spatial Association)

Son representaciones cartográficas que permiten identificar, de manera localizada, la existencia y naturaleza de patrones espaciales significativos en una variable de interés. A diferencia de los indicadores globales, que ofrecen un promedio único para todo el territorio, estos mapas desagregan la autocorrelación espacial, revelando dónde ocurren agrupamientos de valores similares, por ejemplo: altos junto a altos o bajos junto a bajos, así como situaciones de contraste, tales como: valores altos rodeados de bajos, o viceversa. Cada unidad espacial es evaluada en función de su relación con sus vecinas, ponderada por una matriz de pesos que define la estructura de vecindad, ya sea por contigüidad, distancia o flujo. El resultado se somete a pruebas de significancia estadística, lo que permite filtrar ruido y destacar únicamente aquellas asociaciones no atribuibles al azar. Visualmente, estos mapas utilizan colores o símbolos para codificar la dirección, intensidad y significancia del patrón detectado, facilitando la lectura geográfica de dinámicas complejas [55]. Más que una herramienta descriptiva, funcionan como un dispositivo de diagnóstico territorial, orientando la investigación hacia zonas críticas que merecen análisis contextualizados.

3.9.3. Modelo de Regresión Ponderada Geográficamente (GWR)

El Modelo de Regresión Ponderada Geográficamente (GWR) es una técnica de análisis espacial que extiende el modelo de regresión lineal múltiple tradicional al incorporar la ubicación geográfica de las observaciones de forma explícita [56]. Mientras que un modelo de regresión global estima un único conjunto de parámetros para toda el área de estudio, asumiendo que las relaciones son estacionarias en el espacio, el GWR reconoce que dichas relaciones pueden variar espacialmente. Por lo tanto, el GWR genera un conjunto único de coeficientes de regresión para cada entidad geográfica, permitiendo cuantificar y mapear la variación local de las relaciones entre variables.

Su funcionamiento se basa en realizar una regresión lineal independiente para cada punto en el espacio, pero con una diferencia clave: no todas las observaciones tienen el mismo peso en cada solución local. La influencia de una observación en el cálculo de los parámetros de un lugar específico se determina por su cercanía geográfica a ese lugar, utilizando una función de ponderación que asigna pesos que disminuyen con la distancia. Como resultado, el pro-

ceso genera un conjunto único de coeficientes para cada ubicación, cuyos valores reflejan las relaciones entre las variables que predominan en el entorno geográfico inmediato [56]. Este enfoque produce distribuciones espaciales continuas de los parámetros del modelo, transformando coeficientes que normalmente son estáticos en variables geográficas que muestran la heterogeneidad espacial de las relaciones subyacentes. Para validar el modelo, es necesario comparar si la variabilidad observada en estas superficies de parámetros es estadísticamente significativa en relación con la hipótesis nula de estacionariedad espacial.

3.10. Visualización interactiva de datos espaciales

La visualización interactiva de datos espaciales constituye un componente central dentro del análisis geográfico, ya que permite representar simultáneamente información multidimensional y facilitar su interpretación en contextos donde intervienen dinámicas territoriales complejas. A través de mapas dinámicos, controles temporales y gráficos vinculados, estas herramientas favorecen la identificación de patrones espaciales, variaciones temporales y relaciones entre variables que no son evidentes cuando la información se presenta de manera no gráfica. Este enfoque ha demostrado ser especialmente útil para sintetizar información georreferenciada de gran volumen y apoyar procesos de análisis espacial más rigurosos [57].

En el ámbito de la visualización interactiva de datos espaciales, diversas plataformas y librerías especializadas han ampliado significativamente la capacidad para representar información geográfica de manera dinámica. Herramientas basadas en la web, como Google Earth Engine [58], facilitan el procesamiento y despliegue de grandes volúmenes de datos satelitales y geoespaciales directamente en entornos interactivos. Complementariamente, librerías de código abierto como Folium [59] permiten la creación de mapas dinámicos integrados en aplicaciones web mediante la incorporación de controles, capas temáticas y funciones de exploración espacial. Por su parte, Plotly [60] aporta visualizaciones interactivas orientadas tanto a datos espaciales como no espaciales, permitiendo vincular gráficos, mapas y elementos descriptivos en una misma interfaz. En conjunto, estas herramientas potencian la comprensión de patrones territoriales complejos al traducir análisis técnicos y estadísticos en representaciones visuales intuitivas, replicables y accesibles para diversos tipos de usuarios. Su principal valor radica en que convierten procesos analíticos avanzados en representaciones visuales que pueden ser interpretadas con claridad por especialistas y actores institucionales, fortaleciendo la transparencia del análisis territorial y facilitando la toma de decisiones basada en evidencia.

3.11. Antecedentes

A continuación, se hace hincapié en algunos estudios de investigación los cuales se relacionan con la expansión urbana y pobreza utilizando imágenes satelitales resaltando sus aportes y diferencias: **“Mapping Socioeconomic Conditions Using Satellite Imagery: A Computer Vision Approach for Developing Countries”** [61]:

Este estudio utiliza un sistema de visión por computadora basado en redes neuronales para extraer indicadores de pobreza de imágenes satelitales de alta resolución. El método emplea transferencia de aprendizaje, lo que permite identificar elementos geográficos y construcciones visibles en imágenes satelitales de manera económica y escalable. Este enfoque se aplicó en Pakistán y tiene como objetivo mejorar la distribución de fondos y la evaluación de políticas para gobiernos y ONG.

- **Aportes:** El uso de transferencia de aprendizaje en este trabajo facilita una implementación eficiente en términos de costo, lo que es crucial para países en desarrollo. Además, su enfoque escalable podría replicarse en Bogotá D.C. para mapear indicadores de pobreza.
- **Diferencias:** Aunque el contexto de Pakistán tiene similitudes con Colombia en términos de desafíos socioeconómicos, Bogotá D.C. requiere la incorporación de variables geográficas y demográficas específicas para entornos metropolitanos.

“Estimación de índice de pobreza multidimensional (IPM) en Bogotá D.C. y ciudades cercanas usando imágenes satelitales” [62]: Este estudio emplea redes neuronales convolucionales (CNN) y regresión tensorial (GTR) para estimar el IPM en Bogotá D.C. y áreas circundantes. A partir de imágenes de alta resolución de Planet Scope y datos censales de 2018, los resultados muestran que el modelo GTR supera al CNN en la precisión de la clasificación de píxeles para la distribución del IPM, permitiendo una mejor visualización de la pobreza en áreas urbanas.

- **Aportes:** Este modelo combina datos satelitales con información censal, destacándose como una herramienta robusta para abordar la desigualdad urbana.
- **Diferencias:** Aunque está diseñado para Bogotá D.C., el modelo podría ampliarse para abarcar análisis temporales que evalúen cómo cambian las condiciones de pobreza a lo largo del tiempo, integrando nuevos datos de imágenes satelitales y socioeconómicos.

“Monitoring Urban Expansion Using Remote-Sensing Data Aided by Google Earth Engine” [63]: Utiliza Google Earth Engine y un modelo de clasificación de vectores de soporte (SVM) para evaluar la expansión urbana en Isfahan, Irán, entre 1986 y 2019. Logró una precisión del 96 % en la clasificación de áreas de crecimiento urbano, recomendando

futuras investigaciones sobre el impacto ambiental de la expansión urbana, como la formación de islas de calor.

- Aportes: Destaca la capacidad de integrar herramientas de acceso libre como Google Earth Engine con técnicas de aprendizaje automático para realizar análisis históricos de expansión urbana. Esto podría aplicarse para estudiar patrones de crecimiento urbano en Bogotá D.C. y su impacto ambiental.
- Diferencias: Bogotá D.C. presenta un contexto más dinámico y diverso que Isfahan, lo que requeriría adaptar el modelo para incluir las variables adicionales relacionadas con la pobreza y desigualdad.

“Análisis de patrones espaciales en la expansión urbana de ciudades intermedias. El caso de San Fernando, Chile” [64]: Este análisis de expansión urbana en San Fernando utiliza modelos de cambio de uso de suelo (LUCC) y fotografías aéreas. Identifica una configuración policéntrica en el crecimiento de la ciudad y sugiere que la planificación urbana debe considerar estas dinámicas para evitar una expansión no planificada.

- Aportes: Resalta cómo los modelos LUCC pueden proporcionar información sobre patrones espaciales en ciudades intermedias, lo que es valioso para la planificación urbana en zonas periféricas de Bogotá D.C., donde el crecimiento policéntrico está emergiendo.
- Diferencias: Bogotá D.C., al ser una metrópolis, requiere un enfoque integral que considere sus complejos sistemas de transporte, alta densidad poblacional y dinámicas económicas.

“Spatial Monitoring of Urban Expansion Using Satellite Remote Sensing Images: A Case Study of Amman City, Jordan” [3]: Emplea imágenes satelitales Landsat y clasificación supervisada para monitorear la expansión urbana en Amman, Jordania. El estudio analiza cómo la expansión afecta la cobertura vegetal y otras áreas, identificando patrones de crecimiento en torno a rutas de transporte, lo cual es valioso para el monitoreo espacial en entornos urbanos.

- Aportes: Su análisis sobre la relación entre expansión urbana y cobertura vegetal es directamente aplicable en Bogotá D.C., donde la conservación ambiental es una preocupación creciente. La metodología también podría usarse para identificar áreas críticas de cambio urbano.
- Diferencias: En comparación con Amman, Bogotá D.C. enfrenta presiones adicionales, como la interacción entre crecimiento urbano y ecosistemas como los humedales. Por lo tanto, el modelo debe adaptarse para considerar estos aspectos y sus efectos en la pobreza

“Evaluación del efecto de la pandemia sobre la deforestación en Colombia mediante imágenes de satélite y datos de Terra-i” [65]: Este análisis emplea aprendizaje supervisado y datos de Terra-i para evaluar cómo la pandemia impactó la deforestación en Colombia, observando una disminución en la pérdida forestal durante el período de crisis. Se observa cómo la deforestación se ve influenciada por restricciones de movilidad y disminución de actividad económica.

- **Aportes:** Ilustra cómo eventos externos, como la pandemia, pueden alterar significativamente patrones de uso del suelo. Este enfoque puede adaptarse para estudiar cómo las restricciones afectaron el desarrollo urbano en Bogotá D.C. durante el mismo período.
- **Diferencias:** A diferencia de la deforestación rural, Bogotá D.C. enfrenta retos urbanos que requieren un análisis más enfocado en el uso del suelo y la densidad poblacional.

“Urban Expansion and the Urban–Rural Income Gap: Empirical Evidence from China” [66]: Este estudio emplea un modelo de regresión cuantílica de panel para analizar cómo la expansión urbana afecta la desigualdad de ingresos en China. Encuentra una relación en forma de U invertida, sugiriendo que la expansión inicial reduce la brecha de ingresos, pero una vez superado un umbral, puede incrementarla si no se aplican políticas de inclusión adecuadas.

- **Aportes:** Proporciona un marco analítico para estudiar la interacción entre expansión urbana y desigualdad económica, que podría ser útil para entender los efectos socioeconómicos del crecimiento urbano en Bogotá D.C.
- **Diferencias:** En contraste con China, Bogotá D.C. tiene una estructura socioeconómica más fragmentada, por lo cual se tendría que ajustar el modelo para reflejar desigualdades específicas.

“Documento de Bases Teóricas de Algoritmo (ATDB): RAISG-MapBiomias Colombia Colección 2” [67]: Esta iniciativa aplica la clasificación de cobertura del suelo en las distintas regiones geográficas del país utilizando el algoritmo Random Forest, junto con datos satelitales Landsat y filtros temporales que aseguran la coherencia a lo largo del tiempo, en un periodo comprendido entre 1985 y 2023.

- **Aportes:** Ofrece un marco metodológico replicable para la post-clasificación y depuración de mapas de cobertura del suelo, adaptado al contexto colombiano. La integración de filtros temporales permite mejorar la precisión y continuidad de las series, aspecto fundamental para estudios de cambio de uso del suelo y monitoreo territorial.
- **Diferencias:** A diferencia de MapBiomias, que contempla múltiples clases de cobertura y transiciones detalladas, la metodología propuesta en este estudio se enfoca únicamente en dos categorías: urbano y no urbano. Esta simplificación permite centrar el análisis

en procesos de expansión urbana, aunque implica una menor desagregación temática en comparación con la clasificación multiclase de la iniciativa.

4. Construcción de base de datos de imágenes satelitales e indicadores socioeconómicos de pobreza

Con el objetivo de construir una base de datos que consolide información satelital y socioeconómica de Bogotá D.C. durante la última década, se integraron datos de diferentes fuentes. Para el componente espacial, se usaron imágenes del satélite Landsat 8 del periodo 2014–2024, a partir de las cuales se generaron composiciones anuales e índices espectrales relacionados con la cobertura y el uso del suelo. De forma complementaria, se incorporaron los datos de la EM de Bogotá D.C. de los años 2014, 2017 y 2021 consolidadas a nivel de UPZ.

4.1. Procesamiento de imágenes

La Figura 4.1 presenta el diagrama de flujo del proceso de preprocesamiento de las imágenes Landsat 8, donde se resumen las etapas empleadas para construir las bases de datos espaciales utilizadas en este estudio.

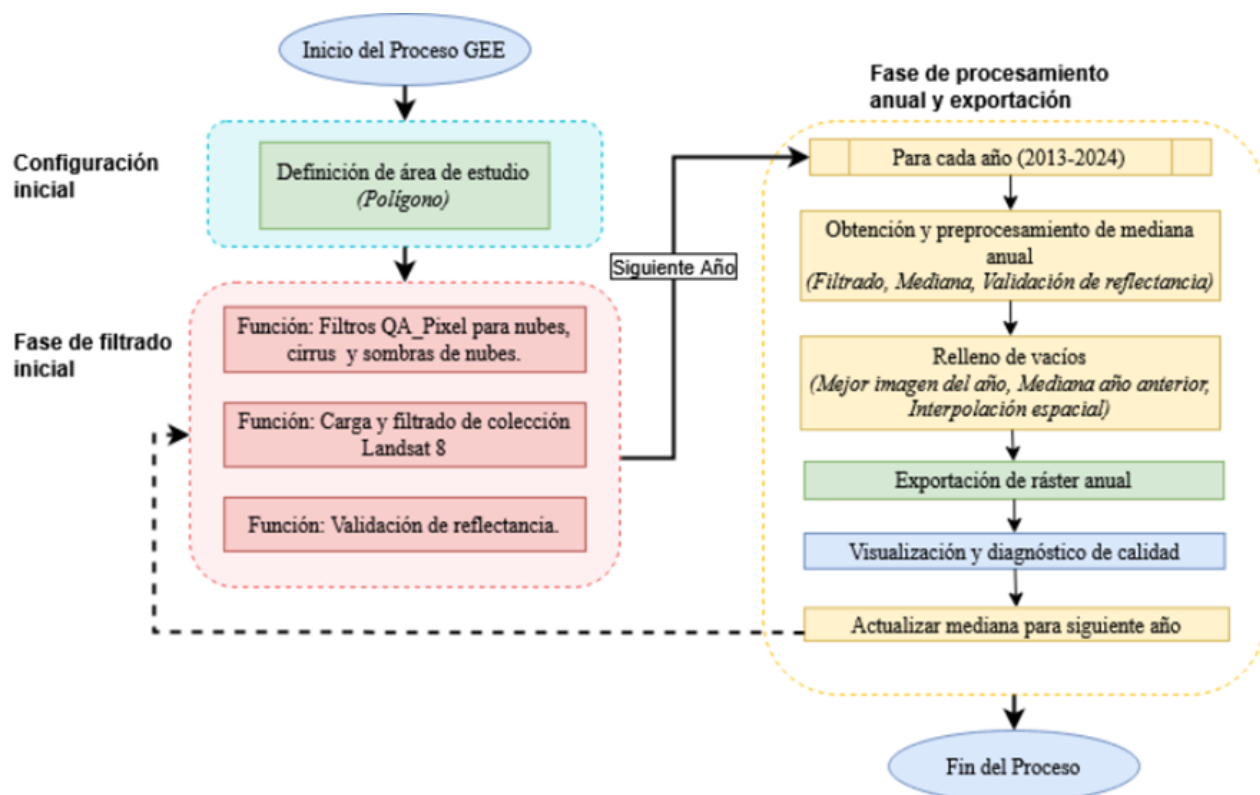


Figura 4.1: Diagrama de flujo preprocesamiento de imágenes Landsat 8

4.1.1. Descripción de las imágenes utilizadas

Para analizar el fenómeno de expansión urbana, se decidió utilizar el satélite Landsat 8, que ofrece registros desde el año 2013, ofreciendo cobertura temporal adecuada para el periodo de análisis de una década. Aunque Landsat 9 [68], (lanzado en 2021) tiene especificaciones similares, su espacio de observaciones es más limitado. Sentinel-2 [69], por su parte, cuenta con mejor resolución espacial y temporal, pero como se mencionó anteriormente su disponibilidad comienza en 2015, lo que reduce la posibilidad de análisis en este caso. Landsat 7, aunque sigue en operación desde 1999, presenta fallos en su sensor ETM+ desde 2003, generando artefactos y pérdida de información en un 22 % de cada escena [70].

Es importante destacar que las primeras escenas de Landsat 8 disponibles en 2013 presentaban limitaciones en cobertura y calidad radiométrica, lo que hacía poco confiable su aplicación. Por esta razón, el periodo de análisis definido corresponde a los años 2014 a 2024, intervalo que ofrece condiciones más estables y homogéneas para la construcción de la base de datos satelital y el seguimiento de la expansión urbana en Bogotá D.C.

4.1.2. Área de estudio

El área de estudio comprende el perímetro urbano delimitado por las Unidades de Planeamiento Zonal (UPZ) de la ciudad. Para su delimitación se empleó el Path 008 y Row 057, que corresponden al sistema de coordenadas WRS-2 utilizado por Landsat 8 para identificar escenas, las cuales cubren un área aproximada de 185×180 km [71].

Sin embargo, dado que esta extensión es mayor a los límites de la ciudad, se decidió utilizar un recorte espacial definido por las longitudes -74.37 a -73.85 y las latitudes 4.43 a 4.92 , en el sistema de referencia EPSG:4326 (WGS 84), como se muestra en la Figura 4.2. Esta delimitación tuvo como objetivo reducir el volumen de datos a procesar y concentrar el análisis en el territorio urbano de Bogotá D.C.

Una vez establecido el mejor sensor a utilizar para el mapeo de áreas urbanas, el siguiente paso es descargar los datos para así construir la base de datos satelitales. Para la descarga de las imágenes, aunque estas se encuentran disponibles en los repositorios del USGS [72], se usó la plataforma Google Earth Engine (GEE) ya que permite realizar el preprocesamiento directamente en la nube, evitando grandes costos de almacenamiento de datos en local [58].

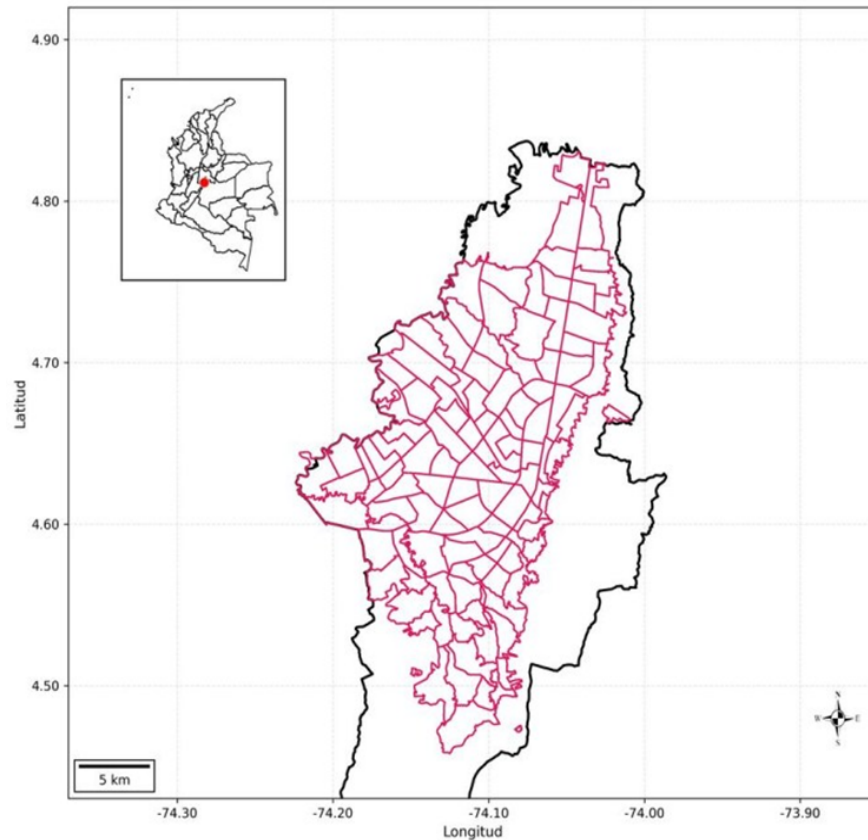


Figura 4.2: Área de estudio seleccionada

4.1.3. Preprocesamiento de las Imágenes

Con el objetivo de generar una base de datos multitemporal de alta calidad para el análisis de la expansión urbana en Bogotá D.C., se implementó un flujo de trabajo de preprocesamiento en la GEE. Este proceso incluye la selección, filtrado, corrección atmosférica, enmascaramiento de nubes, interpolación de datos faltantes y cálculo de índices espectrales clave. A continuación, se describe paso a paso la metodología empleada.

4.1.4. Selección del área de estudio y filtrado de la colección de Landsat-8

Se realizó la selección de la colección Surface Reflectance Collection 2 de Landsat 8 para el periodo 2013-2024, utilizando el área de estudio previamente definida. Esta versión incorpora mejoras en la precisión de la georreferenciación, lo que garantiza que las imágenes se alineen correctamente con las características geográficas reales, así como una corrección atmosférica que reduce los efectos de dispersión de la luz ocasionados por partículas en la atmósfera, como polvo y gotas de agua. Sin dicha corrección, las imágenes podrían presentar variaciones

en el brillo y el contraste atribuibles a condiciones atmosféricas y no a características reales de la superficie terrestre.

Adicionalmente, las imágenes cuentan con corrección de reflectancia superficial, lo que implica que los valores registrados representan la cantidad real de luz reflejada por la superficie, asegurando su comparabilidad entre los años analizados [73].

También se seleccionaron las bandas espectrales 1 a 7 del sensor OLI (Operational Land Imager), que cubren longitudes de onda relevantes para la caracterización del suelo urbano. Estas bandas incluyen el espectro visible (azul, verde y rojo), el infrarrojo cercano (NIR) y dos bandas en el infrarrojo de onda corta (SWIR), las cuales son particularmente sensibles a la reflectancia de materiales artificiales como el concreto, el asfalto, las tejas y otros recubrimientos típicos de las zonas construidas. Además, estas bandas sirven de insumo para calcular los índices espectrales.

4.1.5. Enmascaramientos de calidad

Con el fin de garantizar la calidad de las imágenes, se implementó una función de enmascaramiento de nubes y artefactos atmosféricos basada en la banda QA_PIXEL de los productos de Nivel 2 de Landsat-8 [72]. Esta banda codifica en diferentes bits la presencia de elementos no deseados como nubes, nubes dilatadas, cirros, sombras de nubes o áreas sin información. A partir de estos indicadores se construyó una máscara combinada que elimina del área de estudio los píxeles afectados.

Adicionalmente, se consideraron los niveles de confianza incluidos en los bits 8–9 (nubes) y 10–11 (sombras de nube), cuyo rango va de 0 (sin confianza) a 3 (alta confianza). Para este trabajo, únicamente se mantuvieron los píxeles clasificados con valores de confianza bajos (0–1), excluyendo aquellos con mayor probabilidad de presencia de nubes o sombras.

4.1.6. Construcción de composiciones anuales

Se generaron composiciones anuales de reflectancia de las bandas elegidas de la colección de imágenes de Landsat-8. Para ello se utilizó la mediana anual de los píxeles válidos en cada banda espectral, un estadístico robusto frente a valores extremos que atenúa la influencia de observaciones atípicas (residuos de nubes, sombras o errores radiométricos) y permite sintetizar las condiciones representativas de cada año en una única imagen.

Adicionalmente, para garantizar la correcta interpretación de los valores radiométricos, se aplicó el factor de escala definido en la documentación oficial de Landsat 8 Surface Reflectance

[71], que convierte los valores digitales almacenados en el producto en unidades físicas de reflectancia superficial. La transformación se expresa mediante la siguiente ecuación:

$$\rho = (DN \times 0,0000275) - 0,2 \quad (28)$$

donde:

- ρ es la reflectancia de superficie,
- DN es el valor digital original de la banda,
- 0,00002750 es el factor multiplicativo de escala,
- $-0,2$ Es el factor aditivo de corrección.

Los valores obtenidos deben ajustarse al rango físico esperado de reflectancia (0–1) [74]. Sin embargo, debido a errores radiométricos, condiciones atmosféricas no corregidas o saturación en los sensores, se presentaron valores fuera de dicho rango. Estos valores fueron tratados como anomalías radiométricas y corregidos mediante la misma estrategia aplicada a los vacíos espaciales, evitando afectar la coherencia de las composiciones anuales.

La corrección de vacíos se realizó mediante el siguiente enfoque jerárquico:

- Selección de la mejor escena anual: Los píxeles inválidos se reemplazaron por píxeles con la mejor cobertura nubosa, teniendo en cuenta que cumplieran con los criterios definidos en la banda QA-Pixel.
- Uso de información temporal previa: los vacíos persistentes fueron completados mediante la composición anual del año anterior, asumiendo la ausencia de cambios significativos al interior del área urbana.
- Interpolación espacial local: Finalmente, los valores restantes fueron tratados mediante un proceso de promedio focal, los cuales fueron aplicados a clústeres menores a 50 píxeles, evitando comprometer la calidad de los resultados. Este método estima el valor de un píxel faltante a partir del promedio de los píxeles cercanos, en este caso se usó una ventana de 5 x 5 píxeles.

El proceso de interpolación puede se expresa como [26]:

$$\widehat{X}_{i,j} = \frac{1}{N} \sum_{u=-r}^r \sum_{v=-r}^r X_{i+u,j+v} \quad (29)$$

donde:

- $\widehat{X}_{i,j}$ es el valor interpolado del píxel central,
- $X_{i+u,j+v}$ representa los valores válidos de los píxeles vecinos dentro de la ventana,
- r es el radio del kernel (en este caso $r = 2$, correspondiente a la ventana de 5×5 píxeles),
- y N el número total de píxeles válidos dentro de la ventana.

Con base en las composiciones anuales generadas, se calcularon los índices espectrales NDVI, NDWI y NDBI a partir de las siete bandas extraídas del sensor OLI de Landsat 8. Estos índices, visualizados en la Figura 4.3, fueron ajustados según los percentiles 2 y 98 con el fin de diferenciar de manera más efectiva las coberturas vegetales, los cuerpos de agua y las superficies construidas. Esto permite concentrar el contraste en la variabilidad real de las coberturas presentes en la zona de estudio al reducir la influencia de valores extremos.

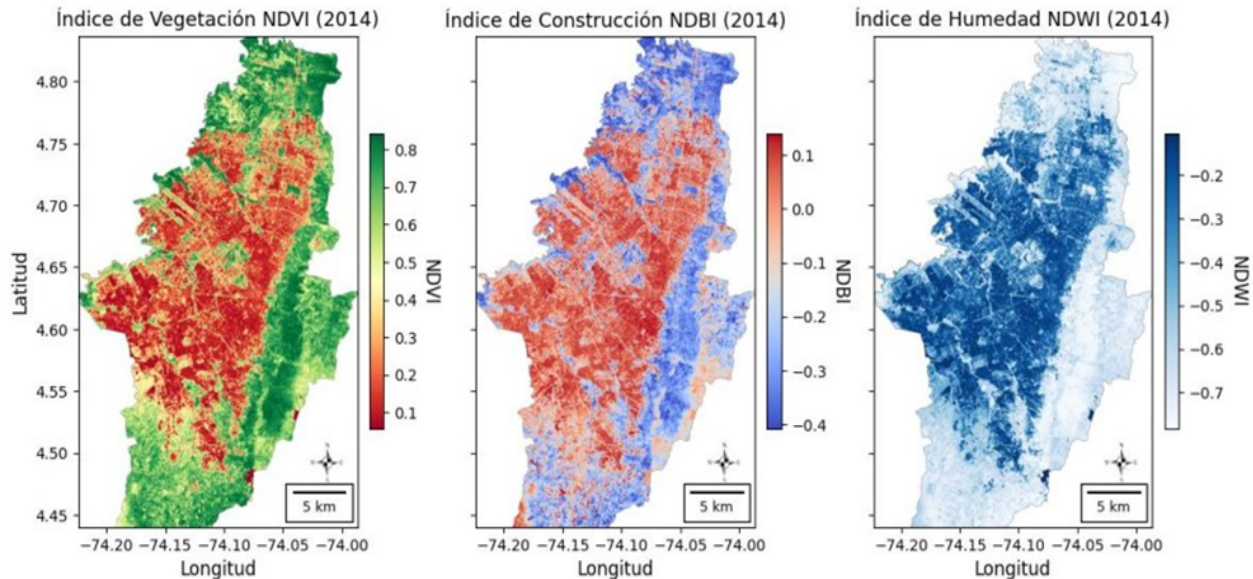


Figura 4.3: Visualización de Índices Espectrales – 2014

La Figura 4.4 presenta las composiciones anuales del área de estudio en Bogotá D.C., cubriendo el periodo 2013–2024. Cada imagen muestra una composición de color natural (RGB) utilizando las bandas 4, 3 y 2 del sensor OLI de Landsat 8, que sirve como base visual para el análisis multitemporal de los cambios urbanos. Se resalta que las áreas en color rojo corresponden a píxeles sin información válida, resultantes tras aplicar la estrategia de relleno de vacíos, los cuales se asocian principalmente a nubes y sombras que no fue posible corregir completamente durante el preprocesamiento.

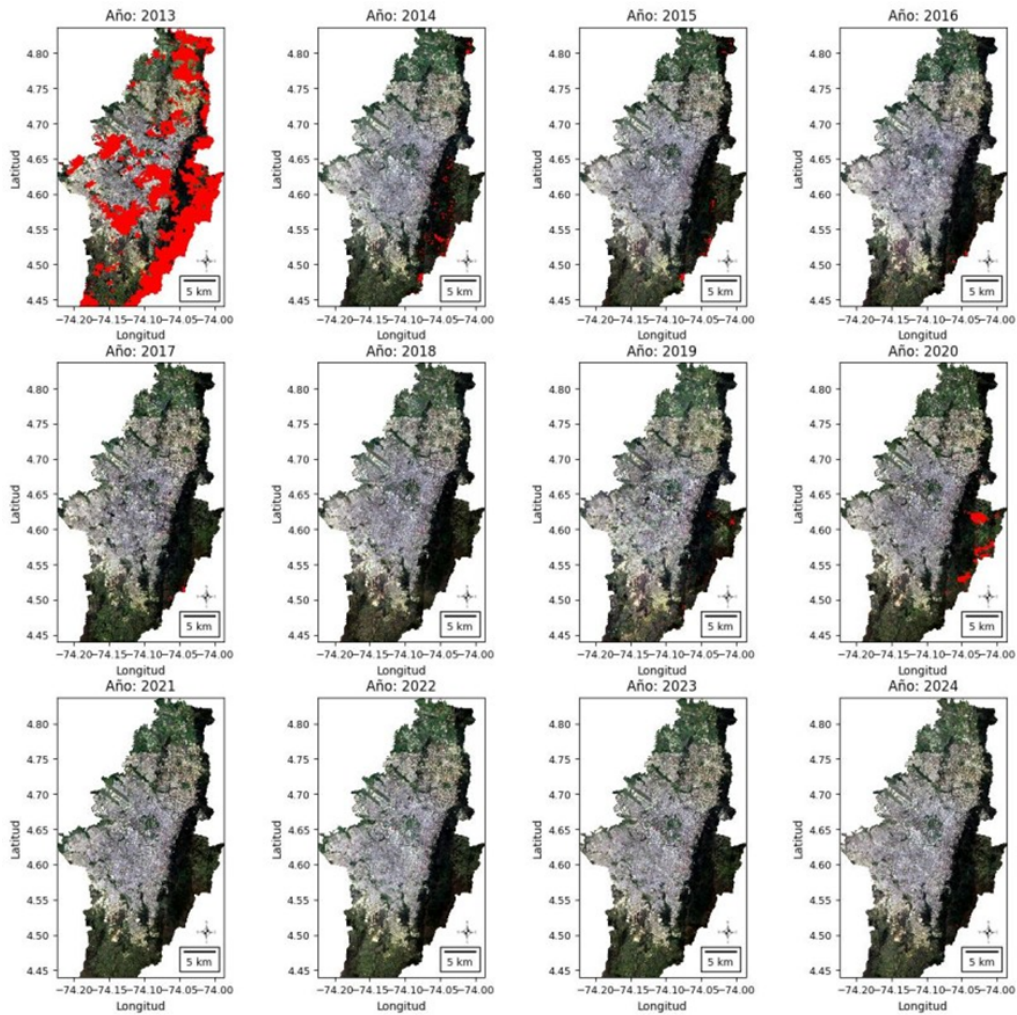


Figura 4.4: Resultados del preprocesamiento de las imágenes Landsat 8 en el periodo 2013-2024

4.2. Procesamiento de bases de datos pobreza

Para el análisis de la pobreza en Bogotá D.C., se utilizó como fuente principal la EM de los años 2014, 2017 y 2021 [75]. Esta encuesta ofreció información detallada sobre las condiciones de vida de la población. A continuación, se detallan sus principales características.

4.2.1. Extracción de los datos de la EM

La integración de los datos de la EM con información geoespacial, como mapas de cobertura del suelo o imágenes satelitales, facilita la identificación de áreas críticas donde se concentran problemas como la pobreza, la falta de acceso a servicios básicos o la degradación ambiental

[2]. Por ejemplo, es posible correlacionar los niveles de pobreza con la presencia de áreas deforestadas o con la falta de infraestructura en zonas rurales.

Luego de descargar las encuestas para los 3 años, se escogieron un total de 25 variables, que comprenden algunos datos de identificación (como el ID de la encuesta, el ID del hogar, el factor de expansión, entre otros) y los distintos subcomponentes para realizar el cálculo del IPM. El proceso completo se detalla en la Figura 4.5.

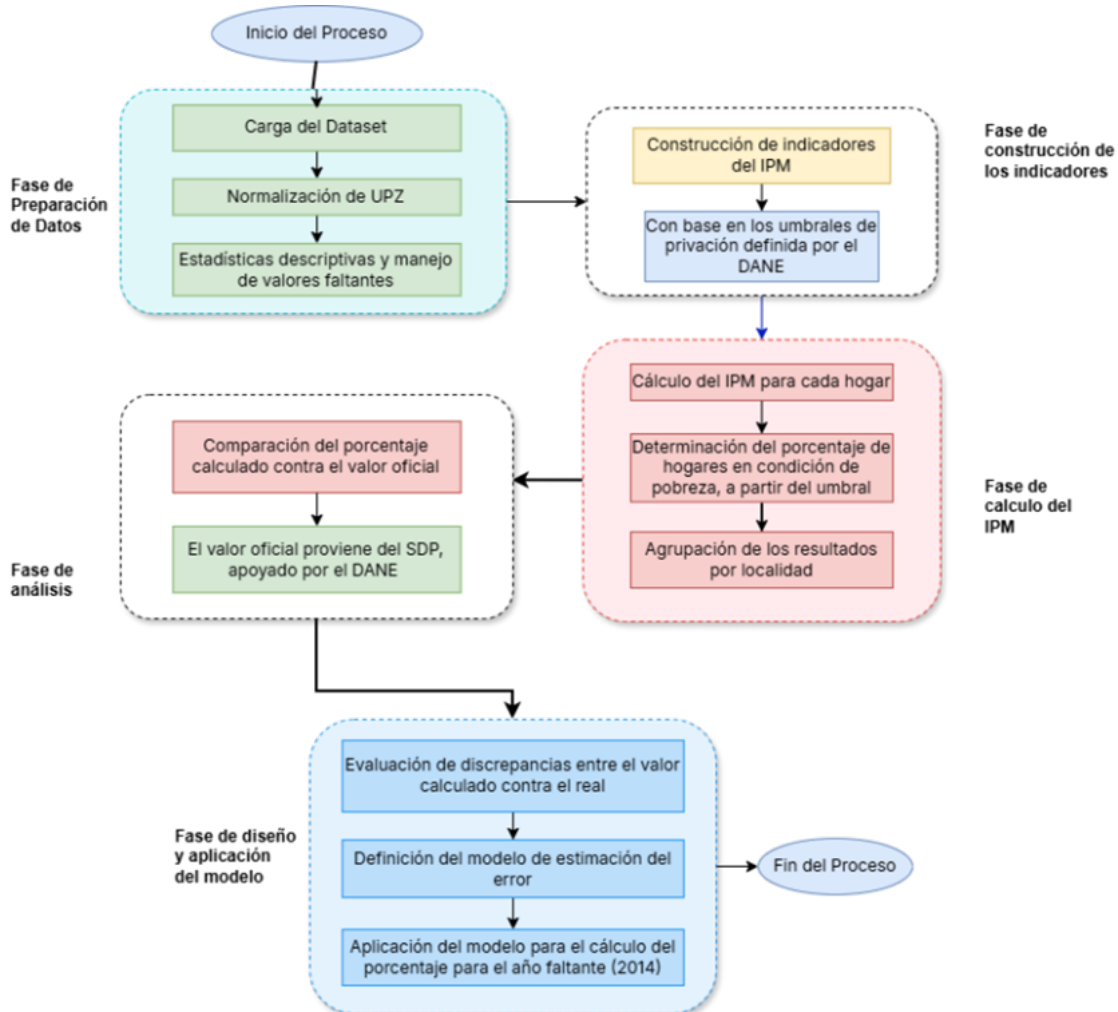


Figura 4.5: Diagrama de flujo metodológico para el cálculo del IPM, para los años 2014, 2017 y 2021

En la fase de preparación de los datos, se realizó la identificación de los tipos de datos correspondientes a cada columna, así como de las posibles respuestas asociadas a estas. Adicionalmente, se procedió a la estandarización de los nombres de las columnas entre las encuestas aplicadas en los distintos años. Durante este proceso, se evidenció que los códigos de

las UPZ no coincidían entre los diferentes periodos, debido a que, posteriormente a la encuesta de 2014, algunas UPZ fueron desagregadas. Por consiguiente, fue necesario reagrupar las UPZ con el fin de garantizar una estructura comparable a lo largo de los años. Posteriormente, se efectuó un análisis descriptivo de la base de datos, identificando un total de 456.912 encuestas. Finalmente, se evaluó la cantidad de valores faltantes en cada una de las columnas, cuyos resultados se presentan en la Tabla 4.1.

Tabla 4.1: Variables con valores faltantes de las encuestas multipropósito

Variable	Faltantes	%
tratam_enf	442,481	96.84
cuidado_menor	434,120	95.01
trabajo_nopag_ultm_sem	273,619	59.88
trabajo_ultm_sem	270,938	59.30
activ_pag_ultm_sem	262,119	57.37
ultm_grado	130,968	28.66
pension	78,881	17.26
ocupacion_ultm_sem	49,854	10.91
read_and_write	22,792	4.99
estudia	22,792	4.99
cod_upz	16,390	3.59

A partir de los formularios empleados para la realización de las encuestas [76], se definieron reglas específicas para la asignación de las etiquetas “No aplica” y “No responde”, con el propósito de completar los valores ausentes y permitir el cálculo del Índice de Pobreza Multidimensional (IPM) de forma coherente y consistente.

En la fase de construcción de los indicadores, los valores se determinaron siguiendo las directrices establecidas en las notas metodológicas del DANE [47]. Para cada componente, se establecieron umbrales que definen el punto a partir del cual un hogar se considera privado de un determinado componente y, en consecuencia, presenta una mayor probabilidad de ser clasificado como hogar en condición de pobreza. Dado que los registros se presentan a nivel de encuesta, fue necesario agrupar la información para obtener resultados consolidados por hogar.

4.2.2. Cálculo del IPM

Una vez calculados todos los componentes por hogar, estos fueron sumados con sus respectivos pesos porcentuales para determinar el valor final del IPM. Un hogar se considera en condición de pobreza cuando su valor supera el umbral de 0,33. A partir de esta clasificación, se estimó el porcentaje de hogares pobres por localidad y por año.

Cabe señalar que, si bien la unidad de análisis definida para el desarrollo del proyecto corresponde a la UPZ, se optó por realizar la agregación a nivel de localidad, dado que la encuesta de 2014 presenta su desagregación en dicho nivel. Por tanto, con el fin de mantener la comparabilidad entre los años analizados, se trabajó con el mismo nivel de desagregación para todas las series temporales.

Con base en ello, se estimó que el porcentaje de hogares en condición de pobreza fue de 14,62 % en 2014, 10,38 % en 2017 y 10,78 % en 2021. Para verificar la fiabilidad de estos resultados, se contrastaron con los porcentajes oficiales reportados por la Secretaría Distrital de Planeación de Bogotá D.C. (SDP) para los mismos años [77], los cuales fueron de 3,75 % en 2017, 4,35 % en 2021 y sin valor disponible para 2014, dado que no se realizó un cálculo oficial para dicho año. Estas diferencias evidencian una discrepancia significativa entre los valores estimados y los reportados oficialmente, lo que hizo necesario aplicar una técnica o modelo que permitiera estimar las “variables latentes” responsables de dichas divergencias, con el fin de obtener una estimación más consistente del porcentaje correspondiente al año 2014.

4.2.3. Aplicación de modelo de estimación

Para la estimación de los valores requeridos, y tras evaluar las distribuciones de los indicadores calculados por localidad y por año (correspondientes a los periodos 2017 y 2021), se identificó que dicho indicador no seguía una distribución normal. Asimismo, se observó que los subgrupos definidos por localidad y año podían presentar distintos grados de sesgo. En consecuencia, se optó por emplear un modelo de estimación del error que no se viera significativamente afectado por dichas características, seleccionándose para este propósito la estimación por Coincidencia de Cuantiles (MQE).

En lugar de usar un único umbral global (0.33) para determinar si un hogar se encuentra en condición de pobreza, en este método se ajusta un pequeño desplazamiento (ε_r) para cada localidad r , de modo que la cantidad de hogares cuya puntuación (IPM) supera $0,333 + \varepsilon_r$ coincida con la cantidad “oficial” de hogares en condición de pobreza para cada región. Posteriormente, suponiendo que dicho desplazamiento varía aproximadamente de forma lineal en el tiempo, se ajusta una línea que pase por $(2017, \varepsilon_{r,2017})$ y $(2021, \varepsilon_{r,2021})$, y se extiende dicha recta hasta el 2014 para obtener $\varepsilon_{r,2014}$. Finalmente, para cada encuesta de 2014, se calculará el “IPM_ajustado” = $IPM_base - \varepsilon_{r,2014}$.

Una vez obtenido el nuevo IPM ajustado (IPM_a), se recalcularon los porcentajes de hogares en condición de pobreza para los años con datos disponibles (2017 y 2021), y se procedió a estimar el valor correspondiente al año 2014. En la Tabla 4.2 se presentan los resultados

comparativos: el porcentaje inicial derivado del IPM base, el porcentaje ajustado calculado mediante el IPM_a , los porcentajes oficiales reportados por la Secretaría Distrital de Planeación (SDP) y la diferencia entre los valores del IPM_a y los porcentajes oficiales para los años 2017 y 2021.

Tabla 4.2: Resumen de porcentajes de personas en condición de pobreza por año

Año	% inicial (IPM base)	% ajustado (IPM_a)	% oficial	% diferencia
2014	14.62	8.90	*	*
2017	10.38	3.75	4.91	1.16
2021	10.78	4.35	5.50	1.15

5. Identificación de dinámicas de expansión urbana usando técnicas de aprendizaje de máquina

Este capítulo tiene como objetivo identificar, caracterizar y delimitar las dinámicas de expansión urbana en Bogotá D.C. durante la última década (2014–2024), integrando metodologías de teledetección con técnicas de inteligencia artificial. A través del análisis de series temporales de imágenes satelitales y la aplicación de métodos de clasificación supervisada y no supervisada de umbralización, se busca mapear los cambios en la cobertura y uso del suelo, diferenciar áreas urbanas de no urbanas con base en valores espectrales e índices derivados, y comprender los factores que impulsan la expansión.

5.1. Clasificación supervisada

La Figura 5.1 muestra el flujo metodológico usado en la aplicación de los modelos supervisados usados para discriminar la cobertura del suelo.

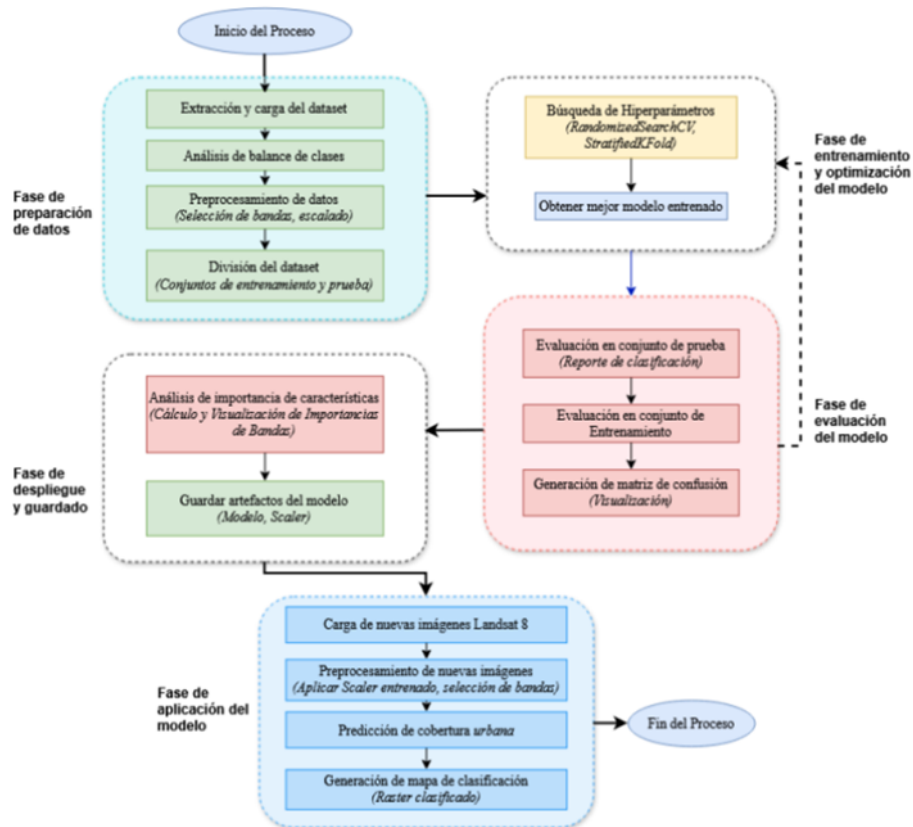


Figura 5.1: Diagrama de flujo metodológico para la clasificación de cobertura urbana en la ciudad de Bogotá D.C.

5.1.1. Generación y preparación de muestras de entrenamiento

Para la extracción de muestras, se consideraron distintas alternativas, entre las que se encontraban datos de la cobertura de Bogotá D.C. generados por distintos programas de monitoreo global. Entre ellos, se evaluó el uso del Copernicus Global Land Service (CGLS-LC100) [78], que ofrece mapas de cobertura terrestre con una resolución espacial de 100 metros para el periodo comprendido entre 2015 y 2019, así como la capa de ESA WorldCover, que proporciona mapas globales de cobertura a una resolución de 10 metros para los años 2020 y 2021 [79].

No obstante, ambos productos presentan limitaciones que comprometen su utilidad el entrenamiento de modelos de clasificación a escala urbana. En el caso de CGLS-LC100, la resolución espacial de 100 metros dificulta la extracción de etiquetas para áreas urbanas, puesto que un solo píxel puede contener múltiples tipos de cobertura, lo cual introduce ruido en los datos de entrenamiento. Por otra parte, aunque ESA WorldCover mejora la resolución espacial a 10 m, su principal limitante radica en la poca cobertura temporal al solo estar disponible para dos años.

Esto resulta problemático, ya que para mejorar la capacidad de generalización de los modelos y capturar variaciones temporales en los patrones de cobertura urbana, es necesario contar con muestras representativas de diferentes años. Esto garantiza que los modelos aprendan a reconocer la cobertura urbana incorporando condiciones variables en el tiempo, como cambios en la iluminación, diferencias en la humedad del suelo o la presencia de sombras proyectadas. Estas condiciones afectan directamente la apariencia de una misma cobertura de un año a otro. En consecuencia, si el conjunto de entrenamiento no incluye muestras que reflejen esta diversidad temporal, el modelo corre el riesgo de sobreajustarse a las condiciones específicas de ciertos años y generalizar de manera errónea en los demás periodos.

Para abordar esta problemática, se optó por extraer etiquetas de píxeles urbanos y no urbanos del área de estudio a partir de las clasificaciones previas proporcionadas por la Colección 2 del proyecto MapBiomias Colombia, desarrollado por la Fundación Gaia Amazonas. Este proyecto ofrece información detallada sobre la cobertura del suelo a nivel nacional para el periodo comprendido entre 1985 y 2023, así como para otros países de Sudamérica. Esta iniciativa fue desarrollada mediante el uso del algoritmo Random Forest, utilizando muestras de entrenamiento derivadas de mapas de referencia como OpenStreetMap, así como de fuentes oficiales como el MGN - Nivel Zona Urbana, integrado al Censo Nacional de Población y Vivienda 2018. Además, se basa en las colecciones de imágenes del satélite Landsat 8, que ofrecen una resolución espacial de 30 metros y se emplean para la generación de los mapas de cobertura y uso del suelo [67]. Debido a su cobertura temporal y espacial, este proyecto constituye fue elegido para la generación de muestras multitemporales destinadas para la aplicación de los modelos supervisados.

Con base en las etiquetas anuales generadas por MapBiomás Colombia, se implementó una metodología de extracción de muestras a partir de imágenes satelitales de Landsat 8 obtenidas desde la plataforma Google Earth Engine (GEE), correspondientes al periodo comprendido entre 2014 y 2023. Para cada año, se realizó la extracción de los valores de reflectancia de las bandas B1 a B7 y de los índices espectrales, considerando únicamente los píxeles clasificados como “Infraestructura Urbana” y las demás categorías fueron agrupadas como “No urbanas” teniendo en cuenta la información derivada de Mapbiomas.

Posteriormente, se aplicó un muestreo aleatorio estratificado, seleccionando 500 píxeles por clase y por año en el periodo 2014–2023, lo que equivale aproximadamente al 0.04 % del total de píxeles del área de estudio en cada año. En total se obtuvieron 10,000 muestras, distribuidas equitativamente entre las dos clases analizadas, con el propósito de garantizar una representación balanceada entre categorías y reducir posibles sesgos en el conjunto de entrenamiento, permitiendo que el modelo aprendiera de manera adecuada la variabilidad espectral y temporal presente en las imágenes Landsat, asociada a distintos años y condiciones atmosféricas.

El enfoque de muestreo aleatorio estratificado descrito garantiza balance entre clases y representatividad temporal, lo cual es adecuado desde una perspectiva de aprendizaje automático clásico. No obstante, al tratarse de datos georreferenciados, es necesario considerar la autocorrelación espacial, fenómeno mediante el cual las observaciones cercanas tienden a ser más similares que aquellas distantes. Esta dependencia espacial puede comprometer la independencia entre los conjuntos de entrenamiento y prueba si ambos contienen píxeles geográficamente próximos, lo que podría inducir un sesgo optimista en las métricas de desempeño del modelo, conocido como spatial leakage.

Para evaluar esta hipótesis, se aplicó el Índice de Moran global a las 10,000 muestras generadas. En la construcción de la matriz de pesos espaciales se adoptó un umbral de vecindad de 500 metros, valor que representa una escala intermedia relevante en el contexto urbano de Bogotá D.C., acorde con la dimensión típica de barrios o manzanas urbanas, y que permite capturar la agrupación espacial sin mezclar zonas con dinámicas territoriales muy distintas. El análisis arrojó un valor de $I = 0.684$ con un p valor menor a 0.001, lo que confirma la presencia de una fuerte autocorrelación espacial positiva. Este resultado indica que los píxeles pertenecientes a la misma clase tienden a agruparse geográficamente, un comportamiento esperable en contextos urbanos como el de Bogotá D.C., donde las áreas construidas y las zonas no urbanas forman manchas espaciales continuas.

Aunque en este trabajo se adoptó una partición aleatoria estratificada, con semilla fija para garantizar reproducibilidad, se reconoce que esta estrategia no aborda explícitamente la estructura espacial inherente a los datos. Esta limitación es común en estudios basados en

píxeles, pero sugiere la necesidad de explorar en futuras investigaciones enfoques de validación espacialmente conscientes, tales como la partición por bloques geográficos disjuntos [80]. A pesar de ello, la inclusión de muestras provenientes de múltiples años y de distintos sectores de Bogotá contribuyen a mitigar el efecto de la autocorrelación espacial, al introducir variabilidad tanto temporal como espacial en el conjunto extraído.

5.1.2. Evaluación de la separabilidad de las clases extraídas

Para verificar la diferenciación espectral entre las clases urbanas y no urbanas, se aplicaron las métricas de distancia Euclidiana, Mahalanobis y Jeffries–Matusita, previamente descritas en el marco teórico. Estas medidas permiten cuantificar la separación entre las distribuciones espectrales y evaluar el grado de superposición entre clases, asegurando así la validez y representatividad del conjunto de entrenamiento empleado en la clasificación supervisada.

Tabla 5.1: Métricas de Separabilidad Espectral entre las Clases

Distancia	Métrica
Distancia Euclidiana	0.6392
Distancia Mahalanobis	3.6422
Distancia Jeffries–Matusita	1.6190

La Tabla 5.1 muestra que las métricas de separabilidad espectral evidencian una diferenciación entre las clases extraídas de las imágenes de Mapbiomas Colombia y los valores de reflectancia de las bandas de Landsat 8. La distancia Euclidiana de 0.63 indica una separación moderada entre los centroides espectrales de ambas clases, mientras que la distancia Mahalanobis, con un valor de 3.64, sugiere una separación significativa considerando la varianza y correlación entre bandas. La distancia de Jeffries–Matusita (1.61) refuerza esta conclusión al indicar una baja superposición entre las distribuciones espectrales, siendo esta última útil por estar acotada entre 0 y 2, donde valores superiores a 1.5 implican alta separabilidad entre clases.

5.1.3. Desarrollo de los modelos supervisados

Las técnicas de aprendizaje supervisado seleccionadas para clasificar en la cobertura del suelo en los años seleccionados fueron Random Forest, SVM Regresión Logística y XGBoost, los cuales fueron implementados a partir de las muestras de suelo extraídas previamente. Esto permitió evaluar y comparar el rendimiento de cada algoritmo en la delimitación y detección de la cobertura urbana.

5.1.4. Entrenamiento de los modelos

La ejecución y el entrenamiento de los modelos se realizaron en el entorno de ejecución de Google Colab [81], que ofrece un entorno gratuito en la nube con acceso a CPU y compatibilidad con la librería Scikit-learn usada para la implementación de los modelos de clasificación y los procesos de validación [82].

El proceso comenzó con la preparación del conjunto de datos, a partir de muestras etiquetadas según la clasificación de MapBiomas y la extracción de los valores de reflectancia de imágenes Landsat 8. Posteriormente, los valores de cada banda e índice espectral fueron normalizados para garantizar que todas las variables se encontraran en la misma escala. A continuación, los datos fueron divididos en conjuntos de entrenamiento y prueba utilizando las muestras extraídas junto con los atributos espectrales calculados. Esta división se realizó mediante una estrategia hold-out estratificada, asignando el 70 % de las muestras al conjunto de entrenamiento y el 30 % al conjunto de prueba, empleando una semilla aleatoria con valor 42 para asegurar la reproducibilidad del experimento.

Sobre el conjunto de entrenamiento se aplicó validación cruzada estratificada con 10 particiones ($k = 10$), incorporando una etapa de mezcla aleatoria (shuffle) de las muestras antes de la generación de los folds y utilizando la misma semilla fija. Este enfoque incorpora una aleatoriedad controlada en la generación de las particiones, lo que permite obtener divisiones consistentes entre distintas ejecuciones del experimento. La validación cruzada permitió estimar de manera más estable el rendimiento del modelo y reducir el riesgo de sobreajuste.

Posteriormente, los hiperparámetros de los modelos fueron ajustados mediante una búsqueda aleatoria, en la que se evaluaron distintas combinaciones de parámetros utilizando validación cruzada estratificada. Para la selección del modelo óptimo se consideró el desempeño medido a través del F1-score sobre el conjunto de prueba, así como el tiempo total de procesamiento. El F1-score fue seleccionado por su capacidad de integrar precisión y recall en una única métrica, lo que permite una evaluación equilibrada del desempeño del clasificador. Adicionalmente, la inclusión del tiempo de procesamiento permitió identificar modelos que no solo ofrecieran un alto rendimiento predictivo, sino que también resultaran computacionalmente eficientes.

Finalmente, se evaluaron tres configuraciones diferentes de variables de entrada con el objetivo de analizar su influencia en el rendimiento de los modelos. En primer lugar, se utilizó únicamente la información proveniente de las bandas espectrales B1 a B7 de Landsat 8. En segundo lugar, se consideraron únicamente los índices espectrales (NDVI, NDBI y NDWI). Finalmente, se probó una combinación de bandas espectrales junto con los índices, permitiendo determinar qué tipo de variables son más relevantes para la clasificación del suelo urbano.

Con el objetivo de mantener la claridad metodológica, los resultados individuales obtenidos en cada una de las 10 particiones de la validación cruzada ($k = 10$) se presentan en los anexos donde se presentan las métricas de desempeño por modelo y configuración. En las siguientes subsecciones, se presentan únicamente los resultados promedio y comparativos que resumen el desempeño global de los modelos evaluados en el conjunto de prueba.

Adicionalmente, en el Anexo E se incluye el enlace al repositorio del proyecto en GitHub, el cual contiene los scripts de procesamiento, entrenamiento y evaluación de los modelos, así como los cuadernos (notebooks) y archivos de configuración empleados en el desarrollo del análisis.

5.1.5. Clasificación mediante Random Forest

El algoritmo Random Forest se aplicó para la clasificación de las coberturas urbano y no urbano, aprovechando su capacidad para manejar relaciones no lineales y reducir el sobreajuste mediante el uso de múltiples árboles de decisión. Su ejecución se realizó bajo el mismo esquema de validación cruzada y división de datos empleado en los demás modelos, con el fin de asegurar comparabilidad en los resultados.

La Tabla 5.2 presenta el espacio de búsqueda definido para la optimización de hiperparámetros del modelo. Se establecieron rangos discretos para cuatro hiperparámetros clave. El número de árboles en el bosque indica cuántos árboles individuales se entrenan y se combinan para generar la predicción final, lo cual influye en la precisión del modelo. La profundidad máxima de los árboles controla hasta qué nivel puede crecer cada uno, lo que permite capturar relaciones complejas, pero también aumenta el riesgo de sobreajuste si no se regula adecuadamente. El número mínimo de muestras requeridas para dividir un nodo interno establece cuántas observaciones son necesarias para permitir una nueva partición, lo que ayuda a regular la complejidad del modelo. Finalmente, el número mínimo de muestras necesarias en una hoja define el tamaño mínimo de los subconjuntos terminales, y su ajuste adecuado contribuye a prevenir que el modelo se ajuste a patrones específicos del conjunto de entrenamiento.

La Tabla 5.3 resume la mejor configuración de hiperparámetros obtenida para el modelo Random Forest según los tres tipos de variable de entrada utilizadas.

Tabla 5.2: Espacio de búsqueda de hiperparámetros usados en Random Forest

Hiperparámetro	Rango de búsqueda
Número de árboles	[100, 150, 200]
Profundidad máxima de los árboles	[10, 12, 15]
Mínimo número de muestras para dividir un nodo	[2, 5, 10]
Mínimo número de muestras en una hoja	[1, 2, 4]

Tabla 5.3: Mejor configuración de hiperparámetros en Random Forest por tipo de entrada

Hiperparámetro	Bandas espectrales (B1–B7)	Índices espectrales (NDVI, NDBI, NDWI)	Combinación (Bandas + Índices)
Número de árboles	100	150	150
Profundidad máxima de los árboles	10	10	12
Mínimo número de muestras para dividir un nodo	5	5	5
Mínimo número de muestras en una hoja	2	1	1

La Tabla 5.4 presenta el desempeño del modelo Random Forest evaluado las tres configuraciones de entrada definidas. Los resultados detallados por partición se incluyen en el Anexo A.

Tabla 5.4: Comparativa de desempeño de Random Forest según tipo de variables de entrada

Métricas	Bandas espectrales (B1–B7)	Índices espectrales (NDVI, NDBI, NDWI)	Combinación (Bandas + Índices)
Tiempo total (s)	185	115.97	255.33
Accuracy	0.9383	0.9293	0.9423
Precision	0.9384	0.9298	0.9424
Recall	0.9383	0.9293	0.9423
F1-score	0.9383	0.9293	0.9423

5.1.6. Clasificación mediante SVM

El modelo Support Vector Machine (SVM) se utilizó como uno de los clasificadores supervisados para la identificación de áreas urbanas en la ciudad. Su implementación se realizó sobre las mismas particiones de validación y conjuntos de entrenamiento definidos previamente, con el fin de garantizar la comparabilidad de resultados entre modelos.

La Tabla 5.5 presenta el espacio de búsqueda definido para la optimización de hiperparámetros en este modelo. El parámetro de regularización controla el equilibrio entre maximizar el margen y minimizar los errores de clasificación; valores más altos pueden reducir los errores en el conjunto de entrenamiento, pero incrementan el riesgo de sobreajuste. El coeficiente del núcleo (gamma) determina la influencia de cada muestra de entrenamiento sobre la frontera de decisión, afectando directamente la complejidad del modelo; valores altos tienden a generar fronteras más ajustadas a los datos. Finalmente, el tipo de núcleo especifica la función de transformación utilizada para proyectar los datos en un espacio de mayor dimensión, siendo las variantes lineal y RBF las evaluadas en este estudio.

Tabla 5.5: Espacio de búsqueda de hiperparámetros usados en SVM

Hiperparámetro	Rango de búsqueda
Parámetro de regularización	[1, 10, 20, 30, 50]
Coficiente del núcleo (gamma)	[0.01, scale]
Tipo de núcleo (kernel)	[rbf, linear]

La Tabla 5.6 resume la mejor configuración de hiperparámetros obtenida para SVM según el tipo de variables de entrada, mientras que la Tabla 5.7 muestra el desempeño alcanzado por el modelo en las tres configuraciones. Los resultados detallados por partición se presentan en el Anexo B.

Tabla 5.6: Mejor configuración de hiperparámetros en SVM por tipo de entrada

Hiperparámetro	Bandas espectrales (B1–B7)	Índices espectrales (NDVI, NDBI, NDWI)	Combinación (Bandas + Índices)
Parámetro de regularización	30	1	30
Coefficiente del núcleo (gamma)	scale	scale	scale
Tipo de núcleo (kernel)	rbf	rbf	rbf

Tabla 5.7: Comparativa de desempeño de SVM según tipo de variables de entrada

Métricas	Bandas espectrales (B1–B7)	Índices espectrales (NDVI, NDBI, NDWI)	Combinación (Bandas + Índices)
Tiempo total (s)	294.02	227.35	262.04
Accuracy	0.9450	0.9293	0.9293
Precision	0.9452	0.9298	0.9476
Recall	0.9450	0.9293	0.9473
F1-score	0.9450	0.9293	0.9473

5.1.7. Clasificación mediante regresión logística

La regresión logística se implementó como modelo base para la clasificación binaria, con el propósito de establecer un punto de comparación frente a los algoritmos no lineales. Su entrenamiento y validación siguieron el mismo procedimiento metodológico, garantizando condiciones equivalentes de evaluación.

La Tabla 5.8 presenta el espacio de búsqueda de hiperparámetros utilizado para este modelo. El inverso de la regularización controla el grado de penalización aplicado a los coeficientes; valores grandes reducen la regularización, permitiendo modelos más complejos que se ajusten mejor a los datos de entrenamiento. El algoritmo de optimización define el método empleado para encontrar los parámetros que maximizan la verosimilitud del modelo, siendo lbfgs y saga los evaluados por su eficiencia en problemas de alta dimensionalidad. El tipo de regularización, en este caso L2, actúa como un factor de control para evitar el sobreajuste al penalizar coeficientes grandes. Finalmente, el número máximo de iteraciones establece el límite de pasos que el optimizador puede ejecutar para lograr la convergencia.

Tabla 5.8: Espacio de búsqueda de hiperparámetros usados en regresión logística

Hiperparámetro	Rango de búsqueda
Inverso de la regularización	[0.01, 0.1, 1, 10, 100]
Algoritmo de optimización	[lbfgs, saga]
Tipo de regularización	[l2]
Iteraciones máximas	[500, 1000]

La Tabla 5.9 presenta la mejor configuración de hiperparámetros obtenida para regresión logística según el tipo de variables de entrada, mientras que la Tabla 5.10 muestra el desempeño del modelo de las tres configuraciones. Los resultados detallados por partición se incluyen en el Anexo C.

Tabla 5.9: Mejor configuración de hiperparámetros en regresión logística por tipo de entrada

Hiperparámetro	Bandas espectrales (B1–B7)	Índices espectrales (NDVI, NDBI, NDWI)	Combinación (Bandas + Índices)
Inverso de la regularización	100	100	100
Algoritmo de optimización	saga	lbfgs	lbfgs
Tipo de regularización	l2	l2	l2
Iteraciones máximas	500	1000	1000

Tabla 5.10: Comparativa de desempeño de Regresión Logística según tipo de variables de entrada

Métricas	Bandas espectrales (B1–B7)	Índices espectrales (NDVI, NDBI, NDWI)	Combinación (Bandas + Índices)
Tiempo total (s)	11.11	4.06	16.87
Accuracy	0.9347	0.9267	0.9363
Precision	0.9347	0.9267	0.9363
Recall	0.9347	0.9267	0.9363
F1-score	0.9347	0.9267	0.9363

5.1.8. Clasificación mediante XGBoost

El modelo XGBoost se implementó para aprovechar su capacidad de aprendizaje secuencial mediante boosting de árboles de decisión, optimizando el rendimiento del modelo a través de la corrección iterativa de errores. Su configuración siguió el mismo esquema de validación y evaluación empleado en los demás clasificadores.

La Tabla 5.10 presenta el espacio de búsqueda utilizado para la optimización de hiperparámetros del modelo. El número de árboles hace referencia a la cantidad total de iteraciones o modelos individuales que se combinan en el ensamble final, lo que influye tanto en la precisión como en el tiempo de entrenamiento. La profundidad máxima de cada árbol limita la complejidad de las interacciones que se pueden modelar, evitando que el modelo se ajuste excesivamente al conjunto de entrenamiento. La tasa de aprendizaje regula la contribución de cada árbol en el proceso de ajuste, siendo valores pequeños más favorables para la generalización, pero requiriendo un mayor número de iteraciones. La proporción de muestras define qué fracción del conjunto de datos se utiliza para entrenar cada árbol, lo que introduce aleatoriedad y ayuda a prevenir el sobreajuste. Por su parte, la fracción de columnas determina la proporción de variables consideradas en cada división, favoreciendo la diversificación de los árboles y mejorando la robustez del modelo.

Tabla 5.11: Espacio de búsqueda de hiperparámetros usados en XGBoost

Hiperparámetro	Rango de búsqueda
Número de árboles	[100, 150, 200]
Profundidad máxima de cada árbol	[3, 6, 10]
Tasa de aprendizaje	[0.01, 0.1, 0.2]
Proporción de muestras	[0.6, 0.8, 1.0]
Fracción de columnas	[0.6, 0.8, 1.0]

La Tabla 5.12 presenta la mejor configuración de hiperparámetros obtenida para XGBoost según el tipo de variables de entrada, mientras que la Tabla 5.13 muestra el desempeño del modelo en las tres configuraciones. Los resultados detallados por partición se incluyen en el Anexo D.

Tabla 5.12: Mejor configuración de hiperparámetros en XGBoost por tipo de entrada

Hiperparámetro	Bandas espectrales (B1–B7)	Índices espectrales (NDVI, NDBI, NDWI)	Combinación (Bandas + Índices)
Número de árboles	200	150	150
Profundidad máxima de cada árbol	10	3	6
Tasa de aprendizaje	0.1	0.1	0.01
Proporción de muestras	1.0	1.0	0.8
Fracción de columnas	1.0	0.8	1.0

Tabla 5.13: Comparativa de desempeño de XGBoost según tipo de variables de entrada

Métricas	Bandas espectrales (B1–B7)	Índices espectrales (NDVI, NDBI, NDWI)	Combinación (Bandas + Índices)
Tiempo total (s)	36.84	18.27	44.56
Accuracy	0.9343	0.9297	0.9403
Precision	0.9343	0.9300	0.9404
Recall	0.9343	0.9297	0.9403
F1-score	0.9343	0.9297	0.9403

5.1.9. Comparativa entre métodos supervisados

En todos los casos analizados, la inclusión de índices espectrales NDVI, NDBI y NDWI mejoró consistentemente las métricas de desempeño de los modelos, evidenciando que la combinación de bandas espectrales con estos índices derivados aporta información complementaria, que favorece la discriminación entre áreas urbanas y no urbanas. Como se observa en la Tabla 5.14, los modelos evaluados lograron sus mejores resultados utilizando la combinación de bandas e índices, mostrando altos valores de F1-score. En general, todos demostraron una buena capacidad para clasificar la cobertura del suelo a partir de estas variables. Random Forest destaca como un modelo que combina de manera efectiva la interpretabilidad, rendimiento predictivo y capacidad para procesar datos de alta dimensionalidad. Su elevado desempeño,

reflejado en un F1-score de 0.9423, junto con su capacidad para cuantificar la importancia relativa de las variables predictoras, lo posicionan como una buena alternativa en aplicaciones donde la precisión y la transparencia del modelo son relevantes.

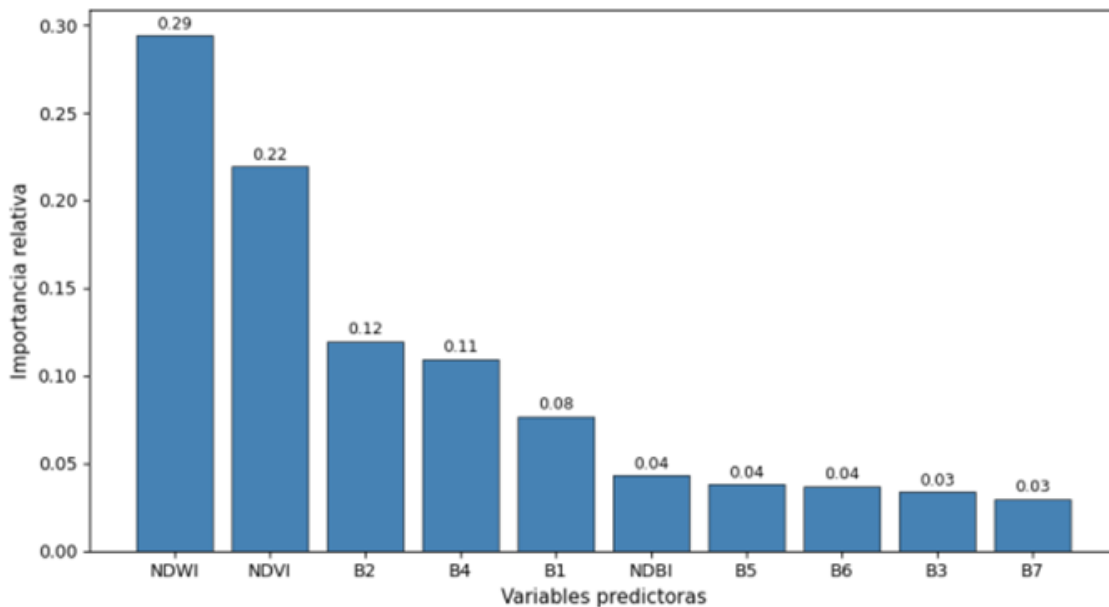


Figura 5.2: Importancia relativa de las variables predictoras en el modelo Random Forest (entrada: bandas+índices).

La Figura 5.2 muestra que el índice NDWI es la variable que más contribuye al modelo, seguido por el NDVI y la banda B2 (azul). En Random Forest, la importancia de una variable se mide por cuánto ayuda a reducir la impureza en los nodos del árbol al momento de tomar decisiones de clasificación. Cuanto mayor sea esta reducción, promediada en todos los árboles del bosque, más influyente resulta la variable para distinguir entre áreas urbanas y no urbanas. Este resultado confirma que los índices espectrales derivados, como el NDWI y el NDVI, aportan información altamente relevante en este contexto.

No obstante, esta ventaja conlleva un mayor tiempo de procesamiento en relación con métodos más simples, como la regresión logística.

Tabla 5.14: Comparativa de mejores modelos según su F1 ponderado

Modelo	Mejor caso (variables)	F1-ponderado	Tiempo [s]
Random Forest	Bandas + Índices	0.9423	255.33
SVM	Bandas + Índices	0.9473	262.04
Regresión logística	Bandas + Índices	0.9363	16.87
XGBoost	Bandas + Índices	0.9403	44.56

5.2. Métodos no supervisados de umbralización

Para validar los resultados de los modelos supervisados, se aplicaron distintos métodos de umbralización sobre el índice NDBI, específicamente los propuestos por Otsu, Li y Lee y Yen, Chang y Chang. Estos métodos permiten establecer de manera automática un valor de corte para diferenciar entre entornos urbanos y no urbanos. Al ser procedimientos no supervisados, clasifican las coberturas de forma rápida y sin requerir entrenamiento previo.

Para su aplicación, se utilizaron los mismos conjuntos de prueba empleados en los modelos supervisados, generando predicciones a partir de los umbrales calculados y evaluando sus resultados frente a las etiquetas reales. Esto permitió realizar una comparación directa con los modelos de aprendizaje automático.

La Tabla 5.15 presenta la comparativa entre el método seleccionado (Random Forest) y los métodos de umbralización del NDBI, mostrando que todos ellos tienen un desempeño similar respecto a su F1-Score, aunque Random Forest destaca por obtener un rendimiento superior respecto a los métodos clásicos, lo cual evidencia su capacidad para identificar correctamente la cobertura del suelo en la ciudad, aunque con un tiempo de procesamiento notablemente mayor.

Tabla 5.15: Comparativa de métodos de umbralización vs Random Forest

Método	F1-score	Tiempo [s]
Random Forest	0.9423	255.33
Otsu	0.8375	0.0010
Li y Lee	0.8157	0.0013
Yen, Chang y Chang	0.8390	0.0003

5.3. Reclasificación de coberturas

A pesar de las métricas reportadas por los modelos de clasificación, es importante resaltar sus limitaciones, como los errores en áreas de transición espectral entre clases y las fluctuaciones año a año que no siempre reflejan la dinámica real del crecimiento urbano. Dado que se trata de una serie temporal, resulta necesario implementar una etapa de reclasificación posterior que permita corregir inconsistencias y estabilizar los mapas de clasificación generados anualmente a partir de la clasificación inicial.

La reclasificación multitemporal y espacial consiste en aplicar un conjunto de reglas y filtros que reevalúan la clase asignada a cada píxel, no como un valor aislado en el espacio y en el tiempo, sino como parte de una trayectoria o patrón de comportamiento. La metodología

aplicada está basada en el enfoque propuesto por la iniciativa MapBiomass que demostró aumentar la precisión y consistencia en la generación de mapas multianuales en varios países de Latino América, garantizando que dichos cambios sean coherentes con las dinámicas propias de cada región [83].

A continuación, se presentan los filtros post-clasificación aplicados a la cobertura de las UPZ de la ciudad de Bogotá D.C de manera jerárquica.

- Filtro espacial: Se aplicó una operación morfológica de apertura binaria para eliminar manchas pequeñas e inconsistentes en la clase urbana. Esto reduce el ruido espacial y asegura que solo agrupaciones significativas de píxeles sean consideradas como dentro de esta clase
- Gap fill (relleno de huecos): Se rellenaron valores vacíos restantes utilizando una función de vecindad 3x3. El valor del píxel central fue reemplazado por la clase más frecuente entre sus vecinos.
- Filtro de ventana temporal: Se utilizó una ventana móvil de tres años (año anterior, actual y siguiente). Si al menos dos de los tres años clasificaban un píxel como urbano, se mantuvo esa clase para el año actual. Este enfoque suaviza variaciones abruptas debidas a errores o condiciones atmosféricas, y refuerza la consistencia entre años consecutivos.
- Filtro de consistencia temporal (persistencia urbana): Si un píxel fue urbano en el año anterior, se mantiene como urbano en el año actual, incluso si la clasificación actual indica lo contrario. Esto evita cambios abruptos e inconsistentes en áreas ya urbanizadas.

6. Evaluación el modelo escogido y caracterización de las variables relacionadas

A continuación, se presentan los resultados del proceso de post-procesamiento aplicado a la clasificación urbana, así como la evolución temporal del área urbana en Bogotá D.C. entre 2014 y 2024. Luego, se presenta una caracterización de los distintos componentes que se han analizado, para entender la relación entre la expansión y el desarrollo urbano, junto con el porcentaje de personas en condición de pobreza. Los factores para analizar son: expansión, urbanización, población, pobreza y densidad poblacional.

La Tabla 6.1 presenta el impacto del post-procesamiento en la estimación del área urbana para cada año. En la mayoría de los casos, se observa un incremento del área urbana tras aplicar las reglas de coherencia espacial y temporal, con aumentos que alcanzan los 23.0 km² en 2021. Este comportamiento refleja la corrección de errores del clasificador original, que tendía a subestimar la cobertura urbana al fragmentar áreas continuas o clasificar erróneamente píxeles urbanos. La única reducción (-1.5 km² en 2016) corresponde a la eliminación de falsos positivos, mostrando que el post-procesamiento no solo agrega superficie urbana, sino que también depura y refina la información, mejorando así la consistencia y credibilidad de la serie temporal.

Tabla 6.1: Impacto del post-procesamiento en la estimación del área urbana

Año	Antes (km²)	Después (km²)	Δ (km²)
2014	255.39	261.57	6.18
2015	259.80	265.33	5.53
2016	269.57	268.04	-1.53
2017	251.03	268.90	17.87
2018	258.39	270.86	12.47
2019	266.87	272.76	5.89
2020	254.59	273.47	18.89
2021	251.19	274.18	22.99
2022	262.62	275.81	13.19
2023	260.77	276.79	16.02
2024	264.93	279.55	14.62

La Tabla 6.2 presenta la evolución de las áreas clasificadas como urbano y no urbano en el total de las UPZ de Bogotá D.C., luego del proceso de reclasificación. Además, se incluye el porcentaje de crecimiento urbano interanual y el porcentaje del área urbanizada con respecto al total, lo cual permite observar de manera consolidada los cambios en la cobertura a lo largo del periodo analizado. En general, se evidencia una expansión progresiva del área urbana, con un aumento acumulado de 6.9% entre 2014 y 2024, reflejando una tendencia sostenida

de urbanización que alcanza el 67.01 % del territorio al final del periodo.

Tabla 6.2: Dinámica de Urbanización y Crecimiento Urbano Anual (2014–2024)

Año	Área Urbana (km²)	Área No Urbana (km²)	Crecimiento Urbano (%)	% de Urbanización
2014	261.6	155.6	0.00	62.70
2015	265.3	151.9	1.44	63.60
2016	268.0	149.1	1.02	64.25
2017	268.9	148.3	0.32	64.46
2018	270.9	146.3	0.73	64.92
2019	272.8	144.4	0.70	65.38
2020	273.5	143.7	0.26	65.55
2021	274.2	143.0	0.26	65.72
2022	275.8	141.4	0.59	66.11
2023	276.8	140.4	0.36	66.35
2024	279.5	137.6	1.00	67.01

La Figura 6.1 presenta la dinámica del cambio urbano entre 2014 y 2024 luego del post procesamiento, resaltando las áreas que transitaron de no urbano a urbano (en rojo) y las zonas urbanas que se mantuvieron estables (en gris). Se evidencia que la expansión urbana se ha concentrado principalmente en los bordes periféricos y en sectores de transición dentro de las UPZ, reflejando un patrón de crecimiento caracterizado por la dispersión y la fragmentación espacial.

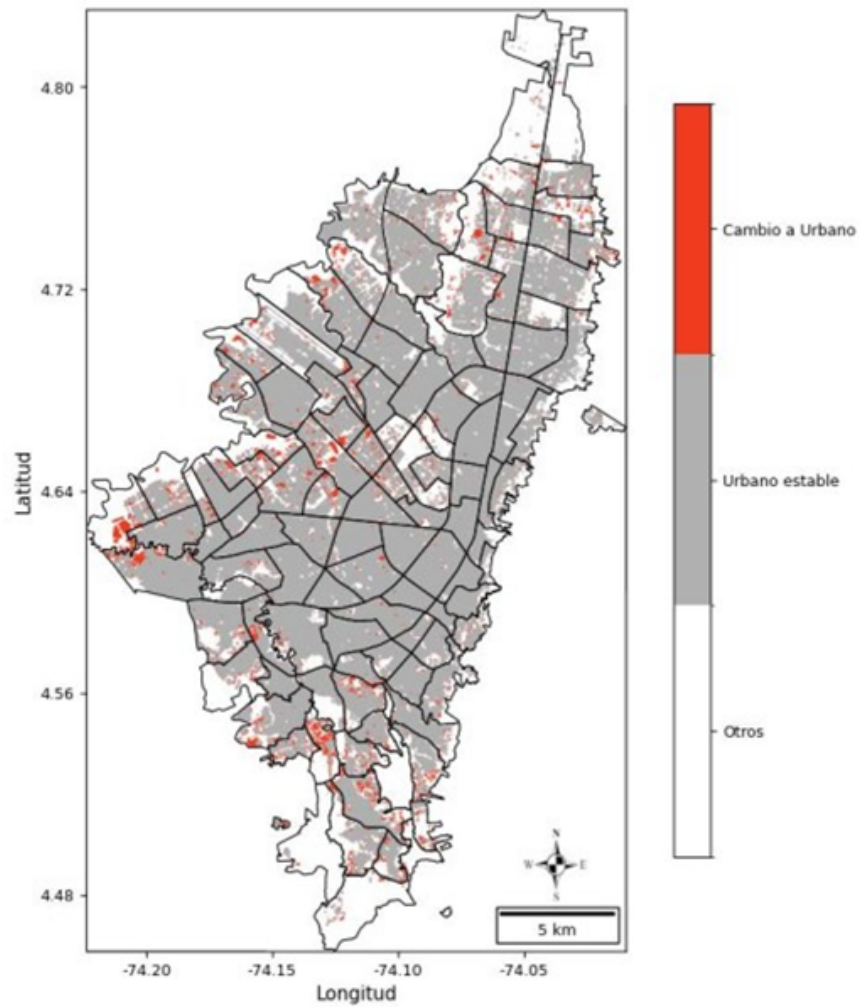


Figura 6.1: Cambios en el área urbana 2014-2024.

A continuación, la Figura 6.2 contiene el diagrama de flujo que explica el proceso que se realizó para entender las caracterizaciones de las variables, las técnicas usadas para evaluar posibles relaciones y un análisis profundo de las causas y estrategias de afrontamiento para las situaciones propias de cada sector de la ciudad de Bogotá D.C.

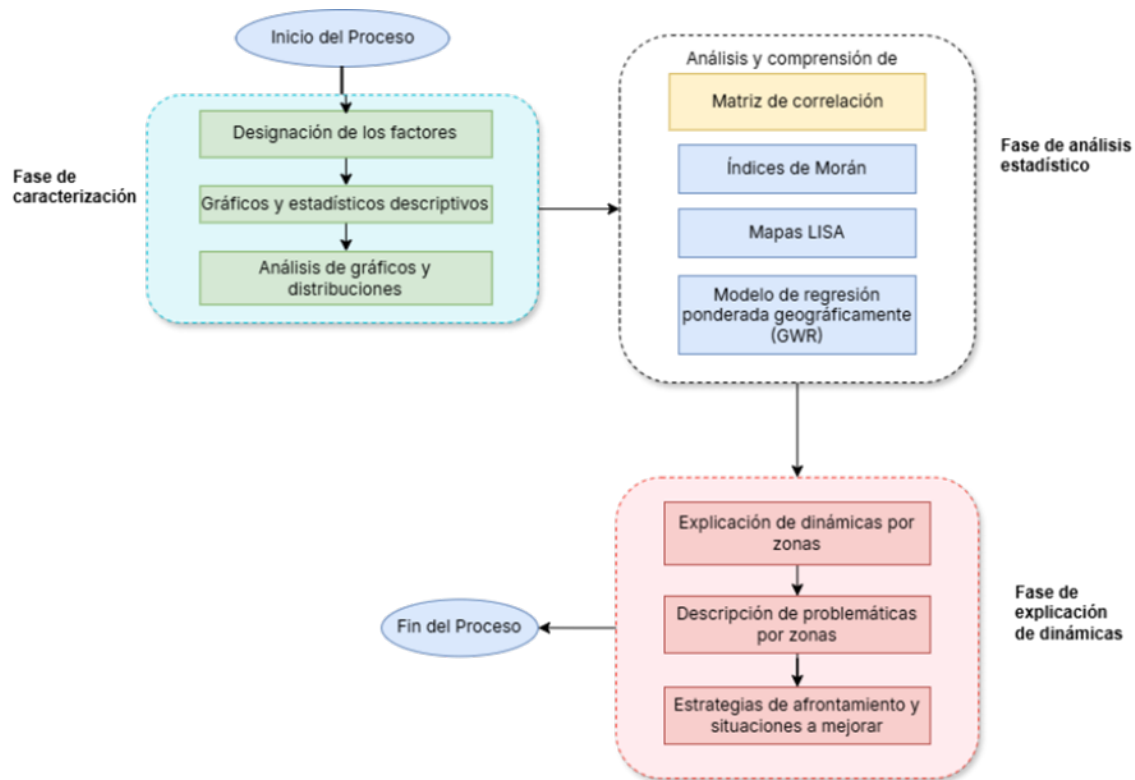


Figura 6.2: Diagrama de flujo metodológico para evaluar la caracterización de las relaciones entre las variables.

6.1. Caracterizaciones

6.1.1. Expansión

La Figura 6.3 muestra la proporción de expansión, el cual revela una distribución bastante asimétrica. La mayoría de las UPZ están registrando valores muy cercanos a cero, lo que sugiere que, en gran parte del territorio de Bogotá D.C, la expansión urbana ha sido mínima o incluso inexistente. Con algunos valores atípicos altos, como uno extremo de 0.84, demuestran que unas pocas áreas han experimentado una expansión urbana desproporcionada en el periodo de tiempo analizado.

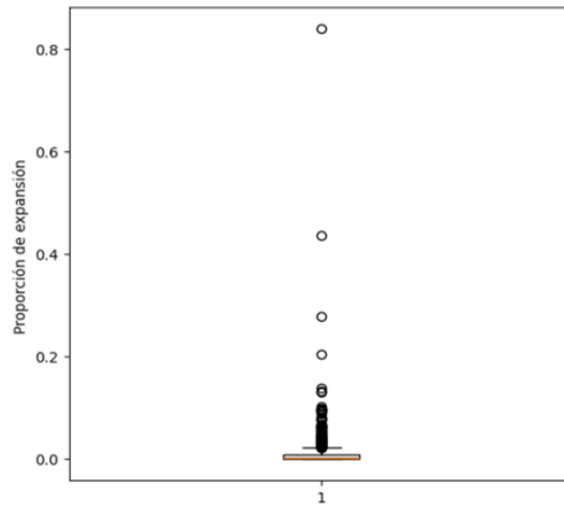


Figura 6.3: Diagrama de caja de proporción de expansión por UPZ en Bogotá D.C. (2014 – 2024) (Con 2014 como año base).

6.1.2. Urbanización

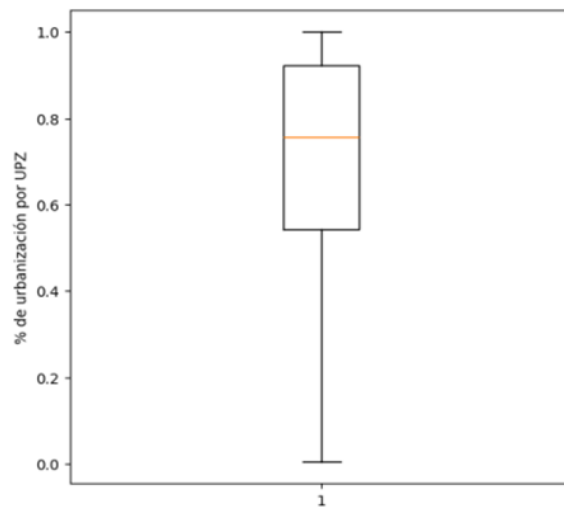


Figura 6.4: Diagrama de caja de % de urbanización por UPZ en Bogotá D.C. (2014 – 2024).

La Figura 6.4 que muestra el porcentaje de urbanización por UPZ, el cual revela una distribución más equilibrada en comparación con la proporción de expansión, aunque hay una variabilidad notable entre las diferentes unidades territoriales. La mediana se sitúa en 0.75, mostrando que al menos la mitad de las UPZ están urbanizadas en más de un 75 %, reflejando así una ciudad que en su mayoría está consolidada. No se observan valores extremos

inferiores fuera del rango esperado (no hay atípicos marcados), lo que sugiere que los casos de baja urbanización son pocos, pero no extremos.

6.1.3. Población

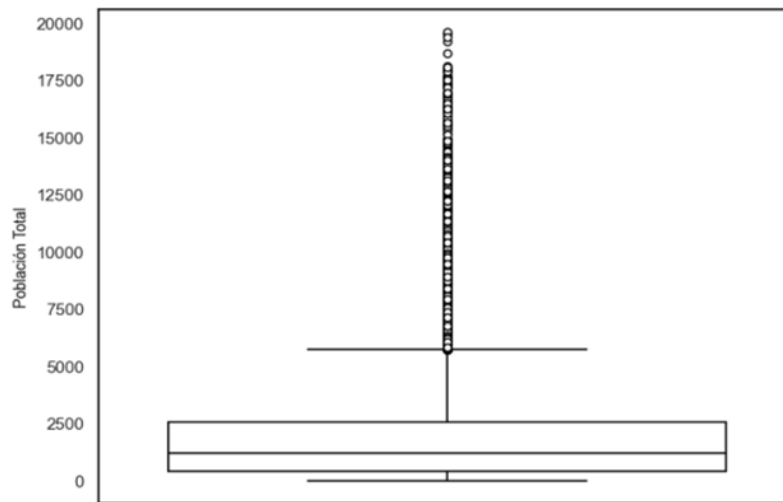


Figura 6.5: Diagrama de caja de la población total en Bogotá D.C. (2014 – 2024).

La Figura 6.5 muestra una distribución que está sesgada hacia la derecha, con un notable número de valores atípicos que superan el límite superior. La mediana se encuentra en 1.209 habitantes, lo que sugiere que al menos la mitad de las UPZ tienen poblaciones relativamente bajas. El valor máximo registrado alcanza los 19.611 habitantes, lo que indica una concentración poblacional significativa en una o pocas UPZ específicas.

6.1.4. Pobreza

La Figura 6.6 revela una tendencia general hacia la reducción de la pobreza, junto con una disminución en la dispersión y en la cantidad de outliers a lo largo del tiempo. En 2014, la mediana del porcentaje de pobreza era de aproximadamente 6.44, con un valor máximo de 26.83, lo que indica una distribución amplia con marcadas desigualdades y varias UPZ en condiciones críticas. Para 2017, la mediana, que bajó a 4.05, y el rango intercuartílico (entre $Q1 = 1.76$ y $Q3 = 6.76$) sugieren una mejora general en las condiciones de vida, aunque todavía hay algunas áreas notablemente más pobres. Finalmente, en 2021 la mediana siguió descendiendo a 3.67 y tuvo un valor máximo de 21.38, lo que evidencia no solo una mejora en los niveles medios de pobreza, sino también una reducción progresiva de los casos extremos.

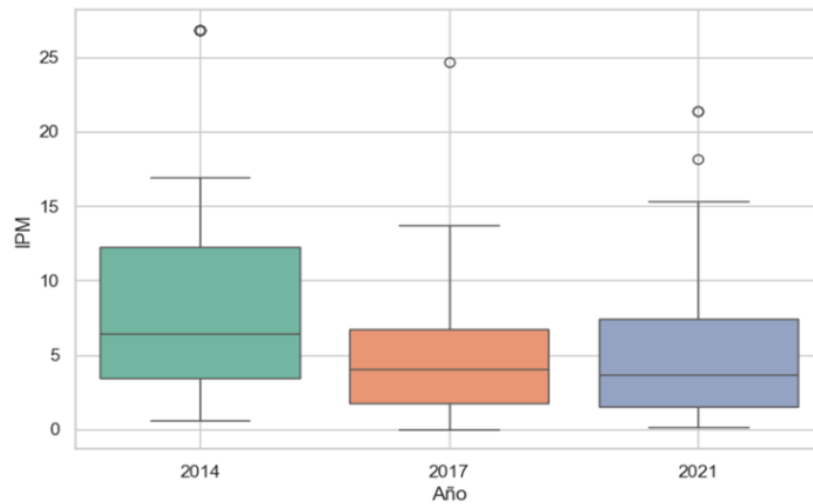


Figura 6.6: Diagrama de caja del IPM por año.

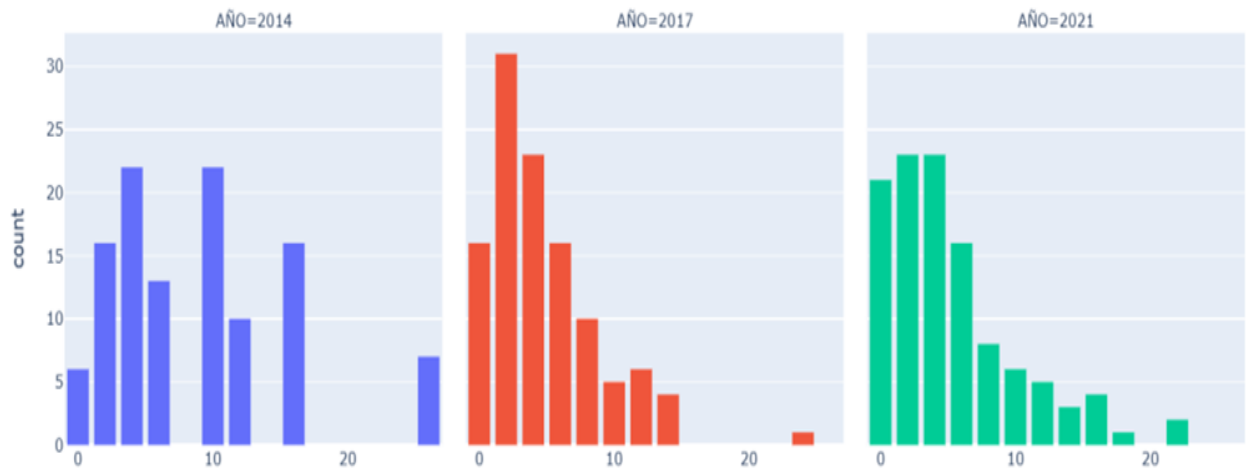


Figura 6.7: Distribución del IPM para los años 2014, 2017 y 2021.

La Figura 6.7 que muestra la distribución del porcentaje de pobreza a lo largo de los años refuerza y complementa el análisis que se han desarrollado con los diagramas de caja, ofreciendo una representación visual del cambio en la pobreza entre 2014 y 2021. En 2014, la distribución era más dispersa y multimodal, con un número considerable de UPZ que presentaban niveles medios y altos de pobreza. Para 2017, se nota un claro desplazamiento hacia la izquierda, con una mayor concentración de UPZ en niveles bajos de pobreza (entre 0 y 5), lo que indica una mejora en las condiciones generales. En 2021, esta tendencia se mantiene, con una distribución aún más inclinada hacia la izquierda, aunque con una cola más larga que en 2017, lo que sugiere que, aunque la mayoría de las UPZ tienen bajos niveles de pobreza, se mantuvieron o se agudizaron algunas áreas con altos valores de pobreza.

6.1.5. Densidad

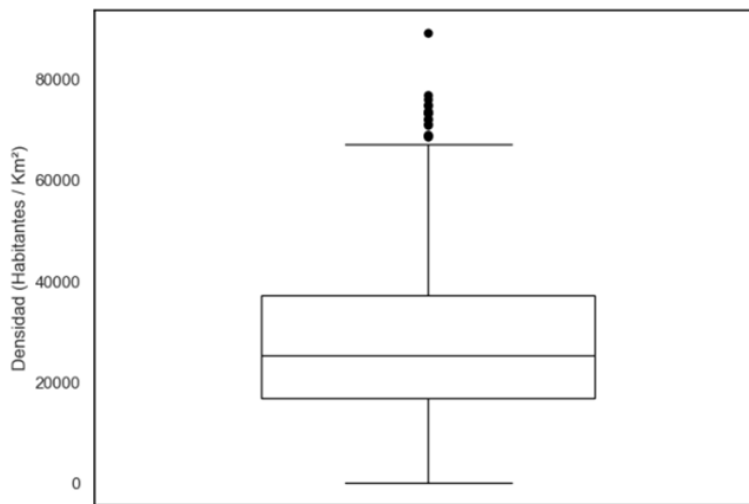


Figura 6.8: Diagrama de caja de densidad poblacional

La Figura 6.8 revela una distribución claramente asimétrica hacia la derecha. La mediana se sitúa alrededor de 25.000 habitantes por kilómetro cuadrado, lo que indica que la mitad de las UPZ analizadas tienen densidades iguales o inferiores a este valor. El rango intercuartílico, que abarca el 50 % central de las observaciones, va desde aproximadamente 15.000 hasta 37.000 hab/km², lo que refleja una variabilidad moderada en la mayoría de las áreas. El límite inferior del rango no atípico se encuentra cerca de 2.000 hab/km², mientras que el superior llega a 70.000 hab/km². Más allá de este umbral, se identifican varios valores atípicos entre 70.000 y 80.000 hab/km², así como un caso extremo que supera los 80.000 hab/km², lo que sugiere que hay UPZ con niveles de concentración poblacional notablemente más altos que el resto.

6.2. Métodos de análisis estadístico

6.2.1. Análisis de correlación

A continuación, se mostrarán algunos gráficos de correlación sobre las variables analizadas, con el objetivo de entender mejor sus comportamientos.

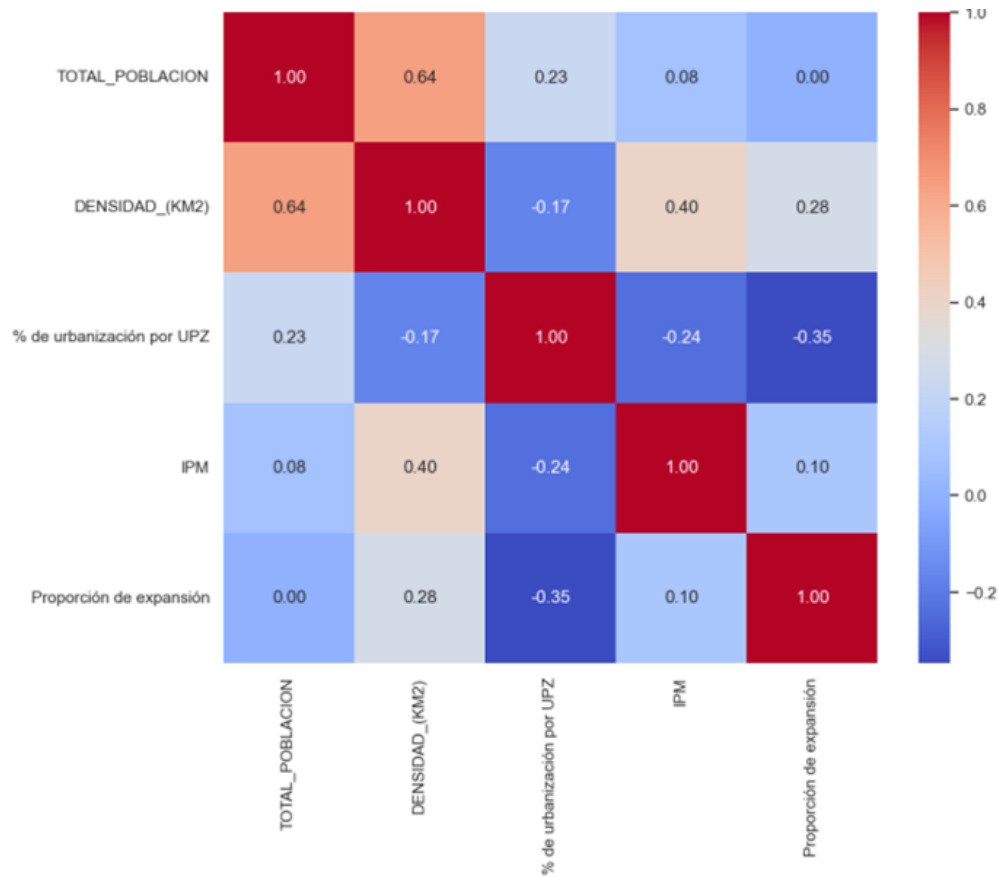


Figura 6.9: Matriz de correlación de variables seleccionadas.

La Figura 6.9 evidencia como las dinámicas urbanas en Bogotá D.C. en el periodo de estudio presentan relaciones diferenciadas. Se puede observar que la población total muestra una fuerte asociación con la densidad urbana (0.64), lo cual refleja la concentración de habitantes en áreas reducidas, mientras que su vínculo con el grado de urbanización es más débil (0.23), indicando que el crecimiento demográfico no siempre se traduce en mayor consolidación urbana. Destaca la correlación positiva moderada entre la densidad urbana y el IPM (0.40), lo que sugiere que en las zonas más compactas se concentra una mayor vulnerabilidad. En contraste, el porcentaje de urbanización presenta una relación negativa débil con la pobreza (-0.24), dando a entender que las UPZ con mayor consolidación urbana tienden a presentar mejores condiciones de vida. Por otra parte, se puede observar que la expansión urbana tiene un vínculo inverso con la urbanización (-0.35), lo que confirma la tensión entre el crecimiento periférico y consolidación de la ciudad; sin embargo, su relación con la pobreza es prácticamente nula (0.10), indicando que los procesos de expansión pueden ocurrir en distintos tipos de entornos urbanos, sin un patrón socioeconómico social definido.

6.3. Índices de Morán

Tabla 6.3: Evolución del índice de Moran global

Año	% de urbanización por UPZ	Densidad	IPM	Proporción de expansión	Total de la población
2014	0.43	0.48	0.62	-	0.32
2017	0.40	0.53	0.53	0.17	0.33
2021	0.39	0.49	0.61	0.19	0.34

La Tabla 6.3 presenta cómo ha evolucionado el índice de Moran global en distintas variables durante los años 2014, 2017 y 2021. Para el caso del IPM, muestra consistentemente el nivel más alto de autocorrelación espacial en comparación con las otras variables. Su valor, que era superior a 0.6 en 2014, disminuyó en 2017 a poco más de 0.5, para luego volver a incrementarse en 2021, alcanzando nuevamente un valor por encima de 0.6. Esto refleja que el agrupamiento geográfico de la pobreza es el más fuerte de todos los fenómenos medidos, aunque su intensidad fluctuó durante el período.

La variable de población total presentó valores más estables, iniciando en 0.32 en 2014, aumentando ligeramente a 0.33 en 2017, y llegando a 0.34 en 2021. Esta tendencia ascendente, aunque moderada, refleja un aumento en la homogeneidad en la distribución poblacional, con áreas densamente pobladas ubicándose cerca de otras similares. El incremento sostenido indica que el crecimiento demográfico no ha sido uniforme, sino que ha favorecido ciertos sectores del territorio, consolidando núcleos urbanos ya existentes.

En cuanto a la proporción de expansión urbana, los valores oscilaron entre 0.17 en 2017 y 0.19 en 2021. La persistencia de autocorrelación positiva sugiere que los procesos de expansión siguen patrones espaciales definidos, donde el crecimiento tiende a extenderse desde áreas ya desarrolladas hacia zonas contiguas, lo cual es común en contextos de desarrollo urbano incremental.

El porcentaje de urbanización por Unidad de Planeación Zonal (UPZ) mostró una caída progresiva: de 0.42 en 2014, a 0.40 en 2017, y finalmente a 0.39 en 2021. Esta disminución constante apunta a una reducción en la concentración espacial de la urbanización, lo que podría interpretarse como una mayor dispersión del proceso de transformación territorial.

Por último, la densidad poblacional por kilómetro cuadrado presentó uno de los niveles estables de agrupamiento espacial en comparación con las demás. Su evolución a lo largo del tiempo es fluctuante: partió de un valor de Moran cercano a 0.49 en 2014, experimentó un aumento hasta aproximadamente 0.52 en 2017 y, finalmente, descendió en 2021 a un nivel de 0.49, similar al del inicio del período. Este comportamiento indica que la tendencia de

las áreas con densidades de población similares a agruparse geográficamente se intensificó a mitad del período antes de volver a su nivel original.

6.4. Mapas LISA

Al aplicar esta técnica al IPM en los años 2014, 2017 y 2021, se revelan transformaciones espaciales significativas en la distribución de la vulnerabilidad social. En el anexo E, se encuentra el repositorio, en el cual está almacenada la matriz de pesos espaciales para los mapas LISA, que se usan para representar matemáticamente las relaciones de proximidad o influencia entre las unidades espaciales

6.4.1. IPM

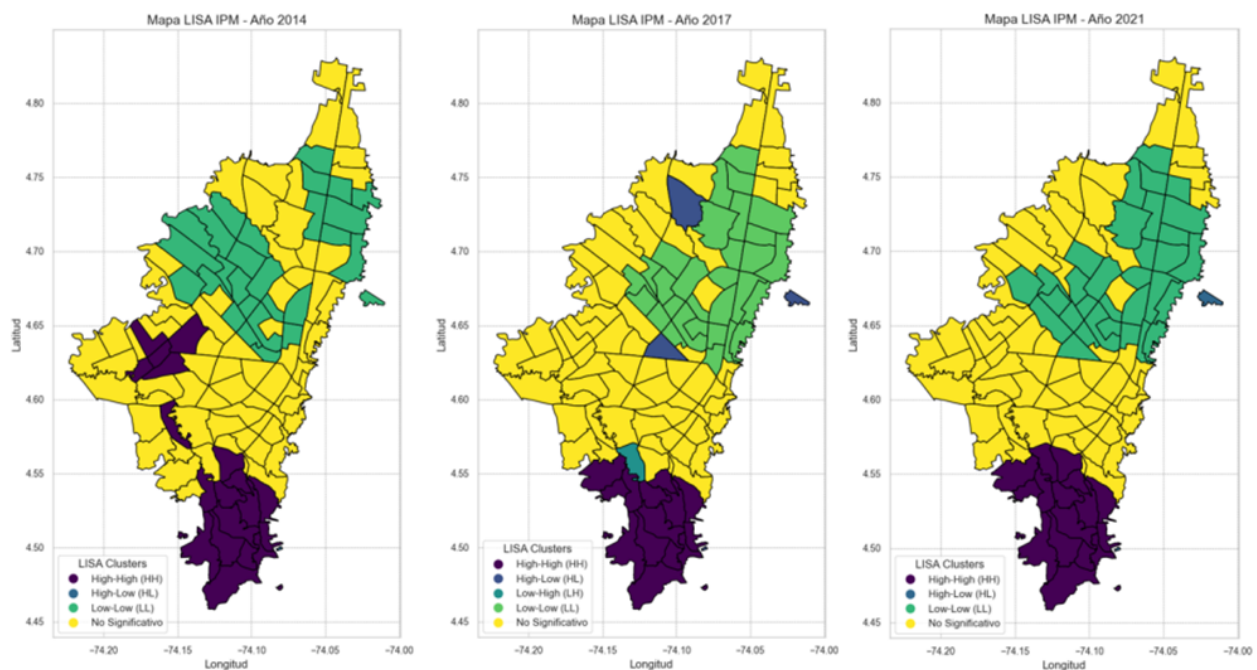


Figura 6.10: Evolución Espaciotemporal de los Clústeres LISA del IPM (2014-2021).

La Figura 6.10 muestra que, de manera general, el sur y sureste concentran históricamente núcleos de alta pobreza (High-High), mientras que el norte y centro se caracterizan por entornos más favorables (Low-Low). Entre ambos polos aparecen zonas de transición (High-Low y Low-High) y sectores sin autocorrelación significativa, que reflejan dinámicas más inestables.

En 2014 se aprecia un patrón muy marcado: el sur y sureste muestran extensas áreas High-High, consolidando núcleos de exclusión social con fuerte concentración de pobreza, mientras el norte y centro presentan grandes superficies Low-Low, asociadas con condiciones de vida estables y mejores oportunidades. Las categorías de transición apenas aparecen, lo que indica que los contrastes abruptos entre territorios vecinos eran poco comunes; en algunos casos aislados se observan Low-High, es decir, “islas” de prosperidad rodeadas por contextos más pobres. Gran parte del territorio intermedio queda clasificado como no significativo, reflejando heterogeneidad sin una tendencia espacial clara.

Para 2017 se produce un cambio hacia una mayor fragmentación. Los clústeres High-High del sur y sureste reducen levemente su extensión, lo que sugiere una dispersión parcial de la pobreza, mientras las áreas Low-Low en el norte y centro también disminuyen. Destaca el aumento de las categorías de transición: High-Low, que evidencian la presencia de núcleos de pobreza rodeados por vecindarios más favorecidos, y Low-High, que muestran zonas relativamente estables siendo cercadas por áreas en deterioro. Al mismo tiempo, las zonas no significativas se redistribuyen, concentrándose más en el centro y norte, donde antes predominaba la baja pobreza.

En 2021 se recompone el panorama con un giro hacia la reconcentración de la pobreza. Los clústeres High-High se intensifican y recuperan terreno en el sur y sureste, lo que coincide con el incremento del índice de Moran global y señala un fortalecimiento de la autocorrelación positiva. Las áreas Low-Low se reducen de forma considerable, sobre todo en el norte, lo que evidencia un retroceso en la extensión de territorios con baja pobreza. Las categorías de transición disminuyen respecto a 2017, lo que sugiere que los contrastes extremos entre áreas vecinas pierden importancia frente a la polarización más amplia entre bloques de alta y baja pobreza. Las zonas no significativas se mantienen, aunque ahora aparecen más focalizadas en regiones intermedias del centro y norte, donde la dinámica social es más cambiante.

6.4.2. Densidad

La Figura 6.11 muestra la autocorrelación espacial local de la densidad, el año 2014, las áreas clasificadas como High-High se encuentran principalmente en el centro y sur del territorio, con una notable presencia también en el oeste y sureste. Estas zonas corresponden a unidades con una alta densidad poblacional, rodeadas por otras con características similares. Por otro lado, las áreas Low-Low aparecen en el norte y este, lo que indica regiones con baja densidad que están rodeadas por entornos con niveles comparables. La mayor parte del área geográfica se clasifica como No Significativo, lo que sugiere que en la mayoría de las UPZs no hay una relación espacial estadísticamente significativa con sus vecinas.

Para 2017, la distribución de los clústeres muestra cambios notables. Los agrupamientos High-High han reducido su extensión y se han concentrado más en el centro y sureste, adoptando una forma más compacta. Las zonas Low-Low han ganado terreno en el norte y este, extendiéndose hacia áreas centrales. El número de unidades clasificadas como No Significativo ha disminuido ligeramente, especialmente en el sur, donde comienzan a aparecer puntos aislados de tipo High-Low. Estas áreas tienen alta densidad rodeadas por zonas de baja densidad, lo que introduce una mayor heterogeneidad en la estructura espacial. También se pueden observar algunas unidades dispersas de tipo Low-High, aunque en menor cantidad.

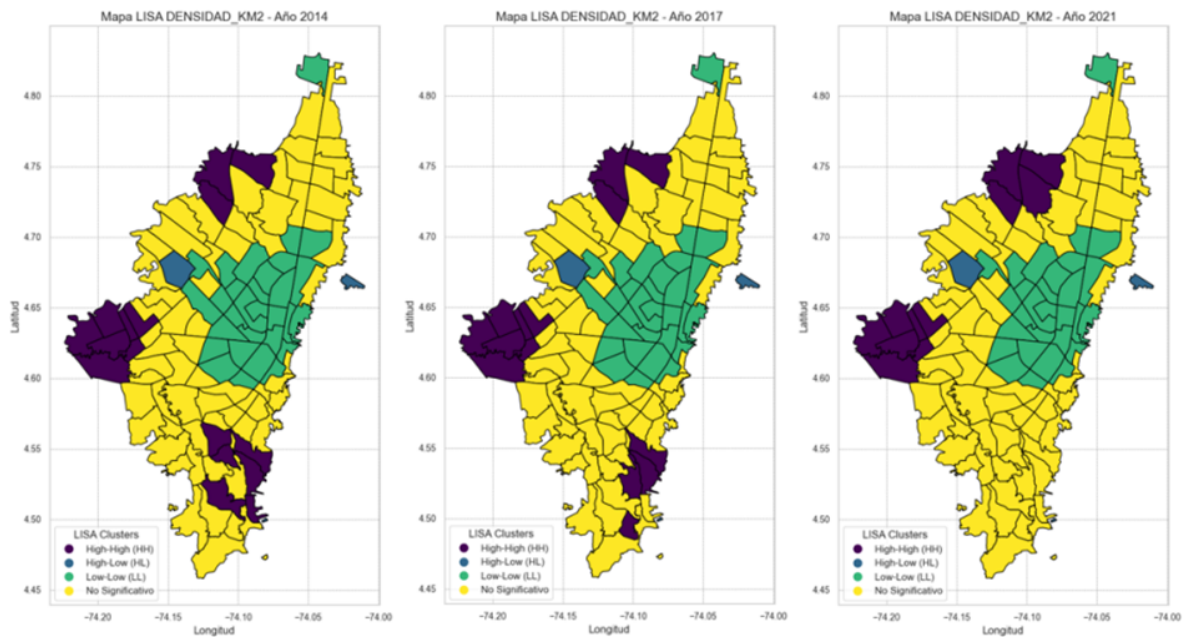


Figura 6.11: Evolución Espaciotemporal de los Clústeres LISA de la densidad poblacional (2014-2021).

En el mapa de 2021, los clústeres High-High se han vuelto a expandir, recuperando cobertura en el centro y sur, con una configuración más coherente y definida que en 2017. Las zonas Low-Low se mantienen estables en el norte y este, sin cambios significativos en su extensión. El número de UPZs clasificadas como No Significativo ha aumentado nuevamente, especialmente en el centro y occidente, lo que sugiere una mayor uniformidad en los valores de densidad entre vecinos. Las áreas High-Low han disminuido en cantidad y se limitan a sectores específicos del sur, mientras que los casos de Low-High siguen presentes de manera dispersa, principalmente en el norte.

6.4.3. Urbanización

De acuerdo con la Figura 6.12, desde 2014 se distingue con claridad un eje centro-sur fuertemente consolidado: un clúster High-High que agrupa unidades con altos niveles de urbanización rodeadas por otras igualmente densas. Este bloque no es solo un continuo físico; es también un continuo funcional, donde convergen infraestructura, servicios públicos, transporte, empleo y actividad económica. Su presencia indica que la urbanización no ocurre de manera aislada, sino por aglomeración: la ciudad se refuerza a sí misma allí donde ya es robusta.

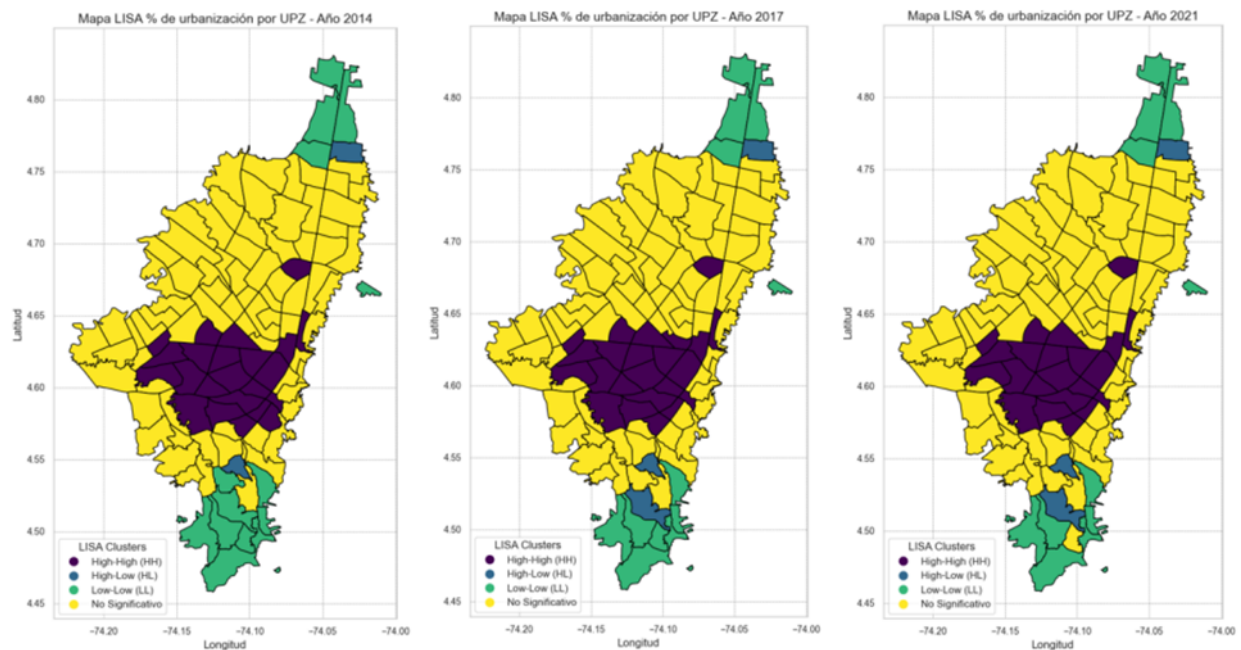


Figura 6.12: Evolución espaciotemporal de los clústeres LISA del porcentaje de urbanización (2014-2021).

Para 2017, este núcleo denso no solo se mantiene, sino que se expande levemente hacia UPZ adyacentes que antes no eran estadísticamente significativas. Mientras tanto, las periferias clasificadas como Low-Low (baja urbanización rodeada de baja urbanización) muestran otra cara de la misma ciudad. Estas áreas, ubicadas sobre todo en el extremo sur y el noreste, se sostienen en niveles reducidos de transformación urbana, lo que sugiere la persistencia de bordes rurales o semi-rurales dentro de la administración urbana formal. Además, aparece un fenómeno que matiza la lectura tradicional centro-periferia: surgen con más fuerza ciertos clústeres High-Low, especialmente hacia el sureste. Estos corresponden a UPZ altamente urbanizadas insertas en entornos poco urbanizados. Ese salto espacial habla de un crecimiento por avance puntual, casi como “salientes” de ciudad que penetran en zonas periféricas.

Y para 2021, el panorama empieza a mostrar señales de madurez y contención. El gran bloque High-High del centro-sur sigue siendo el corazón urbano, pero ya no absorbe nuevas UPZ con la misma intensidad observada en 2017. Eso sugiere que el proceso de densificación interna pudo haber alcanzado un umbral cercano a la saturación física o normativa: el núcleo metropolitano se estabiliza, más que seguir expandiéndose. Al mismo tiempo, se reduce la cantidad de clústeres High-Low dispersos.

6.5. Modelo de regresión ponderada geográficamente (GWR)

Los resultados se presentan de manera espacialmente clara. Cada coeficiente estimado se vincula a la ubicación del centroide de la UPZ correspondiente, lo que permite visualizar cómo cambia la relación entre las variables en diferentes partes del territorio. Esta asignación geográfica, basada en los centroides de las UPZ, asegura una coherencia espacial con la unidad de análisis elegida y facilita la comparación con otros indicadores a ese mismo nivel. Los mapas no solo ilustran la magnitud y dirección (ya sea positiva o negativa) del efecto de cada predictor, sino que también ponen de manifiesto patrones locales, discontinuidades y agrupamientos espaciales que podrían pasar desapercibidos en un enfoque de regresión global. En el anexo E, se encuentra el repositorio, en el cual está almacenada la matriz de pesos espaciales para los mapas GWR.

6.5.1. Proporción de expansión

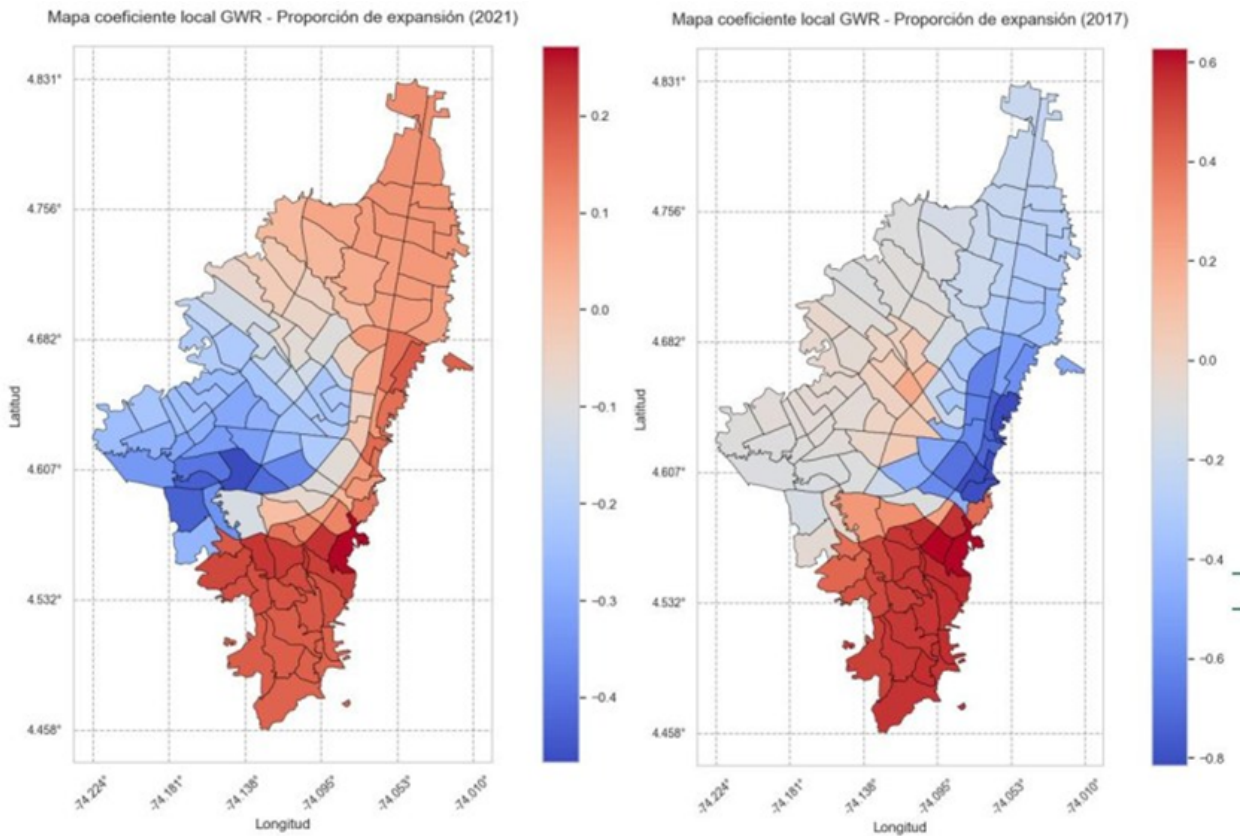


Figura 6.13: Mapa GWR de la proporción de expansión (2017 – 2021).

En 2017, la Figura 6.13 muestra que los efectos fueron más intensos y concentrados: los coeficientes alcanzaban valores de hasta $-0,8$ en zonas centrales y costeras consolidadas, donde la expansión urbana reducía significativamente la pobreza al estar acompañada de infraestructura, servicios y vivienda formal. Sin embargo, en sectores del sur y suroccidente, con coeficientes positivos cercanos a $0,4$, el crecimiento urbano no mitigaba la vulnerabilidad e incluso podía perpetuarla, reflejando procesos informales y deficiencias en la planificación. Para 2021, esta relación se volvió más débil y dispersa: predominaban coeficientes cercanos a $-0,4$ en el centro, sur y occidente (indicando una relación inversa aún presente pero menos intensa), mientras que en el norte y oriente el impacto era nulo o incluso positivo, asociado al crecimiento informal. Es importante señalar que, para el año 2014, la información sobre la proporción de expansión mostró un valor constante de 0 en todas las observaciones. Esta falta de variabilidad en los datos hace que no se pueda aplicar la Regresión Geográficamente Ponderada (GWR), ya que el modelo no tiene la capacidad de calcular coeficientes locales debido a la ausencia de una relación espacial clara. Por lo tanto, este conjunto de observa-

ciones no es suficiente para crear un modelo válido para ese período lo cual resulta complejo y nulo la elaboración de una representación cartográfica.

6.5.2. Densidad

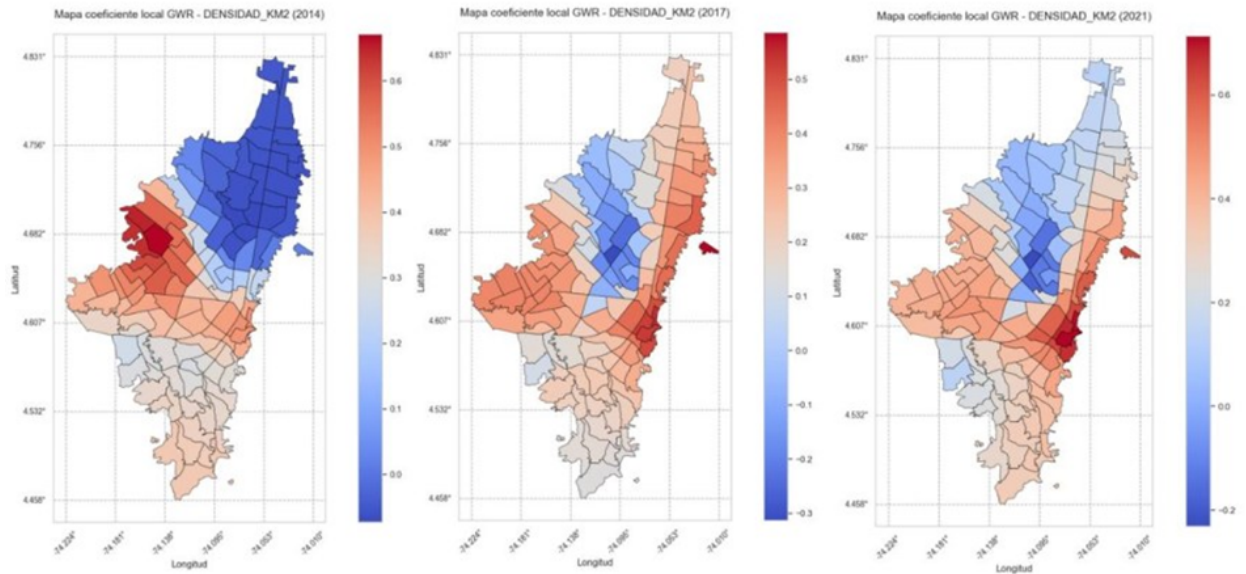


Figura 6.14: Mapa GWR de la densidad poblacional (2014 – 2021).

La Figura 6.14 muestra la relación entre la densidad poblacional y pobreza, en el año 2014, las áreas con coeficientes locales más altos, los cuales están representadas en tonos rojos y anaranjado, se concentran en el sureste y centro del territorio, con valores que alcanzan hasta 0.6. Estas zonas indican una relación positiva entre densidad poblacional y pobreza. El norte y sur predominan en tonos azules claros y grises, con coeficientes cercanos a cero, lo que refleja una ausencia de relación significativa entre ambas variables en esas regiones.

En 2017 mostraba una polarización clara: en el norte y el centro, coeficientes cercanos a cero indicaban que la densidad no incidía significativamente gracias a la presencia de infraestructura, empleo formal y servicios que amortiguaban la vulnerabilidad, mientras que en el sureste y las zonas costeras orientales valores cercanos a 0,8 revelaban un vínculo directo y fuerte, asociado a contextos de informalidad, saturación de servicios y falta de regulación. Para 2021, aunque persiste la desigualdad territorial, el panorama se vuelve más complejo: en el norte, centro y occidente la densidad sigue teniendo un efecto casi nulo debido a condiciones urbanas favorables, pero en el sureste y las costas orientales los coeficientes cercanos a 1,0 reflejan que la concentración poblacional intensifica la pobreza por problemas

estructurales como déficit de vivienda, servicios y sobrepoblación. La distribución bimodal de los coeficientes (con un grupo entre 0,0–0,2 y otro entre 0,8–1,0) confirma esta dicotomía, evidenciando que la densidad puede ser neutral o profundamente agravante según el contexto socioespacial.

6.5.3. Urbanización

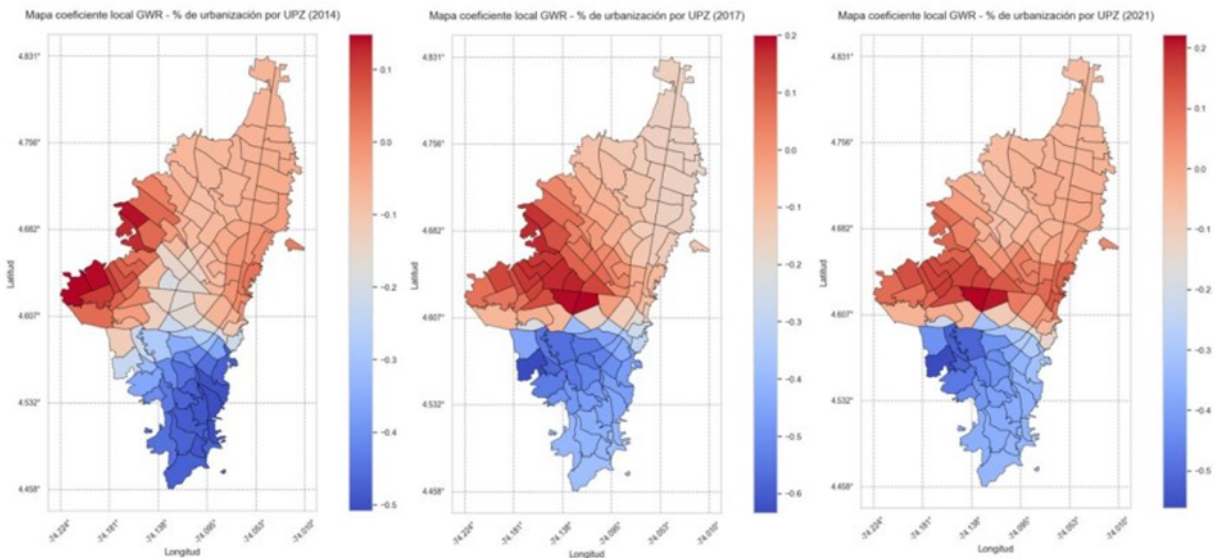


Figura 6.15: Mapa GWR del porcentaje de urbanización (2014 – 2021).

La Figura 6.15 muestra que para el año 2014, las áreas con coeficientes más negativos representados en tonos azules oscuros se concentran en el sur y sureste del territorio, con valores que alcanzan hasta -0.5. Estas zonas indican una relación inversa entre el porcentaje de urbanización y la pobreza: a mayor urbanización, menor pobreza. El centro y norte predominan en tonos rojos y anaranjados, con valores cercanos a cero o ligeramente positivos, lo que refleja una ausencia de relación significativa o incluso una ligera correlación positiva entre ambas variables en esas regiones.

En 2017, el porcentaje de urbanización mostraba efectos fuertemente contrastantes sobre la pobreza: en el sur y sureste, coeficientes muy negativos (hasta -0,5) evidenciaban que el avance urbano se asociaba con reducciones significativas de la vulnerabilidad, impulsadas por la provisión de infraestructura, servicios y programas focalizados, mientras que en el norte y el centro los valores cercanos a -0,1 reflejaban impactos débiles, donde la urbanización no se traducían en mejoras sociales debido a procesos irregulares o a la llegada de población vulnerable sin cambios sustantivos en sus condiciones. Para 2021, el patrón se volvió más heterogéneo

y menos extremo: en el sur y suroeste los coeficientes cercanos a $-0,4$ indicaron que la urbanización seguía contribuyendo a disminuir la pobreza gracias a desarrollos relativamente planificados, mientras que en el centro y noreste los valores alrededor de 0 o ligeramente positivos mostraron que la expansión tenía efectos nulos o incluso coexistía con la vulnerabilidad al producirse de manera informal y sin mejoras estructurales. La atenuación de los extremos y la expansión de áreas con efectos neutros sugieren que el impacto del desarrollo urbano pasó de estar concentrado en focos específicos a distribuirse de forma más homogénea en el territorio, con implicaciones menos intensas, pero más extendidas.

6.6. Dinámicas socioespaciales y económicas

Bogotá D.C. exhibe tres realidades territoriales distintas. En el sur y sureste persisten núcleos de alta densidad poblacional que se superponen con pequeñas áreas de elevada pobreza multidimensional; allí, la autocorrelación espacial de la vulnerabilidad es la más intensa y tiende a recomponerse después de episodios de aparente dispersión. En el centro se observan transiciones: conviven barrios consolidados con zonas en reconfiguración, donde aumentan los contrastes entre UPZ contiguas y se alternan con zonas de mayor y menor vulnerabilidad. En el norte, el espacio disponible para expandirse es cada vez más escaso, la densificación gana terreno y los niveles de pobreza se mantienen relativamente bajos, aunque emergen tensiones asociadas al encarecimiento del suelo y a procesos de sustitución socio-residencial.

En el sur y sureste, el aumento de la densidad y los niveles altos del IPM se refuerzan mutuamente. Los mapas LISA muestran clústeres High-High persistentes, y el Moran global del IPM regresó en 2021 a valores por encima de 0,6 tras una caída en 2017, señalando la reconcentración de la pobreza en bloques territoriales amplios. Esta persistencia se explica por la combinación de crecimiento demográfico sin infraestructura proporcional, déficits de servicios y oportunidades, y una informalidad urbana que abarca más del 30% del área urbana, con 9.016 hectáreas de origen informal según el SDHT [84]. La respuesta pública reciente ha avanzado por dos vías: la institucionalización/regularización de asentamientos (por ejemplo, la legalización de 32 barrios en Bosa, Kennedy, Engativá y otras localidades para habilitar inversión en vías, alumbrado, seguridad y parques) y la producción de vivienda social (VIS/VIP) como alternativa a la expansión informal [85]. En términos de relaciones espaciales, la GWR confirma que, cuando la expansión urbana ocurre con acompañamiento de infraestructura y vivienda formal, puede asociarse con reducciones del IPM; en cambio, donde la expansión es informal, el efecto es nulo o incluso positivo sobre la pobreza, reforzando vulnerabilidades locales.

En el centro y en el norte, la restricción física para expandirse hacia afuera ha desplazado la dinámica hacia adentro mediante densificación y recambio del tejido construido. En el norte

se concentran los valores más altos por metro cuadrado de la ciudad, impulsados por la alta demanda en barrios exclusivos y la escasez de suelo urbanizable: en 2024, La Cabrera y El Chicó promediaron cerca de \$9 millones/m², y sectores como Antiguo Country o Santa Ana se ubicaron en rangos similares (\$8 - 9 millones/m²) [86]. Ya en 2016, proyectos nuevos en Chicó/Chicó Alto rondaban 11 millones/m², y hacia 2018–2020 el agotamiento de lotes en las zonas más exclusivas elevó aún más los precios de reventa y desplazó a compradores con menores presupuestos hacia áreas adyacentes o más al sur [87]. En el centro, (La Candelaria, Santa Fe, Teusaquillo, Los Mártires, entre otros) la trayectoria ha sido mixta: tras una desvalorización de fines del siglo XX, varios barrios se revalorizaron por procesos de recuperación urbana, oferta cultural y vivienda para estratos medios; así, en junio de 2024 zonas como Santa Sofía y San Martín registraron precios cercanos a \$8,8 millones/m², comparables con áreas de estratos altos. Barrios más tradicionales como La Merced, La Macarena y Teusaquillo también incrementaron precios por su localización y revitalización patrimonial y comercial, y, de forma más amplia, incluso los estratos 1–2 mostraron en 2022 una valorización superior al 12 % [88]. En términos espaciales, esto se traduce en un aumento de categorías de transición en los LISA (High-Low y Low-High) desde 2017, con una mezcla fina de condiciones socio-urbanas y pérdida de fronteras claras entre zonas catalogadas como “homogéneas”.

La problemática dominante en estos sectores no es la informalidad sino la gentrificación: la rápida apreciación del suelo y de los arriendos dificulta la posibilidad de los hogares de menores ingresos de permanecer y acelera la sustitución social. Casos locales ilustran dichos mecanismo: en Chapinero Alto y Park Way, antiguamente asequibles para estudiantes, los cánones hoy en día se equiparan con los de zonas de estrato 5; en La Candelaria, la transformación de viviendas patrimoniales en espacios para uso turístico elevó los arriendos y expulsó a los antiguos residentes; en Quinta Camacho, el m² pasó de valer \$2,6 a \$4 millones en pocos años durante el auge comercial [87]; y en La Macarena/Teusaquillo se reportaron alzas del 20 - 30 % tras mejoras del espacio público o esfuerzos para la conservación. Además, parte de la oferta se redirige a alquiler turístico (p.ej., Airbnb) u oficinas, reduciendo la oferta de viviendas para arriendo y presionando los precios, con desplazamientos hacia las periferias en busca de rentas más bajas [86].

La migración internacional alteró con fuerza la demografía reciente. Hasta 2016, Bogotá D.C. tenía participaciones históricamente bajas de población extranjera; a partir de 2017, la crisis venezolana convirtió a Colombia en receptor principal y la ciudad pasó de 1 % de población migrante (2017) a 4,3 % (2020) y hasta 6,1 % en 2021, equivalentes a más de 500.000 personas [88]. Estas llegadas se acomodaron preferentemente en zonas urbanas con vivienda asequible, y a menudo informal, del sur y suroccidente, presionando servicios, empleo e infraestructura y elevando la densidad en sectores ya tensionados. La política pública debe reconocer la centralidad de este problema: sin una oferta suficiente y bien localizada de

viviendas asequibles (como las VIS/VIP), y sin una regularización enfocada con servicios efectivos, la migración termina anclándose en circuitos de precariedad que los indicadores espaciales señalan como clústeres persistentes.

La pandemia de COVID-19 introdujo un choque súbito sobre estas trayectorias. En 2020, las cuarentenas y cierres de fronteras redujeron 72,2 % los movimientos internacionales respecto a 2019; en 2021 el ingreso repuntó casi en la misma magnitud y las salidas de colombianos crecieron 92,2 % [89] [89]. A escala local, incluso se estimó crecimiento poblacional negativo en 2020 por salidas temporales hacia municipios cercanos. En lo social, Bogotá D.C. pasó de un IPM cercano al 4,1 % en 2018 a 7,1 % en 2019 y 7,5 % en 2020, revirtiendo algunos avances previos. De hecho, todas las dimensiones del IPM se resintieron y la ciudad concentró 31,3 % del aumento nacional de nuevos pobres durante la pandemia [90] [91].

La política habitacional ha respondido con ciclos de expansión y retrocesos. Entre 2014 y 2018 se impulsaron VIS/VIP con subsidios, proyectos masivos (p. ej., Campo Verde) y reasentamientos de hogares en riesgo; hubo 28.000 viviendas VIP gestionadas y un fuerte aumento de iniciaciones VIS [92] [93]. Entre 2019 y 2024 continuaron los subsidios y programas como “Mi Ahorro, Mi Hogar”, con énfasis en mujeres cabeza de hogar, aunque con algunos desafíos en su ejecución (p. ej., bajo avance de Plan Terrazas) y una desaceleración importante del mercado VIS en 2024 [84] [94]. Si bien el déficit habitacional se redujo de significativamente en el periodo de 2003 a 2017 (17,4 % a 4,6 %), volvió a subir para los últimos años, quedando en un 7 % para 2024. Además, las ocupaciones informales casi se duplicaron entre 2009 y 2024, donde uno de cada cinco bogotanos percibía la política de vivienda como insuficiente [95] [96]. Por tanto, la oferta social no ha compensado la demanda generada por crecimiento, migración y recambio interno, y la informalidad permanece como medida de escape cuando la ciudad formal no provee soluciones oportunas.

Desde la perspectiva analítica, los modelos espaciales ayudan a dirimir por qué las mismas palancas producen efectos distintos según el territorio. La GWR muestra que la expansión urbanística asociada a provisión efectiva de servicios y vivienda formal se vincula a menores niveles de pobreza (coeficientes negativos, especialmente en 2017, y todavía presentes en 2021 en el sur y suroeste), mientras que la densidad es “condicional”: en el norte, centro y occidente su efecto es cercano a cero, pero en el sureste y oriente densidades altas se asocian a mayor pobreza. La urbanización, por su parte, ha pasado de impactos muy negativos concentrados (2017) a efectos menos extremos y más extendidos (2021), con zonas donde incluso resulta neutra o coexiste con vulnerabilidad cuando se produce de forma irregular.

7. Desarrollo de visualizaciones interactivas

En este capítulo se desarrollan y presentan las visualizaciones interactivas correspondientes a los resultados obtenidos para cada uno de los objetivos planteados. Además, se detalla la herramienta utilizada para su elaboración, describiendo sus principales funcionalidades y la forma en que facilita la exploración de los datos. Cada visualización se acompaña de una explicación sobre la información que representa, los indicadores que se muestran y la manera en que contribuye a la interpretación de los resultados obtenidos en el estudio.

7.1. Selección del entorno

Para la presentación de las visualizaciones se seleccionó GitHub Pages [97], una plataforma gratuita que permite publicar sitios web estáticos directamente desde un repositorio de GitHub, facilitando la integración del código, las visualizaciones y la documentación del proyecto. Su compatibilidad con herramientas interactivas desarrolladas en Python, junto con librerías de JavaScript para dashboards y mapas dinámicos, permite una visualización fluida en cualquier navegador sin requerir software adicional. Asimismo, admite la inserción de aplicaciones creadas en Google Earth Engine Apps mediante archivos HTML embebidos, integrando de forma unificada los mapas y herramientas interactivas dentro del entorno web del proyecto.

En el siguiente enlace se puede acceder a la plataforma donde se encuentran publicadas las visualizaciones desarrolladas: <https://christianveram.github.io/ExpansionUrbana/index.html>

7.2. Desarrollo de visualizaciones

El dashboard está compuesto por cuatro pestañas principales. La primera corresponde a la clasificación del suelo obtenida mediante el modelo seleccionado, la segunda presenta la distribución de la pobreza derivada de la EM, la tercera muestra la expansión urbana a lo largo del periodo de análisis y la cuarta integra el análisis conjunto entre la expansión urbana y la pobreza, permitiendo explorar la relación espacial entre ambos fenómenos. Cada una de estas secciones se explica a continuación.

7.2.1. Clasificación del suelo urbano

En esta sección se presenta la clasificación del suelo urbano en Bogotá D.C., elaborada a partir del procesamiento de datos geospaciales en la plataforma Google Earth Engine. La

visualización utiliza imágenes satelitales compuestas por las bandas 4, 3 y 2 de Landsat 8, permitiendo una interpretación visual cercana a la percepción natural del territorio. Sobre estas imágenes se aplicó el modelo de Random Forest, complementado con filtros de post clasificación para mejorar la coherencia espacial y temporal de los resultados.

La herramienta incorpora funcionalidades interactivas que permiten comparar la imagen original compuesta con las bandas RGB y la representación clasificada obtenida por el modelo y las reglas de post-procesamiento aplicadas. Asimismo, el usuario puede seleccionar una o varias Unidades de Planeamiento Zonal para analizar las áreas clasificadas como urbanas o no urbanas y observar su evolución entre los años 2014 y 2024. Esta capacidad de exploración temporal y espacial facilita la comprensión de los cambios en la ocupación del territorio a lo largo del periodo analizado.

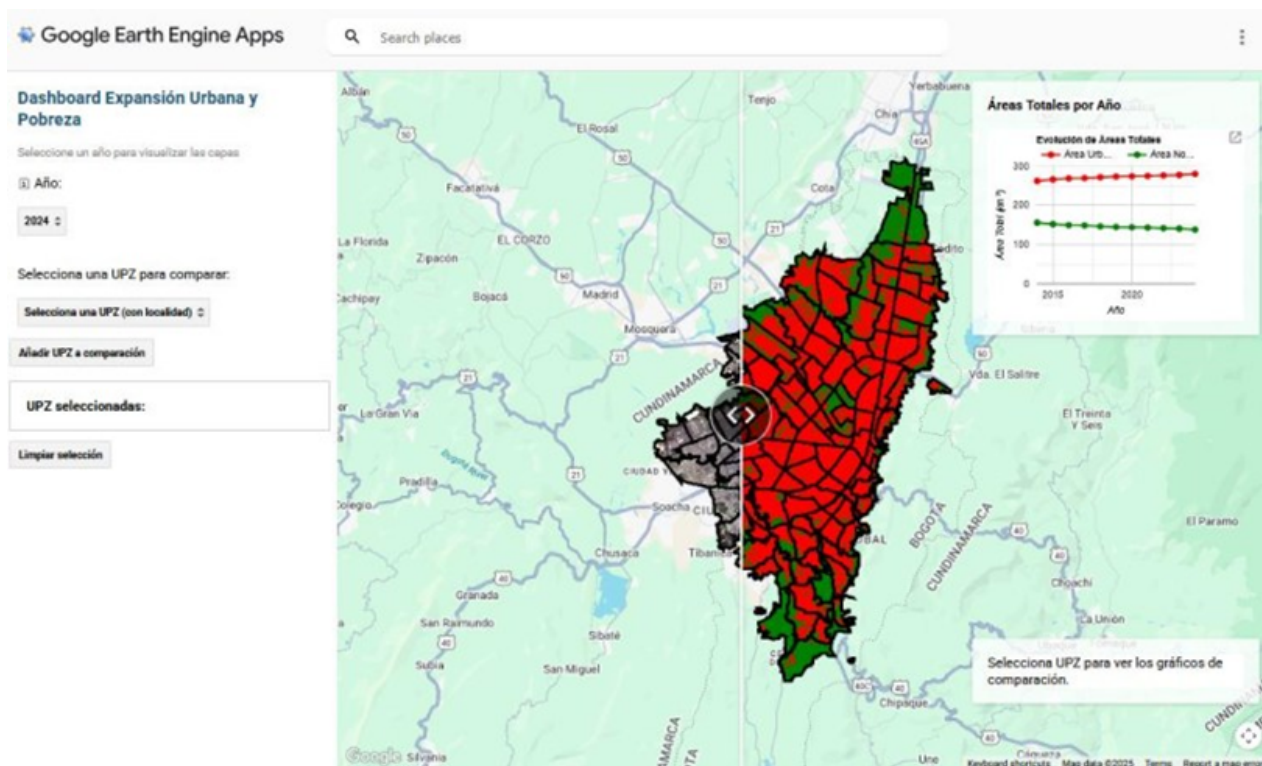


Figura 7.1: Visualización de clasificación del suelo urbano por año.

Adicionalmente, se incluye un gráfico de barras que muestra las Unidades de Planeamiento Zonal con mayor superficie clasificada como urbana o no urbana para cada año del periodo estudiado. Este resumen visual permite identificar de manera rápida las zonas que concentran la mayor extensión de cada categoría, facilitando su comparación y el análisis de tendencias en la distribución espacial del uso del suelo. En la Figura 7.1 se observa la interfaz principal de la aplicación con el mapa interactivo y el panel lateral de control. En la Figura 7.2 se

muestra el gráfico de barras que resume las UPZ con mayor área urbana o no urbana por año.

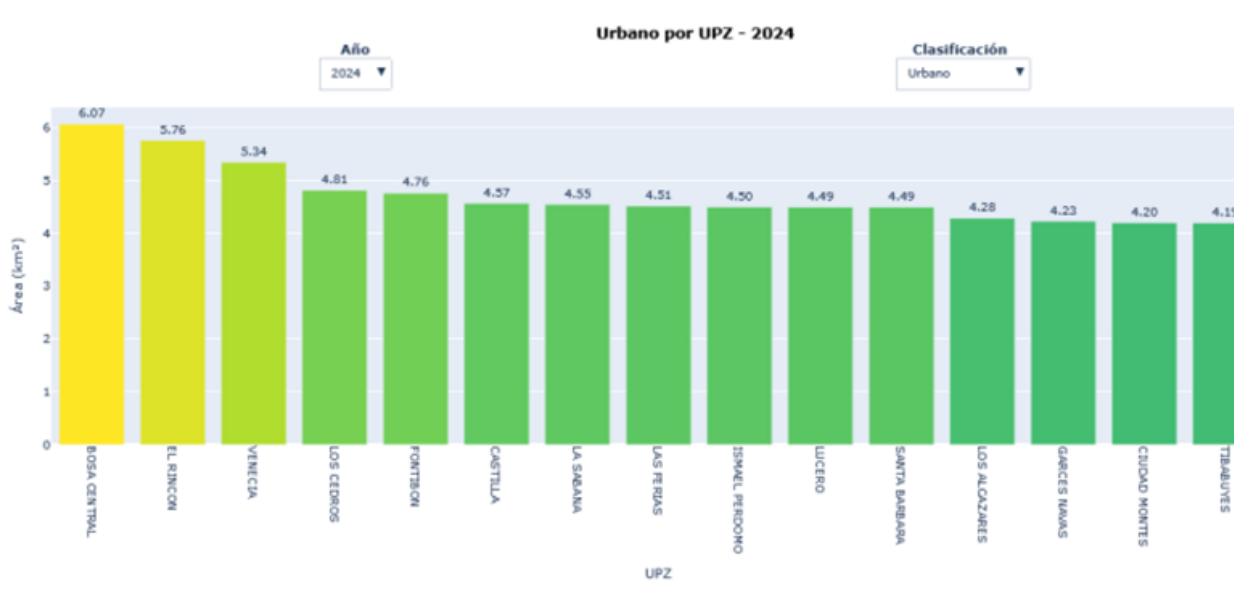


Figura 7.2: Visualización gráfica de barras de área urbana y no urbana para ejemplo en el año 2024.

7.2.2. Población en condición de pobreza 2014, 2017 y 2021

En esta sección se presentan los resultados de la extracción de índices de pobreza en Bogotá D.C., correspondientes al porcentaje de población en condición de pobreza para los años 2014, 2017 y 2021, desagregados por Unidades de Planeamiento Zonal (UPZ). Los datos utilizados provienen de la EM del Departamento Administrativo Nacional de Estadística (DANE), los cuales fueron procesados y visualizados mediante la librería Plotly en Python, permitiendo generar representaciones interactivas y dinámicas.

Las visualizaciones incluyen mapas temáticos que muestran la distribución espacial de la pobreza en cada año analizado, permitiendo identificar las zonas con mayor concentración de población en situación de vulnerabilidad económica y analizar su evolución temporal dentro del periodo de estudio.

Adicionalmente, se presenta un gráfico de barras que resume las UPZ con mayores porcentajes de población en condición de pobreza para cada año evaluado. Este recurso facilita la comparación entre unidades territoriales y permite identificar las áreas más vulnerables. En la Figura 7.3 se observan los tres mapas temáticos correspondientes a los años 2014, 2017 y 2021, mientras que en la Figura 7.4 se muestra el gráfico de barras con las UPZ que registraron los mayores niveles de pobreza en 2021.

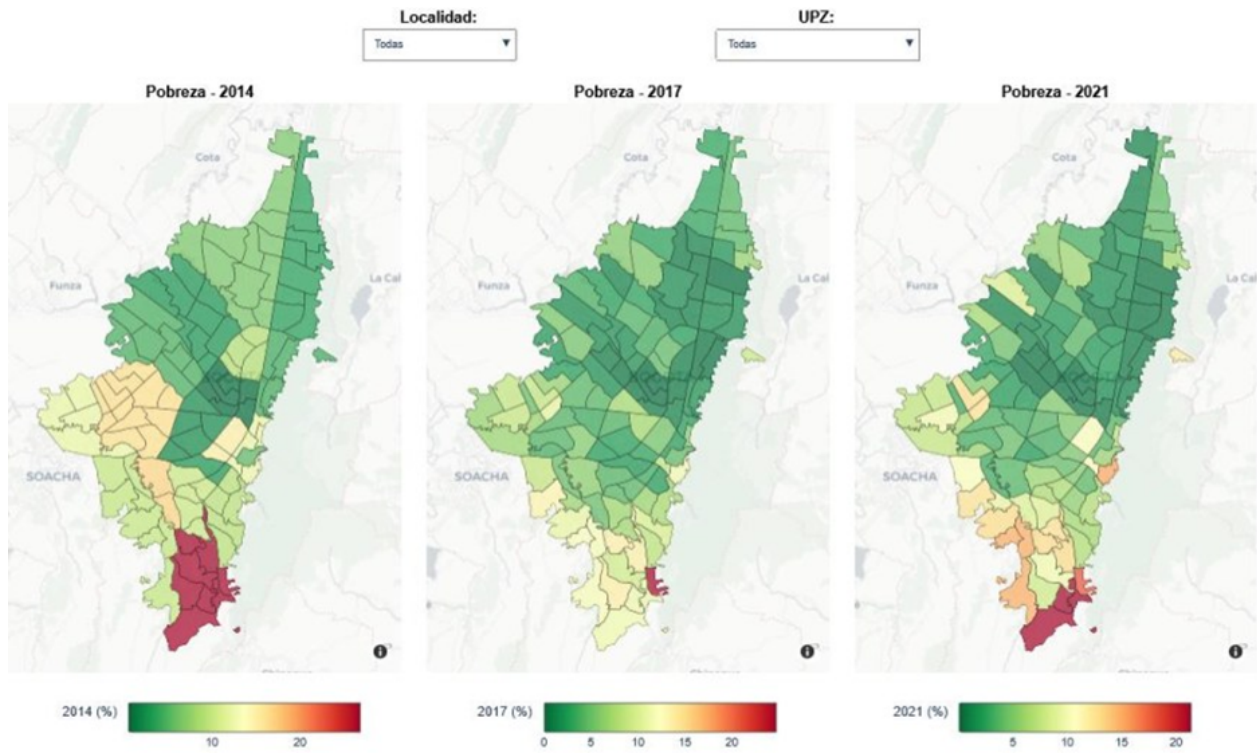


Figura 7.3: Visualización población en condición de pobreza (2014,2017 y 2021).

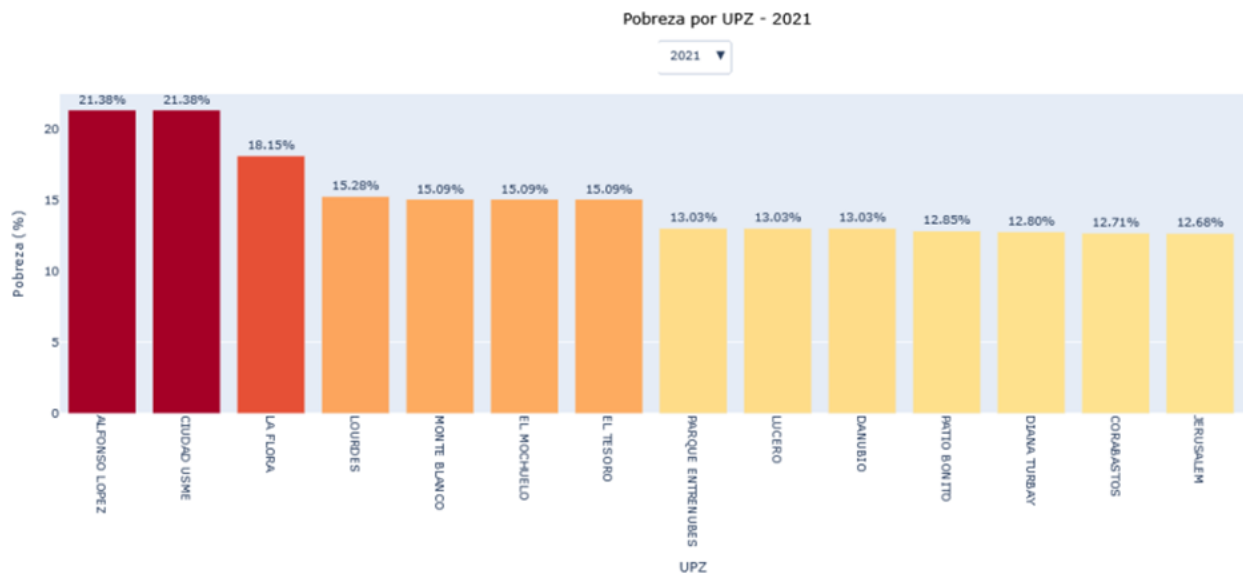


Figura 7.4: Visualización población en condición de pobreza para el año 2021.

7.2.3. Expansión Urbana

En esta sección se presenta la visualización de la expansión urbana en Bogotá D.C. entre los años 2014 y 2024. La representación muestra las áreas donde ha ocurrido crecimiento urbano acumulado a partir de una imagen de referencia del año 2014, identificando tres categorías: suelo no urbano, suelo urbano estable y nuevas zonas urbanizadas. Esta visualización fue desarrollada utilizando la librería Folium en Python, integrando capas raster generadas a partir de imágenes satelitales reclasificadas y datos vectoriales de las Unidades de Planeamiento Zonal.

La interfaz incluye un mapa interactivo con un control desplegable que permite al usuario seleccionar el año de visualización, con el año 2024 establecido como valor predeterminado. Además, al hacer clic sobre cualquier UPZ, se despliega una tabla con la evolución temporal de su superficie urbana en kilómetros cuadrados, así como su porcentaje de crecimiento. Estas funcionalidades facilitan la exploración espacial y temporal del fenómeno de expansión urbana en la ciudad.

En la Figura 7.5 se observa el mapa interactivo que representa el crecimiento urbano acumulado desde 2014 hasta 2024, incluyendo la leyenda temática y el panel de selección de años.

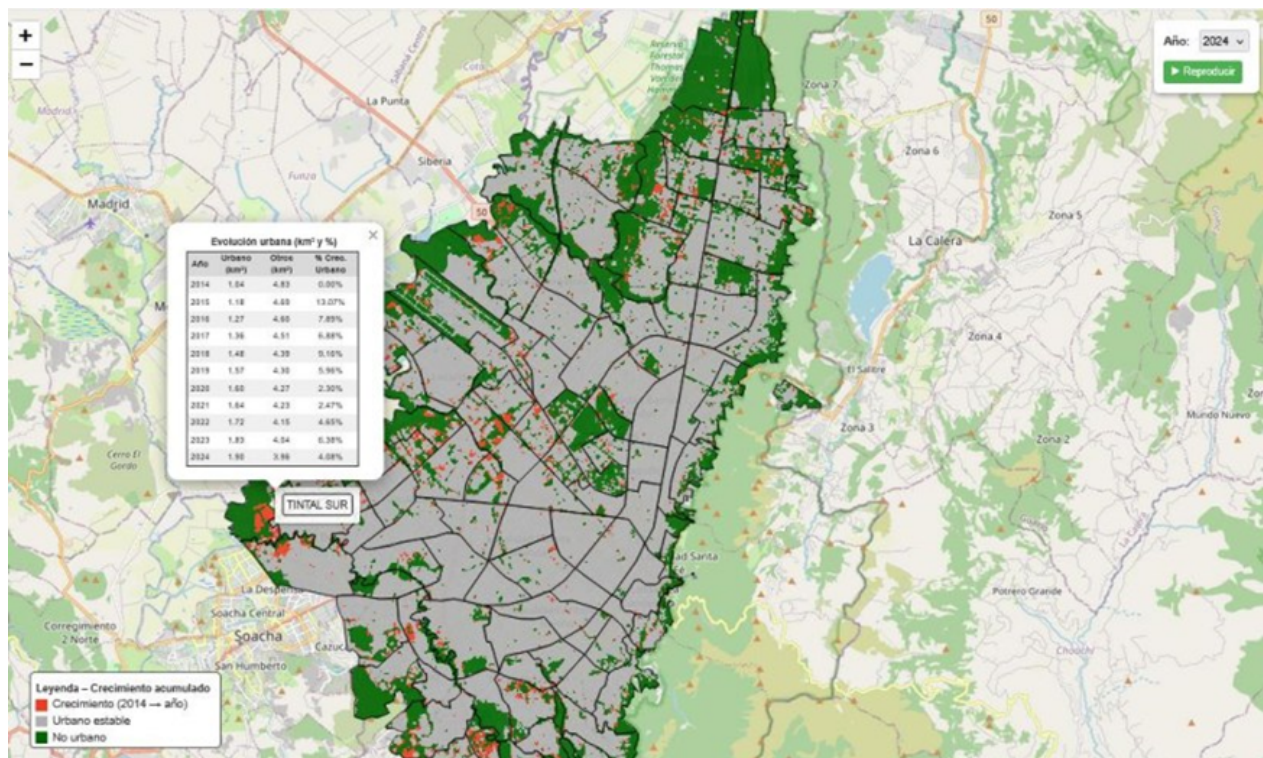


Figura 7.5: Visualización de la evolución del crecimiento urbano por UPZ.

7.2.4. Relación Expansión Urbana y Pobreza

En esta sección se presenta la caracterización de las variables económicas y sociodemográficas (IPM, densidad poblacional, urbanización y proporción de expansión urbana), junto con un análisis detallado de las dinámicas e interacciones, cuyo contexto varía dependiendo de cada zona. Esta visualización fue desarrollada utilizando la librería Plotly en Python (con apoyo de otras librerías), integrando los resultados encontrados tanto en los mapas LISA como los GWR.

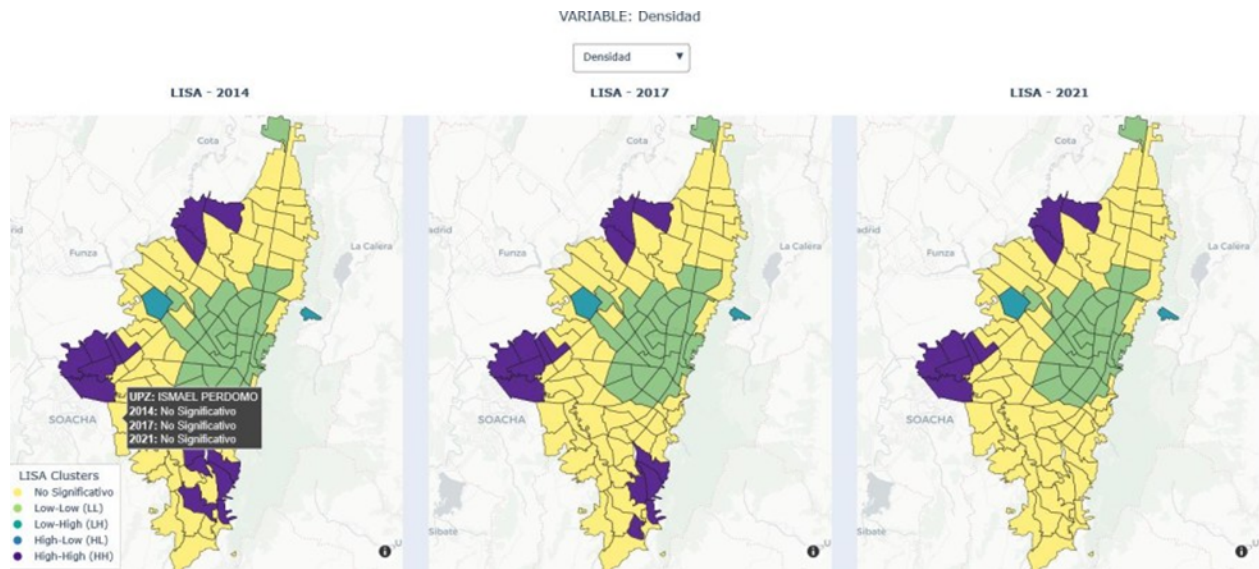


Figura 7.6: Visualización del mapa LISA para la densidad poblacional.

La interfaz incluye un mapa interactivo con un control desplegable que permite al usuario seleccionar el año de visualización, con el año 2017 establecido como valor predeterminado. Además, al hacer clic sobre cualquier UPZ, se despliega información adicional, dependiendo del mapa. Estas funcionalidades facilitan la comprensión de las interacciones y los comportamientos demográficos y económicos de cada parte de la ciudad.

En la Figura 7.6 se observa un ejemplo de uno de los mapas interactivos que se presentarán en esta pestaña

8. Conclusiones y trabajos futuros

8.1. Conclusiones

A lo largo del documento se integró una base de datos que combina series temporales de imágenes Landsat 8 (2014–2024) con indicadores socioeconómicos de la EM (2014, 2017, 2021), con el fin de seguir la expansión urbana de Bogotá D.C. y su vínculo con la pobreza. El preprocesamiento de las imágenes, realizadas en Google Earth Engine, permitió disponer de series temporales homogéneas. Este proceso incluyó la corrección atmosférica, la eliminación de nubes y sombras, y la creación de composiciones anuales mediante la mediana de los píxeles válidos. Con respecto a los datos socioeconómicos, la menor desagregación espacial de 2014 presentó retos para homogeneizar la unidad de análisis, llevando a la necesidad de aplicar la técnica de coincidencia de cuantiles (MQE), para alinear las estimaciones del IPM con las cifras oficiales de la Secretaría de Planeación y así asegurar la comparabilidad en toda la serie.

Sobre esta base, se entrenó y validó un clasificador Random Forest que distinguió con alta precisión las áreas urbanas y no urbanas (F1-score = 0.9423), superando otras alternativas, tanto de métodos supervisados como no supervisados. La inclusión de índices espectrales (NDVI, NDBI, NDWI) resultó fundamental para mejorar el poder discriminativo de la cobertura del suelo urbano. La aplicación posterior de filtros espaciales y temporales de post-clasificación incrementó la coherencia interna de la serie y corrigió fragmentaciones, permitiendo estimar un crecimiento acumulado del 6,9

Al evaluar la relación entre la expansión urbana y la pobreza, se evidenciaron distintos patrones dependiendo del contexto. Los análisis LISA y la regresión ponderada geográficamente evidenciaron que la asociación varía según el territorio: en el sur y sureste, la urbanización informal tiende a consolidar clústeres de alta pobreza. Los aumentos en la densidad poblacional de estas zonas amplifican las carencias donde coinciden con déficits persistentes en servicios y calidad de la vivienda. Sin embargo, cuando la expansión se acompaña de infraestructura y vivienda formal, se llegan a observar reducciones del IPM y mejoras en el bienestar de los hogares en distintos componentes. Por otro lado, dichos efectos son más bien neutros en el norte y el centro, donde las variaciones tanto en la densidad poblacional como en los porcentajes de urbanización no parecían afectar tan severamente las condiciones socioeconómicas de los habitantes de estos sectores. Debido a que las problemáticas de estas zonas se relacionan más con los aumentos en el costo del suelo y las revalorizaciones, a causa de procesos como la recuperación urbana y la gentrificación.

Por último, para facilitar la interpretación y el uso de la evidencia, se desarrolló un dashboard interactivo con capas temporales, selección por UPZ y comparaciones entre imágenes satelita-

les y clasificaciones, lo que habilita exploraciones interactivas de posibles interesados técnicos y de política pública que pudiesen estar interesados en conocer más sobre las interacciones vistas en la ciudad con relación a las variables analizadas.

Si bien los resultados obtenidos son robustos, la investigación enfrentó limitaciones técnicas y de alcance que deben ser mencionadas. En primer lugar, la resolución espacial de las imágenes del Landsat 8 son de 30 metros, lo que dificultaba capturar con mayor detalle las dinámicas intraurbanas más finas, como cambios al interior de manzanas, pasajes o asentamientos de muy pequeña escala, lo que puede subestimar procesos de densificación o informalidad muy localizada. En segundo lugar, las restricciones de tiempo y procesamiento solo permitieron aplicar la metodología al caso de Bogotá D.C., dejando por fuera la posibilidad de contrastar estos patrones con otras ciudades intermedias o con el conjunto del país, lo que habría permitido evaluar mejor la generalización de los hallazgos y sus posibles variaciones regionales. Finalmente, es importante volver a mencionar que el cálculo del IPM oficial no estaba disponible para el año 2014, por lo que fue necesario estimarlo; contar con esta medición directa habría permitido construir comparaciones más precisas y robustas a lo largo del periodo analizado.

A pesar de ello, la investigación ofrece una base metodológica con alta potencialidad de aplicación práctica. El enfoque integrado puede ser empleado por entidades de planeación y sectores sociales para monitorear el cumplimiento de metas urbanas y de reducción de pobreza, identificar UPZ o sectores prioritarios para intervenciones en infraestructura, vivienda o programas sociales, y evaluar ex ante y ex post el impacto territorial de instrumentos de ordenamiento como planes parciales, macroproyectos o programas de renovación urbana. Asimismo, el esquema propuesto podría adaptarse y escalarse a otras ciudades colombianas o de la región, permitiendo construir sistemas de alerta temprana sobre expansión informal, orientar la focalización de recursos públicos y apoyar decisiones de política que busquen que la urbanización sea un vehículo efectivo de inclusión social y no un factor adicional de reproducción de la pobreza.

En conjunto, estos resultados demuestran que es posible construir una línea de base confiable y comparable para medir la expansión urbana y relacionarla con el IPM, y que la dirección de dicha relación depende del tipo de urbanización y de las características demográficas y sociales propias de cada zona. La combinación de teledetección, aprendizaje automático y análisis espacial fino no solo mejora la estimación del crecimiento urbano, sino que también ofrece insumos para evaluar las desventajas presentes en la ciudad y potenciar los contextos donde la urbanización formal contribuye a reducir la pobreza. Este enfoque invita a mirar la ciudad como un espejo cambiante, donde el crecimiento urbano no solo transforma el paisaje, sino también las formas de vida y las oportunidades que allí se hayan.

8.2. Trabajos futuros

A partir de la temática planteada y teniendo en cuenta las limitaciones técnicas y temporales, se proponen algunas ideas para trabajos futuros que permitiesen profundizar en temáticas relacionadas del presente documento.

8.2.1. Integración de imágenes satelitales de mayor resolución espacial y temporal

Si bien las imágenes Landsat 8 ofrecen una cobertura temporal amplia y gratuita, su resolución espacial (30 metros) limita la detección de detalles finos en áreas densamente urbanizadas. Se recomienda incorporar sensores de mayor resolución espacial, como Sentinel-2, PlanetScope o WorldView, que permitan obtener resultados más precisos en la delimitación de zonas construidas, corredores viales y espacios verdes. Con ello, se contribuiría a perfeccionar los modelos de clasificación y a realizar un seguimiento más detallado de los procesos de urbanización informal presentes en la zona sur de la ciudad.

8.2.2. Ampliación del análisis a nivel regional o nacional

Otra línea de trabajo consiste en extender la metodología aplicada a otras ciudades o regiones del país, con el fin de identificar patrones comunes y diferencias en los procesos de expansión urbana y pobreza. Esto permitiría construir una visión más completa del fenómeno a nivel nacional y aportar información útil para la formulación de políticas públicas en materia de desarrollo urbano sostenible.

8.2.3. Incorporación de variables ambientales y de sostenibilidad

Para avanzar hacia un enfoque integral, se sugiere incluir indicadores ambientales como la cobertura vegetal, la contaminación del aire o la disponibilidad de recursos hídricos. De esta manera, sería posible evaluar no solo la expansión física de la ciudad, sino también sus implicaciones ecológicas y la sostenibilidad de los nuevos desarrollos urbanos. Buscando así plantear estrategias de mitigación y adaptación frente a los impactos ambientales derivados del crecimiento urbano.

8.2.4. Desarrollo de modelos predictivos y escenarios de crecimiento urbano

El uso de modelos de simulación espacial permitiría proyectar la expansión urbana futura bajo diferentes escenarios de planificación. Estas proyecciones serían de gran utilidad para anticipar las zonas de posible crecimiento informal y diseñar políticas preventivas orientadas a un desarrollo urbano más equilibrado.

Anexos

En el presente anexo se incluyen los resultados detallados obtenidos durante la evaluación de los modelos aplicados bajo las tres configuraciones de variables elegidas.

Las tablas que se presentan a continuación muestran las métricas de desempeño obtenidas en cada partición del proceso de validación cruzada, incluyendo precisión, recall, F1-score y exactitud. Estos resultados complementan la información resumida en el cuerpo principal del documento y permiten una visualización más completa del comportamiento del modelo en cada escenario de entrenamiento.

Anexo A. Resultados del modelo Random Forest

A.1 Resultados del modelo Random Forest con bandas espectrales

Tabla A.1: Resultados del modelo Random Forest con bandas espectrales

Fold	Accuracy	Precision	Recall	F1-Score
1	0.9386	0.9394	0.9386	0.9385
2	0.9229	0.9230	0.9229	0.9229
3	0.9357	0.9357	0.9357	0.9357
4	0.9486	0.9486	0.9486	0.9486
5	0.9186	0.9190	0.9186	0.9186
6	0.9286	0.9289	0.9286	0.9286
7	0.9571	0.9572	0.9571	0.9571
8	0.9571	0.9377	0.9371	0.9371
9	0.9343	0.9343	0.9343	0.9343
10	0.9429	0.9429	0.9429	0.9429
Promedio ± Desv. Est.	0.9364 ± 0.0109	0.9367 ± 0.0108	0.9364 ± 0.0109	0.9364 ± 0.0109

A.2 Resultados del modelo Random Forest con índices y bandas espectrales

Tabla A.2: Resultados del modelo Random Forest con índices y bandas espectrales

Fold	Accuracy	Precision	Recall	F1-Score
1	0.9400	0.9405	0.9400	0.9400
2	0.9243	0.9246	0.9243	0.9243
3	0.9386	0.9386	0.9386	0.9386
4	0.9429	0.9429	0.9429	0.9429
5	0.9157	0.9158	0.9157	0.9157
6	0.9314	0.9316	0.9314	0.9314
7	0.9586	0.9586	0.9586	0.9586
8	0.9429	0.9432	0.9429	0.9428
9	0.9414	0.9415	0.9414	0.9414
10	0.9500	0.9506	0.9500	0.9500
Promedio \pm Desv. Est.	0.9386 \pm 0.0116	0.9388 \pm 0.0117	0.9386 \pm 0.0116	0.9386 \pm 0.0116

A.3 Resultados del modelo Random Forest con índices espectrales

Tabla A.3: Resultados del modelo Random Forest con índices espectrales

Fold	Accuracy	Precision	Recall	F1-Score
1	0.9271	0.9282	0.9271	0.9271
2	0.9129	0.9141	0.9129	0.9128
3	0.9314	0.9315	0.9314	0.9314
4	0.9357	0.9361	0.9357	0.9357
5	0.9014	0.9017	0.9014	0.9014
6	0.9186	0.9198	0.9186	0.9185
7	0.9343	0.9344	0.9343	0.9343
8	0.9286	0.9297	0.9286	0.9285
9	0.9286	0.9288	0.9286	0.9286
10	0.9300	0.9306	0.9300	0.9300
Promedio \pm Desv. Est.	0.9249 \pm 0.0102	0.9255 \pm 0.0101	0.9249 \pm 0.0102	0.9248 \pm 0.0102

Anexo B. Resultados del modelo SVM

B.1 Resultados del modelo SVM con bandas espectrales

Tabla B.1: Resultados del modelo SVM con bandas espectrales

Fold	Accuracy	Precision	Recall	F1-Score
1	0.9414	0.9425	0.9414	0.9414
2	0.9271	0.9273	0.9271	0.9271
3	0.9457	0.9458	0.9457	0.9457
4	0.9571	0.9573	0.9571	0.9571
5	0.9243	0.9246	0.9243	0.9243
6	0.9357	0.9363	0.9357	0.9357
7	0.9471	0.9471	0.9471	0.9471
8	0.9443	0.9446	0.9443	0.9443
9	0.9343	0.9344	0.9343	0.9343
10	0.9529	0.9530	0.9529	0.9529
Promedio \pm Desv. Est.	0.9410 \pm 0.0101	0.9413 \pm 0.0100	0.9410 \pm 0.0101	0.9410 \pm 0.0101

B.2 Resultados del modelo SVM con índices y bandas espectrales

Tabla B.2: Resultados del modelo SVM con índices y bandas espectrales

Fold	Accuracy	Precision	Recall	F1-Score
1	0.9429	0.9438	0.9429	0.9428
2	0.9314	0.9321	0.9314	0.9314
3	0.9471	0.9472	0.9471	0.9471
4	0.9486	0.9488	0.9486	0.9486
5	0.9214	0.9215	0.9214	0.9214
6	0.9357	0.9365	0.9357	0.9357
7	0.9443	0.9444	0.9443	0.9443
8	0.9443	0.9449	0.9443	0.9443
9	0.9414	0.9416	0.9414	0.9414
10	0.9514	0.9518	0.9514	0.9514
Promedio \pm Desv. Est.	0.9409 \pm 0.0086	0.9413 \pm 0.0085	0.9409 \pm 0.0086	0.9408 \pm 0.0086

B.3 Resultados del modelo SVM con índices espectrales

Tabla B.3: Resultados del modelo SVM con índices espectrales

Fold	Accuracy	Precision	Recall	F1-Score
1	0.9300	0.9319	0.9300	0.9299
2	0.9143	0.9162	0.9143	0.9142
3	0.9229	0.9237	0.9229	0.9228
4	0.9400	0.9409	0.9400	0.9400
5	0.9057	0.9058	0.9057	0.9057
6	0.9200	0.9231	0.9200	0.9199
7	0.9400	0.9401	0.9400	0.9400
8	0.9300	0.9316	0.9300	0.9299
9	0.9271	0.9279	0.9271	0.9271
10	0.9371	0.9383	0.9371	0.9371
Promedio ± Desv. Est.	0.9267 ± 0.0107	0.9280 ± 0.0106	0.9267 ± 0.0107	0.9267 ± 0.0107

Anexo C. Resultados del modelo Regresión Logística

C.1 Resultados del modelo Regresión Logística con bandas espectrales

Tabla C.1: Resultados del modelo Regresión Logística con bandas espectrales

Fold	Accuracy	Precision	Recall	F1-Score
1	0.9329	0.9333	0.9329	0.9328
2	0.9300	0.9303	0.9300	0.9300
3	0.9443	0.9445	0.9443	0.9443
4	0.9400	0.9400	0.9400	0.9400
5	0.9143	0.9145	0.9143	0.9143
6	0.9257	0.9261	0.9257	0.9257
7	0.9429	0.9440	0.9429	0.9428
8	0.9343	0.9343	0.9343	0.9343
9	0.9300	0.9300	0.9300	0.9300
10	0.9486	0.9488	0.9486	0.9486
Promedio \pm Desv. Est.	0.9343 \pm 0.0096	0.9346 \pm 0.0097	0.9343 \pm 0.0096	0.9343 \pm 0.0096

C.2 Resultados del modelo Regresión Logística con índices y bandas espectrales

Tabla C.2: Resultados del modelo Regresión Logística con índices y bandas espectrales

Fold	Accuracy	Precision	Recall	F1-Score
1	0.9400	0.9404	0.9400	0.9400
2	0.9286	0.9287	0.9286	0.9286
3	0.9371	0.9373	0.9371	0.9371
4	0.9543	0.9543	0.9543	0.9543
5	0.9243	0.9243	0.9243	0.9243
6	0.9329	0.9333	0.9329	0.9328
7	0.9443	0.9444	0.9443	0.9443
8	0.9400	0.9401	0.9400	0.9400
9	0.9329	0.9329	0.9329	0.9329
10	0.9500	0.9502	0.9500	0.9500
Promedio \pm Desv. Est.	0.9384 \pm 0.0089	0.9386 \pm 0.0089	0.9384 \pm 0.0089	0.9384 \pm 0.0089

C.3 Resultados del modelo Regresión Logística con índices espectrales

Tabla C.3: Resultados del modelo Regresión Logística con índices espectrales

Fold	Accuracy	Precision	Recall	F1-Score
1	0.9243	0.9247	0.9243	0.9243
2	0.9129	0.9138	0.9129	0.9128
3	0.9286	0.9287	0.9286	0.9286
4	0.9357	0.9357	0.9357	0.9357
5	0.9000	0.9000	0.9000	0.9000
6	0.9200	0.9211	0.9200	0.9199
7	0.9443	0.9443	0.9443	0.9443
8	0.9243	0.9246	0.9243	0.9243
9	0.9243	0.9243	0.9243	0.9243
10	0.9357	0.9360	0.9357	0.9357
Promedio ± Desv. Est.	0.9250 ± 0.0119	0.9253 ± 0.0118	0.9250 ± 0.0119	0.9250 ± 0.0119

Anexo D. Resultados del modelo XGBoost

D.1 Resultados del modelo XGBoost con bandas espectrales

Tabla D.1: Resultados del modelo XGBoost con bandas espectrales

Fold	Accuracy	Precision	Recall	F1-Score
1	0.9386	0.9389	0.9386	0.9386
2	0.9229	0.9234	0.9229	0.9228
3	0.9329	0.9329	0.9329	0.9329
4	0.9386	0.9387	0.9386	0.9386
5	0.9186	0.9188	0.9186	0.9186
6	0.9329	0.9335	0.9329	0.9328
7	0.9571	0.9572	0.9571	0.9571
8	0.9400	0.9401	0.9400	0.9400
9	0.9400	0.9401	0.9400	0.9400
10	0.9471	0.9472	0.9471	0.9471
Promedio \pm Desv. Est.	0.9369 \pm 0.0105	0.9371 \pm 0.0104	0.9369 \pm 0.0105	0.9368 \pm 0.0105

D.2 Resultados del modelo XGBoost con índices y bandas espectrales

Tabla D.2: Resultados del modelo XGBoost con índices y bandas espectrales

Fold	Accuracy	Precision	Recall	F1-Score
1	0.9429	0.9438	0.9429	0.9428
2	0.9229	0.9232	0.9229	0.9228
3	0.9371	0.9374	0.9371	0.9371
4	0.9486	0.9486	0.9486	0.9486
5	0.9229	0.9231	0.9229	0.9228
6	0.9314	0.9319	0.9314	0.9314
7	0.9514	0.9520	0.9514	0.9514
8	0.9343	0.9350	0.9343	0.9343
9	0.9457	0.9459	0.9457	0.9457
10	0.9500	0.9502	0.9500	0.9500
Promedio \pm Desv. Est.	0.9387 \pm 0.0100	0.9391 \pm 0.0097	0.9387 \pm 0.0100	0.9387 \pm 0.0100

D.3 Resultados del modelo XGBoost con índices espectrales

Tabla D.3: Resultados del modelo XGBoost con índices espectrales

Fold	Accuracy	Precision	Recall	F1-Score
1	0.9186	0.9204	0.9186	0.9185
2	0.9129	0.9153	0.9129	0.9127
3	0.9343	0.9345	0.9343	0.9343
4	0.9343	0.9345	0.9343	0.9343
5	0.8957	0.8961	0.8957	0.8957
6	0.9171	0.9198	0.9171	0.9170
7	0.9343	0.9344	0.9343	0.9343
8	0.9257	0.9266	0.9257	0.9257
9	0.9300	0.9306	0.9300	0.9300
10	0.9357	0.9367	0.9357	0.9357
Promedio ± Desv. Est.	0.9239 ± 0.0122	0.9249 ± 0.0119	0.9239 ± 0.0122	0.9238 ± 0.0122

Anexo E. Repositorio del proyecto y código fuente

En este anexo se presenta el enlace al repositorio del proyecto alojado en GitHub, que contiene el código fuente, los cuadernos (notebooks) y los scripts desarrollados durante las diferentes etapas del análisis, incluyendo el procesamiento de las Encuestas Multipropósito, la extracción y preprocesamiento de imágenes satelitales Landsat 8, así como la implementación de modelos supervisados y no supervisados y sus respectivos análisis. El repositorio tiene como propósito garantizar la transparencia y la reproducibilidad del estudio, permitiendo acceder al conjunto completo de recursos utilizados en la implementación y evaluación de los modelos.

Enlace: <https://github.com/christianveram/ExpansionUrbana>

9. Referencias

- [1] UN-Habitat, “The fundamentals of urbanization: Evidence base for policy making,” 2016. [Online]. Available: <https://unhabitat.org/the-fundamentals-of-urbanization-evidence-base-for-policy-making>
- [2] R. Jordán, L. Riffo, and A. Prado, *Desarrollo sostenible, urbanización y desigualdad en América Latina y el Caribe: Dinámicas y desafíos para el cambio estructural*, C. E. para América Latina y el Caribe (CEPAL), Ed., 9 2017. [Online]. Available: <https://www.cepal.org/es/publicaciones/42141-desarrollo-sostenible-urbanizacion-desigualdad-america-latina-caribe-dinamicas>
- [3] H. Al-Bilbisi, “Spatial monitoring of urban expansion using satellite remote sensing images: A case study of amman city, jordan,” *Sustainability*, vol. 11, 2019. [Online]. Available: <https://www.mdpi.com/2071-1050/11/8/2260>
- [4] A. M. Ferrero and I. Astorkiza, “Expansión urbana y sostenibilidad: una dicotomía difícil de conciliar,” *Revista Española de Control Externo*, Tech. Rep.
- [5] A. Asokan and J. Anitha, “Machine learning based image processing techniques for satellite image analysis -a survey,” in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 2019, pp. 119–124.
- [6] T. Kuchkorov, S. Urmanov, M. Kuvvatova, and I. Anvarov, “Satellite image formation and preprocessing methods,” in *2020 International Conference on Information Science and Communications Technologies (ICISCT)*, 2020, pp. 1–4.
- [7] USGS, “What is remote sensing and what is it used for?” 3 2025. [Online]. Available: <https://www.usgs.gov/faqs/what-remote-sensing-and-what-it-used>
- [8] N. R. Canada, “Fundamentals of remote sensing,” 2025. [Online]. Available: https://natural-resources.canada.ca/sites/nrcan/files/earthsciences/pdf/resource/tutorial/fundam/pdf/fundamentals_e.pdf
- [9] M. de Transportes y Movilidad Sostenible, “Programa landsat,” <https://pnt.ign.es/programa-landsat>, 2025, consultado el 22 de diciembre de 2025.
- [10] NASA, “History.” [Online]. Available: <https://landsat.gsfc.nasa.gov/satellites/history/>
- [11] NASA, “Landsat 3,” 1978. [Online]. Available: <https://landsat.gsfc.nasa.gov/satellites/landsat-3/>
- [12] NASA, “Timeline.” [Online]. Available: <https://landsat.gsfc.nasa.gov/satellites/timeline>

- [13] NASA, “Landsat 8,” 2013. [Online]. Available: <https://landsat.gsfc.nasa.gov/satellites/landsat-8>
- [14] T. D. Acharya and I. Yang, “Exploring landsat 8,” *International Journal of IT, Engineering and Applied Sciences Research (IJIEASR)*, vol. 4, pp. 4–10, 2015. [Online]. Available: https://www.researchgate.net/profile/Tri-Acharya/publication/311901147_Exploring_Landsat_8/links/589c0de6458515e5f4549e58/Exploring-Landsat-8.pdf
- [15] R. H. Puerta Tuesta, J. A. Iannacone Oliver, and M. E. Reátegui Inga, “Una revisión del uso de imágenes sentinel-2 para el monitoreo de la cobertura boscosa a nivel global,” *Ingeniería y competitividad*, vol. 25, no. 3, pp. e–30 612 708, 2023. [Online]. Available: <https://doi.org/10.25100/iyc.v25i3.12708>
- [16] C. Justice, E. Vermote, J. Townshend, R. Defries, D. Roy, D. Hall, V. Salomonson, J. Privette, G. Riggs, A. Strahler, W. Lucht, R. Myneni, Y. Knyazikhin, S. Running, R. Nemani, Z. Wan, A. Huete, W. van Leeuwen, R. Wolfe, L. Giglio, J. Muller, P. Lewis, and M. Barnsley, “The moderate resolution imaging spectroradiometer (modis): land remote sensing for global change research,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 4, pp. 1228–1249, 1998.
- [17] U. Cossio, E. d. J. Ceballos Guerra, and C. Cuartas Vélez, “Aplicación del tratamiento de imágenes satelitales en el estudio geológico de un sector del suroeste de antioquia,” *Boletín de Ciencias de la Tierra*, no. 14, pp. 79–93, 2002.
- [18] S. D. R., P. D. Shenoy, and V. K. R., “Remote sensing satellite image processing techniques for image classification: A comprehensive survey,” *International Journal of Computer Applications*, vol. 161, pp. 24–37, 3 2017. [Online]. Available: <https://ijcaonline.org/archives/volume161/number11/27193-2017913306/>
- [19] T. V. Tran, R. Reef, and X. Zhu, “A review of spectral indices for mangrove remote sensing,” *Remote Sensing*, vol. 14, 2022. [Online]. Available: <https://www.mdpi.com/2072-4292/14/19/4868>
- [20] Y. Zha, J. Gao, and S. Ni, “Use of normalized difference built-up index in automatically mapping urban areas from tm imagery,” *International Journal of Remote Sensing*, vol. 24, pp. 583–594, 2003. [Online]. Available: <https://doi.org/10.1080/01431160304987>
- [21] R. H. Lozano and N. P. Pavon, “Índices para el monitoreo de cuerpos de agua usando sensores remotos,” *Acta Universitaria*, vol. 34, p. e3814, 2024. [Online]. Available: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0188-62662024000100108&lng=es&nrm=iso

- [22] A. Asokan and J. Anitha, “Change detection techniques for remote sensing applications: a survey,” *Earth Science Informatics*, vol. 12, pp. 143–160, 2019. [Online]. Available: <https://doi.org/10.1007/s12145-019-00380-5>
- [23] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, 1st ed. John Wiley & Sons, 1973.
- [24] J. A. Richards, *Remote Sensing Digital Image Analysis*, 6th ed. Springer Nature Switzerland AG, 2022. [Online]. Available: <https://doi.org/10.1007/978-3-030-82327-6>
- [25] J. A. Richards and X. Jia, *Feature Reduction*. Springer Berlin Heidelberg, 2006, pp. 267–294. [Online]. Available: https://doi.org/10.1007/3-540-29711-1_10
- [26] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. Pearson, 2018.
- [27] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, pp. 62–66, 1979.
- [28] C. H. Li and C. K. Lee, “Minimum cross entropy thresholding,” *Pattern Recognition*, vol. 26, pp. 617–625, 1993. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/003132039390115D>
- [29] J.-C. Yen, F.-J. Chang, and S. Chang, “A new criterion for automatic multilevel thresholding,” *IEEE Transactions on Image Processing*, vol. 4, pp. 370–378, 1995.
- [30] V. N. Vapnik, *The nature of statistical learning theory*. Springer-Verlag New York, Inc., 1995.
- [31] S. E. Jozdani, B. A. Johnson, and D. Chen, “Comparing deep neural networks, ensemble classifiers, and support vector machine algorithms for object-based urban land use/land cover classification,” *Remote Sensing*, vol. 11, 2019. [Online]. Available: <https://www.mdpi.com/2072-4292/11/14/1713>
- [32] M. A. Friedl and C. E. Brodley, “Decision tree classification of land cover from remotely sensed data,” *Remote Sensing of Environment*, vol. 61, pp. 399–409, 1997. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425797000497>
- [33] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, “Random forests for land cover classification,” *Pattern Recognition Letters*, vol. 27, pp. 294–300, 2006. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167865505002242>
- [34] K. C. Seto and R. K. Kaufmann, “Using logit models to classify land cover and land-cover change from landsat thematic mapper,” *International Journal of Remote Sensing*, vol. 26, pp. 563–577, 2005. [Online]. Available: <https://doi.org/10.1080/01431160512331299270>

- [35] Z. Shao, M. N. Ahmad, and A. Javed, “Comparison of random forest and xgboost classifiers using integrated optical and sar features for mapping urban impervious surface,” *Remote Sensing*, vol. 16, 2024. [Online]. Available: <https://www.mdpi.com/2072-4292/16/4/665>
- [36] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Information Processing & Management*, vol. 45, pp. 427–437, 2009. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306457309000259>
- [37] M. Gonzalvo-Cirac and M. V. Roqué-Sánchez, “Demografía, población vulnerable y bioética,” *Persona y Bioética*, vol. 19, no. 2, pp. 220–237, nov 2015. [Online]. Available: <https://personaybioetica.unisabana.edu.co/index.php/personaybioetica/article/view/5319>
- [38] J. López-Jiménez, “Las densidades y la forma urbana resultante en el planeamiento urbanístico para la vida e interacción social: el caso de alicante,” *Ciudad y Territorio Estudios Territoriales*, vol. 54, no. 211, p. 37–56, sep. 2021. [Online]. Available: <https://recyt.fecyt.es/index.php/CyTET/article/view/91429>
- [39] B. Lamy, J. Ramirez, and A. Sanchez, “Migración interna y sus efectos en el crecimiento urbano del municipio de querétaro,” *Quivera Revista de Estudios Territoriales*, vol. 21, p. 49, 10 2019.
- [40] F. R. Mendoza, “La gentrificación en los estudios urbanos: una exploración sobre la producción académica de las ciudades,” *Cadernos Metrópole*, vol. 18, pp. 697–719, 12 2016. [Online]. Available: <https://www.scielo.br/j/cm/a/wJPJJP9xn8NvcrPdY6scwSb/?format=pdf&lang=es>
- [41] M. Janoschka, “Gentrificación en España reloaded,” *Papers: Institut d’Estudis Regionals i Metropolitans de Barcelona*, no. 60, pp. 24–33, Jun. 2018. [Online]. Available: <https://ddd.uab.cat/record/194953>
- [42] J. A. M. Clemente, V. M. G. Antonio, and J. M. P. Abarca, “Efectos de la pobreza y de los factores sociodemográficos en la educación superior: un modelo probit aplicado a México,” *Nova Scientia*, vol. 10, pp. 539–568, 2018. [Online]. Available: https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-07052018000100539
- [43] DANE, “Metodologías oficiales y arreglos institucionales para la medición de la pobreza en Colombia,” Departamento Administrativo Nacional de Estadística y Departamento Nacional de Planeación, Tech. Rep., 5 2012. [Online]. Available: <https://colaboracion.dnp.gov.co/cdt/conpes/social/150.pdf>

- [44] D. de Metodología y Producción Estadística, “Colombia - medición de pobreza monetaria y desigualdad,” Departamento Administrativo Nacional de Estadística, Tech. Rep., 7 2024. [Online]. Available: <https://microdatos.dane.gov.co/index.php/catalog/874/related-materials>
- [45] S. Alkire and J. Foster, “Counting and multidimensional poverty measurement,” *Journal of Public Economics*, vol. 95, pp. 476–487, 2011. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0047272710001660>
- [46] Consejo Nacional de Política Económica y Social, “Metodologías oficiales y arreglos institucionales para la medición de la pobreza en colombia,” Departamento Nacional de Planeación, Bogotá, Colombia, Tech. Rep. 150, May 2012. [Online]. Available: <https://colaboracion.dnp.gov.co/CDT/Conpes/Economicos/150.pdf>
- [47] DANE, “Boletín técnico - medida de pobreza multidimensional municipal,” Departamento Administrativo Nacional de Estadística, Tech. Rep., 1 2020. [Online]. Available: https://www.dane.gov.co/files/investigaciones/condiciones_vida/pobreza/2018/informacion-censal/nota-metodologia-censal-pobreza-municipal-2018.pdf
- [48] N. Sgouropoulos, Q. Yao, and C. Yastremiz, “Matching a distribution by matching quantiles estimation,” *Journal of the American Statistical Association*, vol. 110, pp. 742–759, 4 2015.
- [49] D. A. N. de Estadística (DANE), “Boletín técnico: Encuesta multipropósito bogotá - cundinamarca (em) 2021,” DANE, Tech. Rep., 2021. [Online]. Available: https://www.dane.gov.co/files/investigaciones/multi/Boletin_EM_2021.pdf
- [50] Y. D. Wei and R. Ewing, “Urban expansion, sprawl and inequality,” *Landscape and Urban Planning*, vol. 177, pp. 259–265, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169204618303992>
- [51] C. H. Wheeler, “Urban decentralization and income inequality: Is sprawl associated with rising income segregation across neighborhoods?” FRB of St. Louis, Tech. Rep., 5 2006. [Online]. Available: <https://ssrn.com/abstract=904384>
- [52] P. Zhao, “The impact of urban sprawl on social segregation in beijing and a limited role for spatial planning,” *Tijdschrift voor Economische en Sociale Geografie*, vol. 104, pp. 571–587, 2013. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/1/tesg.12030>
- [53] S. L. Bahna and S. A. Conrad, “Annals of allergy, asthma & immunology,” *Annals of Allergy, Asthma & Immunology*, vol. 103, p. S2–S60, 9 2009. [Online]. Available: <https://www.clinicalkey.es/#!/browse/toc/1-s2.0-S1081120610X6046X/null/journalIssue>

- [54] C. Chasco Yrigoyen, “Econometría espacial aplicada a la predicción-extrapolación de datos microterritoriales,” Tesis doctoral, Universidad Autónoma de Madrid, Madrid, España, 2002, facultad de Ciencias Económicas y Empresariales, Instituto L.R. Klein. [Online]. Available: <http://hdl.handle.net/10486/4077>
- [55] J. P. Celemín, “Autocorrelación espacial e indicadores locales de asociación espacial. importancia, estructura y aplicación,” *Revista Universitaria de Geografía*, vol. 18, no. 1, pp. 11–31, 2009. [Online]. Available: https://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1852-42652009000100002
- [56] H. A. J. Duque, Juan Carlos Velásquez, “Infraestructura pública y precios de vivienda: una aplicación de regresión geográficamente ponderada en el contexto de precios hedónicos,” *Ecos de Economía*, 2011. [Online]. Available: <https://www.redalyc.org/articulo.oa?id=329027268005>
- [57] J. Wood, J. Dykes, A. Slingsby, and K. Clarke, “Interactive visual exploration of a large spatio-temporal dataset: Reflections on a geovisualization mashup,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, pp. 1176–1183, 11 2007. [Online]. Available: <https://doi.org/10.1109/tvcg.2007.70570>
- [58] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, “Google earth engine: Planetary-scale geospatial analysis for everyone,” *Remote Sensing of Environment*, vol. 202, pp. 18–27, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425717302900>
- [59] P. V. Project, “Folium: Python data. leaflet.js maps,” 2015.
- [60] P. T. Inc., “Collaborative data science,” 2015.
- [61] A. Arshad, J. Zulfiqar, M. H. Zaib, A. Khan, and M. J. Khan, “Mapping socioeconomic conditions using satellite imagery: A computer vision approach for developing countries,” *Journal of Economy and Technology*, vol. 1, pp. 144–163, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2949948823000100>
- [62] O. M. J. Sebastián, “Estimación de índice de pobreza multidimensional (ipm) en bogotá d.c. y algunas ciudades cercanas usando imágenes satelitales,” 2024. [Online]. Available: <https://repositorio.unal.edu.co/handle/unal/86359>
- [63] M. Aghlmand and G. Kaplan, “Monitoring urban expansion using remote-sensing data aided by google earth engine,” 2021. [Online]. Available: <https://api.semanticscholar.org/CorpusID:234025375>
- [64] M. Morales-Soto and F. Maturana-Miranda, “Análisis de patrones espaciales en la expansión urbana de ciudades intermedias. el caso de san fernando,

- chile,” *Cuadernos de Vivienda y Urbanismo*, vol. 12, 2019. [Online]. Available: <https://doi.org/10.11144/Javeriana.cvu12-24.apee>
- [65] M. Anduquia and J. Garcés, “Evaluación del efecto de la pandemia sobre la deforestación en Colombia por medio de imágenes de satélite y usando los datos de terra-i,” Ph.D. dissertation, Pontificia Universidad Javeriana Cali, 2024.
- [66] S. Zhong, M. Wang, Y. Zhu, Z. Chen, and X. Huang, “Urban expansion and the urban–rural income gap: Empirical evidence from China,” *Cities*, vol. 129, p. 103831, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0264275122002700>
- [67] F. G. Amazonas, “Documento de bases teóricas de algoritmo (atdb): Raisg-mapbiomas Colombia colección 2, apéndice Colombia – colección 2 de mapas anuales de cobertura y uso del suelo de Colombia,” Fundación Gaia Amazonas, Tech. Rep., 2024. [Online]. Available: <https://colombia.mapbiomas.org/wp-content/uploads/sites/3/2024/11/Apendice-1-Colombia-Coleccion-2.0.pdf>
- [68] U. G. Survey, “Landsat 9 data users handbook,” 11 2023. [Online]. Available: <https://www.usgs.gov/media/files/landsat-9-data-users-handbook>
- [69] E. S. Agency, “Sentinel-2 mission overview,” 2023. [Online]. Available: <https://documentation.dataspace.copernicus.eu/Data/SentinelMissions/Sentinel2.html>
- [70] USGS, “Landsat 7 scan line corrector-off (slc-off) gap-filled products: Phase one methodology,” 5 2003. [Online]. Available: <https://www.usgs.gov/media/files/landsat-7-slc-gap-filled-products-phase-one-methodology>
- [71] U. G. Survey, “Landsat 8 (18) data users handbook,” 11 2019. [Online]. Available: https://d9-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/media/files/LSDS-1574_Landsat8_Data_Users_Handbook.pdf
- [72] NASA and USGS, “Landsat 8-9 collection 2 level 2 science product guide,” 11 2023. [Online]. Available: <https://www.usgs.gov/media/files/landsat-8-9-collection-2-level-2-science-product-guide>
- [73] N. Flood, “Seasonal composite Landsat TM/ETM+ images using the medoid (a multi-dimensional median),” *Remote Sensing*, vol. 5, pp. 6481–6500, 2013. [Online]. Available: <https://www.mdpi.com/2072-4292/5/12/6481>
- [74] Y. Zhang, C. E. Woodcock, P. Arévalo, P. Olofsson, X. Tang, R. Stanimirova, E. Bullock, K. R. Tarrio, Z. Zhu, and M. A. Friedl, “A global analysis of the spatial and temporal variability of usable Landsat observations at the pixel scale,”

- Frontiers in Remote Sensing*, vol. Volume 3 - 2022, 2022. [Online]. Available: <https://www.frontiersin.org/journals/remote-sensing/articles/10.3389/frsen.2022.894618>
- [75] DANE, “Encuesta multipropósito (em),” 7 2022. [Online]. Available: <https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/encuesta-multiproposito>
- [76] SDP, “Formularios encuesta multipropósito,” 2021. [Online]. Available: <https://www.sdp.gov.co/gestion-estudios-estrategicos/informacion-estadisticas/encuesta-multiproposito/formularios>
- [77] S. de Planeación de Bogotá, “Visor de la encuesta multipropósito de bogota y la región.” [Online]. Available: <https://experience.arcgis.com/experience/dfa5a8a94d9547d1a4336e6975a13c0d/page/Indicadores?views=5.-UPZ>
- [78] M. Buchhorn, B. Smets, L. Bertels, B. D. Roo, M. Lesiv, N.-E. Tsendbazar, L. Li, and A. Tarko, “Copernicus global land service: Land cover 100m: version 3 globe 2015-2019: Product user manual,” 9 2020. [Online]. Available: <https://doi.org/10.5281/zenodo.3938963>
- [79] D. Zanaga, R. V. D. Kerchove, D. Daems, W. D. Keersmaecker, C. Brockmann, G. Kirches, J. Wevers, O. Cartus, M. Santoro, S. Fritz, M. Lesiv, M. Herold, N.-E. Tsendbazar, P. Xu, F. Ramoino, and O. Arino, “Esa worldcover 10 m 2021 v200,” 10 2022. [Online]. Available: <https://doi.org/10.5281/zenodo.7254221>
- [80] D. R. Roberts, V. Bahn, S. Ciuti, M. S. Boyce, J. Elith, G. Guillera-Arroita, S. Hauenstein, J. J. Lahoz-Monfort, B. Schröder, W. Thuiller, D. I. Warton, B. A. Wintle, F. Hartig, and C. F. Dormann, “Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure,” *Ecography*, vol. 40, pp. 913–929, 2017. [Online]. Available: <https://nsojournals.onlinelibrary.wiley.com/doi/abs/10.1111/ecog.02881>
- [81] Google, “Google colab,” 2024. [Online]. Available: <https://colab.research.google.com>
- [82] “Scikit-learn: Machine learning in python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 11 2011.
- [83] C. M. Souza, J. Z. Shimbo, M. R. Rosa, L. L. Parente, A. A. Alencar, B. F. T. Rudorff, H. Hasenack, M. Matsumoto, L. G. Ferreira, P. W. M. Souza-Filho, S. W. de Oliveira, W. F. Rocha, A. V. Fonseca, C. B. Marques, C. G. Diniz, D. Costa, D. Monteiro, E. R. Rosa, E. Vélez-Martin, E. J. Weber, F. E. B. Lenti, F. F. Paternost, F. G. C. Pareyn, J. V. Siqueira, J. L. Viera, L. C. F. Neto, M. M. Saraiva, M. H.

- Sales, M. P. G. Salgado, R. Vasconcelos, S. Galano, V. V. Mesquita, and T. Azevedo, “Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine,” *Remote Sensing*, vol. 12, 2020. [Online]. Available: <https://www.mdpi.com/2072-4292/12/17/2735>
- [84] Alcaldía Mayor de Bogotá D. C., “Superficie de vivienda de origen informal,” Secretaría Distrital del Hábitat, Bogotá, Colombia, Tech. Rep., Aug. 2023, subdirección de Información Sectorial. [Online]. Available: <https://observatoriohabitad.org/>
- [85] A. de Bogota, “Bogotá legaliza 32 nuevos barrios, asentamientos humanos de origen informal,” Alcaldia de Bogota, Tech. Rep., 12 2023. [Online]. Available: <https://bogota.gov.co/mi-ciudad/planeacion/bogota-legaliza-32-nuevos-barrios-que-n-acieron-de-manera-informal>
- [86] Properati. (2024, 9) ¿cuánto ha subido el precio de la vivienda en bogotá durante 2024? [Online]. Available: <https://blog.properati.com.co/precio-vivienda-bogota-durante-pri-mer-semester-2024>
- [87] P. Harris, “Chapinero’s second ‘renaissance’ as a bogotá real estate star,” *The City Paper*, 3 2016.
- [88] C. Osorio, “Salir de la miseria sin abrazar la gentrificación: el barrio san bernardo busca un futuro distinto para el centro de bogotá,” *El Pais*, 5 2025. [Online]. Available: <https://elpais.com/america-colombia/2025-05-25/salir-de-la-miseria-sin-abrazar-la-gentrificacion-el-barrio-san-bernardo-busca-un-futuro-distinto-para-el-centro-de-bogota.html>
- [89] B. P. U. Contreras, L. T. Oyola, Ángela Patricia Landaeta Vega, M. O. González, S. L. M. Mayorga, A. R. Pisco, J. A. S. Villa, J. P. C. Torres, and F. R. M. Ospina, “Anuario de movimientos internacionales,” DANE, Tech. Rep., 2022. [Online]. Available: https://www.dane.gov.co/files/investigaciones/poblacion/migracion/Anuario_21_movimientos-internacionales.pdf
- [90] F. Bogota, “Pobreza en bogotá: más allá de la pandemia,” *El Tiempo*, 5 2021. [Online]. Available: <https://www.eltiempo.com/bogota/pobreza-en-bogota-mas-alla-de-la-pandemia-opinion-587088>
- [91] P. A. Avendaño, S. Sofía, and G. Castellanos, “Índice de Pobreza Multidimensional (IPM) en Bogotá en 2024,” Observatorio de Desarrollo Económico, Bogotá, Colombia, Tech. Rep., Apr. 2025. [Online]. Available: https://observatorio.desarrolloeconomico.gov.co/wp-content/uploads/2025/05/ne_ipm2024.pdf
- [92] Alcaldía de Bogotá, “Bogotá legaliza 32 nuevos barrios, asentamientos humanos de origen informal,” Alcaldía de Bogotá, Tech. Rep., Dec. 2023. [Online]. Available:

<https://bogota.gov.co/mi-ciudad/planeacion/bogota-legaliza-32-nuevos-barrios-que-nacieron-de-manera-informal>

- [93] A. Bernal Llanos, “Vivienda en Bogotá se renueva con proyectos en todas las localidades,” *Portal Bogotá*, Mar. 2018. [Online]. Available: <https://bogota.gov.co/mi-ciudad/habitat/metas-en-vivienda-social-y-ciudad-sostenible-por-la-alcaldia-en-el-201>
- [94] K. Barrero, “Más de 69 mil hogares en bogotá se han beneficiado con subsidios de vivienda,” *Blog de Bogota*, Tech. Rep., 7 2023. [Online]. Available: <https://bogota.gov.co/mi-ciudad/habitat/alcaldia-de-bogota-ha-entregado-mas-de-69-mil-subsidios-de-vivienda>
- [95] J. Cortés, “Bogotá tuvo el déficit de vivienda más bajo de los últimos 12 años,” *Alcaldía Mayor de Bogotá*, Bogotá, Colombia, Tech. Rep., Sep. 2015. [Online]. Available: <https://bogota.gov.co/mi-ciudad/habitat/bogota-tuvo-el-deficit-de-vivienda-mas-bajo-de-los-ultimos-12-anos>
- [96] A. N. Diaz, “En 14 años se duplicó el número de viviendas informales en bogotá: ¿a qué se debe este preocupante aumento?” *El Tiempo*, 8 2025.
- [97] GitHub, “About github pages,” 2025. [Online]. Available: <https://docs.github.com/en/pages/getting-started-with-github-pages/about-github-pages>