

El machine learning como herramienta de apoyo al seguimiento de la Política Pública de Productividad, Competitividad y Desarrollo Socioeconómico de Bogotá (2011-2031)

Jhon Alexander Hernandez Araque

Trabajo de grado para optar por el título de Magíster en Ciencia de Datos y Analítica

Directora

Mireya García

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

2026

Resumen

La evaluación de las políticas públicas distritales presenta limitaciones cuando depende únicamente de métodos tradicionales de seguimiento, particularmente frente a la heterogeneidad territorial y a las relaciones no lineales entre variables socioeconómicas. El presente escrito analiza comparativamente modelos de aprendizaje automático supervisado y no supervisado, con el fin de valorar su utilidad como herramientas complementarias de apoyo al seguimiento y análisis de indicadores de la Política Pública de Productividad, Competitividad y desarrollo Socioeconómico de Bogotá (2011–2031), en variables estratégicas como empleo, productividad, dinamismo empresarial y pobreza multidimensional por localidad. Los resultados muestran desempeños distintos según el indicador analizado. En el caso de Random Forest destacó en pobreza por localidad, la regresión logística en dinamismo empresarial, K-means en agrupamientos sectoriales y SVR en desocupación, mientras que el MLP presentó limitaciones asociadas al tamaño muestral.

Palabras clave: machine learning, políticas públicas, desarrollo socioeconómico, seguimiento institucional, Bogotá.

Abstract

The evaluation of district public policies faces limitations when it relies solely on traditional monitoring methods, particularly given territorial heterogeneity and the nonlinear relationships between socioeconomic variables. In this regard, this paper comparatively analyzes supervised and unsupervised machine learning models to assess their usefulness as complementary tools for monitoring and analyzing indicators of Bogotá's Public Policy on Productivity, Competitiveness, and Socioeconomic Development (2011–2031), focusing on strategic variables such as employment, productivity, business dynamism, and multidimensional poverty by locality. The results show different performances depending on the indicator analyzed. Random Forest stood out for poverty by locality, logistic regression for business dynamism, K-means for sectoral clustering, and SVR for unemployment, while MLP showed limitations related to sample size.

Keywords: machine learning, public policy, economic development, evaluation, Bogotá.

Tabla de Contenido

Introducción	10
Justificación	12
Objetivos	15
Objetivo General	15
Objetivos Específicos.....	15
Planteamiento del Problema	16
Evaluación de Políticas Públicas de Desarrollo Económico.....	22
Métodos Tradicionales de Evaluación de Políticas Públicas	27
Evaluación de Política Pública en Bogotá	30
Machine Learning Aplicado a la Evaluación de Políticas Públicas.....	33
Ventajas y Desafíos de Machine Learning en Evaluación de Políticas Públicas	36
Casos de Estudio de Machine Learning en Desarrollo Socioeconómico	41
Marco Teórico.....	44
Desarrollo y Crecimiento Económico: Análisis Teórico desde un Enfoque del Desarrollo.....	44
Productividad y Competitividad	48
Modelos de Machine Learning en Política Pública	60
Modelos de Machine Learning para Evaluar Política.....	62
Metodología y Aplicación Metodológica	66
Política Pública de Productividad, Competitividad y Desarrollo Socioeconómico de Bogotá	73

Resultados	80
Aprendizaje Supervisado	80
Regresión Lineal	80
Regresión Logística	86
Support Vector Regression SVR	93
Árbol de Decisiones, Random Forest y Gradient Boosting.....	103
Aprendizaje No Supervisado	113
K-means	113
Redes Neuronales MLP	115
Discusión.....	123
Conclusiones	134
Recomendaciones	143
Referencias Bibliográficas	144

Lista de Tablas

Tabla 1 <i>Elementos Centrales del Proceso de Evaluación de Políticas Públicas en Bogotá</i>	30
Tabla 2 <i>Tipos de Evaluación en Política Pública</i>	32
Tabla 3 <i>Objetivo de los Tipos de Evaluación</i>	32
Tabla 4 <i>Modelos de Machine Learning en Políticas Públicas</i>	62
Tabla 5 <i>Modelos de Machine Learning para Evaluar Política Pública</i>	70
Tabla 6 <i>Marco Conceptual de la Actualización del a Política Pública</i>	74
Tabla 7 <i>PIB, Ocupación y Tasa de Producción Per Capita 2005 -2025</i>	81
Tabla 8 <i>Datos</i>	83
Tabla 9 <i>Métricas Regresión Lineal</i>	84
Tabla 10 <i>Creación Neta de Empresas: Creadas-canceladas</i>	87
Tabla 11 <i>Empresas Vigentes</i>	87
Tabla 12 <i>Métricas Regresión Logística</i>	91
Tabla 13 <i>Clasificación de Rendimiento (alto, Medio, Bajo)</i>	92
Tabla 14 <i>Datos</i>	94
Tabla 15 <i>Ocupadas Real-predicho 2011-2025</i>	97
Tabla 16 <i>Desocupadas Real y Predicho 2011 2025</i>	98
Tabla 17 <i>Gini Real y Predicho</i>	100
Tabla 18 <i>Métricas SVR</i>	102
Tabla 19 <i>Base de Datos para Árbol de Decisiones, Random Forest y Gradient Boosting</i>	104
Tabla 20 <i>Métricas de Árbol de Decisión, Random Forest, Gradient Boosting</i>	111
Tabla 21 <i>Descripción de Clúster</i>	114
Tabla 22 <i>Especificación de Hiperparámetros del Modelo MLP</i>	117

Tabla 23 <i>Comparación de Modelos de Machine Learning Implementados</i>	121
---	-----

Lista de Figuras

Figura 1 <i>Tasa de Crecimiento del PIB entre 2019 y 2024 Colombia y Bogotá</i>	16
Figura 2 <i>Índice de Pobreza Multidimensional de la Universidad de Oxford</i>	58
Figura 3 <i>Procedimiento para el Procesamiento de Datos</i>	72
Figura 4 <i>Objetivos de la Política Pública de Productividad, Competitividad y Desarrollo Socioeconómico</i>	76
Figura 5 <i>Avance Acumulado de la Política Pública 2023–2025</i>	78
Figura 6 <i>Comparación del Avance Acumulado por Objetivos Estratégicos</i>	79
Figura 7 <i>PIB Bogotá: Serie Histórica y Predicción (regresión Lineal)</i>	82
Figura 8 <i>PIB de Bogotá y Proyección 2026–2028</i>	82
Figura 9 <i>Regresión Lineal: Valores Reales vs. Predichos</i>	85
Figura 10 <i>Desempeño Empresarial Actual-predicho (regresión Logística)</i>	90
Figura 11 <i>Desempeño Empresarial por Localidad en Bogotá 2023–2025</i>	92
Figura 12 <i>SVR: Personas Ocupadas Reales y Predichas</i>	97
Figura 13 <i>SVR: Personas Desocupadas Reales y Predichas</i>	99
Figura 14 <i>Índice de Gini: Valores Reales y Predichos (SVR)</i>	101
Figura 15 <i>Árbol de Decisión para la Pobreza en Bogotá</i>	106
Figura 16 <i>Importancia de Variables — Random Forest</i>	107
Figura 17 <i>Random Forest: Pobreza Real vs. Predicha</i>	108
Figura 18 <i>Gradient Boosting: Pobreza Real y Predicha</i>	109
Figura 19 <i>Importancia de Variables — Gradient Boosting</i>	110
Figura 20 <i>Comparación de Modelos Supervisados</i>	112
Figura 21 <i>Método del Codo — K-means</i>	113

Figura 22 <i>Clústeres Sectoriales de Bogotá — K-means</i>	114
Figura 23	119
Figura 24 <i>MLP (64,32): PIB Real y Predicho — Bogotá 2008–2025</i>	120

Introducción

Las políticas públicas constituyen uno de los principales instrumentos mediante los cuales el Estado interviene para atender problemáticas y orientar el desarrollo de los territorios. En este contexto, Bogotá ocupa un lugar central, tanto por su peso económico y poblacional como por su incidencia en la actividad productiva, el empleo, la inversión y las decisiones institucionales que influyen en el desarrollo económico de la ciudad y del país.

Sin embargo, a pesar del dinamismo, aún persisten retos en materia de pobreza, desigualdad, desempleo, informalidad y disparidades en las localidades. De esta forma, se surge la Política Pública de Productividad, Competitividad y Desarrollo Socioeconómico de Bogotá, formulada inicialmente en 2011 y actualizada posteriormente mediante el documento en CONPES D.C. 25 en el año 2022, cuyo propósito se orienta a fortalecer el tejido productivo, promover la inclusión productiva, aprovechar las economías de aglomeración, articular actores y contribuir al mejoramiento de la calidad de vida de la población.

Como resultado de la actualización de la política pública, se reconocen dificultades históricas asociadas a la ausencia de un plan de acción y de instrumentos robustos de seguimiento y análisis, lo que ha limitado la posibilidad de valorar con claridad sus avances y resultados.

Por esta razón, surge la necesidad de explorar herramientas complementarias como el machine learning, que permiten identificar patrones, modelar relaciones entre variables y generar aproximaciones predictivas útiles para el análisis de las políticas públicas.

A partir de lo anterior, el estudio se orienta por la siguiente pregunta, ¿Qué modelos de aprendizaje automático supervisado y no supervisado ofrecen mayor capacidad predictiva y explicativa para apoyar el seguimiento y análisis de indicadores de la Política Pública de

Productividad, Competitividad y Desarrollo Socioeconómico de Bogotá 2011 a 2031, en sus indicadores estratégicos de empleo, productividad, dinamismo empresarial y pobreza multidimensional desagregada por localidad?

Esta pregunta se articula con el objetivo general de evaluar comparativamente la capacidad predictiva y explicativa de modelos de machine learning, para valorar su utilidad como herramientas de apoyo al seguimiento y análisis de la política pública de desarrollo económico en Bogotá 2011–2031.

El documento se desarrolló bajo un enfoque cuantitativo y aplicado, apoyado tanto en la revisión documental de la política pública como en la construcción de bases de datos provenientes de fuentes oficiales. Entre estas se encuentran el DANE, la Secretaría Distrital de Desarrollo Económico, la Secretaría Distrital de Planeación, el Observatorio de Desarrollo Económico de Bogotá y otras fuentes institucionales relevantes para el problema de investigación.

De esta forma, se implementan distintos modelos de machine learning, tanto supervisados como no supervisados, entre ellos regresión lineal, regresión logística, Support Vector Regression (SVR), Árbol de Decisión, Random Forest, Gradient Boosting, K-means y una red neuronal MLP, con el propósito de comparar su desempeño frente a indicadores estratégicos del desarrollo económico y valorar su utilidad como herramientas de apoyo al seguimiento y análisis institucional de la Política Pública de Productividad, Competitividad y Desarrollo Socioeconómico de Bogotá 2011–2031.

Justificación

Las nociones de desarrollo en los Estados han sido asimiladas a la mejora de condiciones sociales que involucra directamente en establecer un progreso en la calidad de vida de los ciudadanos en los territorios, sin embargo, las visiones de desarrollo no son independientes a la conceptualización de crecimiento económico y aumento de la economía del país. Frente a lo anterior, el crecimiento económico permite establecer condiciones óptimas para garantizar procesos de desarrollo en un país, por lo que las políticas de Estado han reorientado el crecimiento económico e integrado con los aspectos de desarrollo.

La implementación de acciones para el desarrollo económico en Bogotá reviste gran importancia debido a los retos estructurales que enfrenta la ciudad en materia de empleo, productividad, competitividad, pobreza y desigualdad en sus diferentes localidades.

En este contexto, la evaluación de las políticas públicas constituye una herramienta fundamental para orientar la toma de decisiones, optimizar la asignación de recursos y validar un impacto real efectivo en la calidad de vida de la población Bogotana.

La implementación de modelos de machine learning puede fortalecer la formulación, implementación y evaluación de las acciones gubernamentales, al permitir análisis basados en datos para una toma de decisiones más informada.

En particular los modelos predictivos de machine learning, se utilizan para apoyar las decisiones humanas con el objetivo de mejorar la eficiencia del uso de los recursos, la eficacia de las intervenciones y la equidad de los resultados (Amarasinghe et al., 2023, p. 2).

A pesar de la limitación bibliográfica sobre la implementación de estas herramientas en las políticas públicas, se ha logrado determinar que la aplicación de los modelos machine

learning en la evaluación de políticas públicas ha sido integrada en diversos sectores, incluyendo salud, educación e incluso proyectos de seguridad ciudadana.

No obstante, esta integración del machine learning en la evaluación de políticas también presenta desafíos como la transparencia, el sesgo y la privacidad de los datos, que deben abordarse para garantizar una aplicación ética que legitime el ejercicio (Chy y Buadi, 2024).

En este contexto, los sistemas de machine learning están apoyando cada vez más las decisiones políticas de alto riesgo en áreas como la justicia penal, la educación, la atención sanitaria y los servicios sociales (Amarasinghe et al., 2023, p. 2).

En particular las herramientas de machine learning en la evaluación de las políticas públicas y en concreto para determinar los efectos del desarrollo económico de ciudades como Bogotá, pueden contribuir en la valoración de indicadores, predicciones de eventos en variables como la variación del PIB, la pobreza, la desigualdad y el dinamismo empresarial en las diferentes localidades.

Adicionalmente, la pertinencia de este escrito se enmarca considerando que infortunadamente existe una limitada exploración académica que haya registro, en la que se realice la articulación e implementación de modelos de aprendizaje automático para la evaluación de una política pública de desarrollo económico específica, con comparación sistemática de múltiples modelos y desagregación territorial por localidades, lo que omite el potencial analítico de estas herramientas para valorar los impactos diferenciados de las acciones gubernamentales a nivel territorial.

Por lo anterior es importante resaltar que una toma de decisiones colaborativa para los asuntos públicos del Estado, el machine learning puede tener un profundo impacto en los procesos para la toma de decisiones políticas (Amarasinghe et al., 2023).

En consecuencia, el estudio de la implementación de modelos de machine learning en la evaluación de políticas públicas resulta pertinente para analizar la evolución del desarrollo económico en la ciudad de Bogotá y valorar los impactos de dicha implementación, con el propósito de establecer acciones de mejora que contribuyan al crecimiento económico y a la reducción de las inequidades sociales, a fin de configurar medios que permitan garantizar condiciones óptimas para los habitantes de la ciudad.

Es entonces como el presente escrito explora la necesidad de implementar herramientas analíticas que fortalezcan la evaluación de la política pública Productividad, Competitividad y Desarrollo Socioeconómico de Bogotá, en lo particular porque se requiere establecer alternativas metodológicas integrales a las herramientas tradicionales.

Esto como un complemento que permita aportar al análisis, la identificación de patrones y la comprensión de indicadores estratégicos para comprender los avances reales de la implementación de políticas públicas.

Objetivos

Objetivo General

Evaluar comparativamente la capacidad predictiva y explicativa de modelos de machine learning supervisados y no supervisados, aplicados a indicadores estratégicos de empleo, productividad, dinamismo empresarial y pobreza multidimensional, para valorar su utilidad como herramientas complementarias de apoyo al seguimiento y análisis de la Política Pública de Productividad, Competitividad y Desarrollo Socioeconómico de Bogotá (2011–2031).

Objetivos Específicos

Analizar la Política Pública de Productividad, Competitividad y Desarrollo Socioeconómico de Bogotá en el marco de sus indicadores estratégicos.

Identificar los modelos de machine learning pertinentes para el seguimiento y análisis de políticas públicas de desarrollo económico.

Aplicar modelos de machine learning a indicadores de empleo, productividad, dinamismo empresarial y pobreza multidimensional para el caso de Bogotá, a partir de fuentes oficiales y de la desagregación territorial disponible.

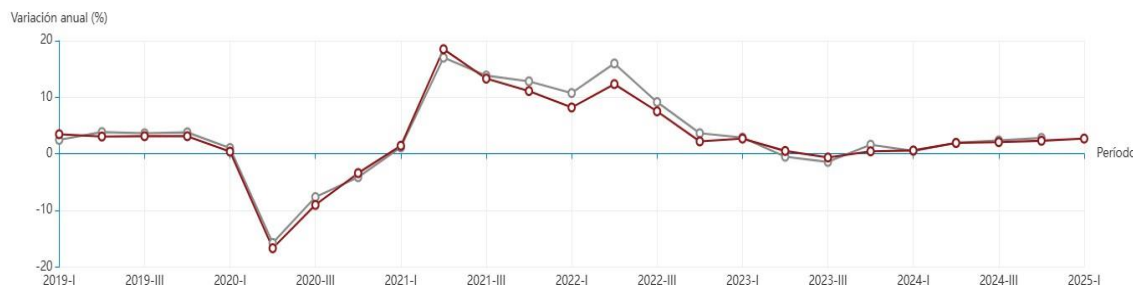
Comparar el desempeño de los modelos implementados mediante métricas de ajuste, para valorar su utilidad como herramientas de apoyo al seguimiento de la política pública.

Planteamiento del Problema

Figura 1

Tasa de Crecimiento del PIB entre 2019 y 2024 Colombia y Bogotá

Tasa de crecimiento anual, por trimestre, del PIB entre 2019 y 2024 Colombia y Bogotá



Nota. Fuente: Cámara de Comercio de Bogotá, con base en información del DANE, 28 de abril de 2025..

Bogotá se mantiene como la principal economía urbana del país, con una participación de aproximadamente el 25 % del PIB nacional (DANE, 2025), consolidándose además como uno de los centros metropolitanos más grandes de América Latina. Adicionalmente, la ciudad ha mantenido un fuerte mercado de consumo y continúa siendo un punto focal para los avances tecnológicos y comunicaciones a nivel local y regional (Aguilera et al., 2020).

No obstante, a pesar del crecimiento y las dinámicas de desarrollo en la región, Bogotá aún enfrenta retos en temas como el desarrollo de industrias, comercio, pobreza, desigualdad, y desempleo. Es importante enmarcar que las acciones para reducir estas disparidades en los indicadores de las desigualdades sociales y el crecimiento heterogéneo en la ciudad representa retos en materia de acciones públicas y gubernamentales para abordar estas dificultades que se reflejan en la ciudad.

Los desafíos de desarrollo económico en Bogotá incluyen altas tasas de desempleo, un aumento progresivo de la informalidad de forma generalizada y baja productividad por cuenta propia. Además, la ciudad enfrenta retos de crecimiento desigual en los sectores productivos y

particularmente en las localidades ya que se presenta segregación a las zonas periféricas ha profundizado en desigualdades sociales en la ciudad y naturalmente ocasionando efectos adversos en el desarrollo socioeconómico en la ciudad.

Estos retos se reflejan en las cifras. Según el Departamento Administrativo Nacional de Estadística DANE, en 2024 la incidencia de pobreza multidimensional en Bogotá se ubicó en 5,4 %, lo que representa un aumento de 1,8 puntos porcentuales frente al 3,6 % reportado en 2023 (DANE, 2025).

Asimismo, en el registro oficial del observatorio de salud de Bogotá evidenció que entre 2019 y 2020 este índice de pobreza multidimensional presentó incrementos, ubicándose en el 7,1% y 7,5%, respectivamente para el año 2019 y 2020 (Observatorio de salud de Bogotá, 2024)

Teniendo en cuenta los indicadores, el distrito en las últimas décadas ha diseñado políticas públicas orientadas al fomento del crecimiento económico, esto de forma integral realizando consideraciones de inclusión con perspectivas desarrollo social en la ciudad. Sin embargo, a pesar de estas acciones, en la ciudad persisten retos significativos en cuanto a la valoración de los impactos de las políticas y la validación de su efectividad para aplicación de las acciones gubernamentales

Por tal sentido, la evaluación de las políticas públicas se encuentra limitada por el uso de metodologías tradicionales que, en muchos casos, no logran capturar la complejidad de los fenómenos sociales y económicos.

Particularmente para el caso de Bogotá, el avance de las metodologías de evaluación, particularmente en torno a nociones clave como el desarrollo económico en los territorios, demanda la integración de modelos que permitan medir con mayor precisión la efectividad de los

resultados alcanzados, destacándose en este propósito la incorporación de herramientas como los modelos de machine learning.

No obstante, hay dilemas, debido a que resulta problemático que exista poca información sobre la aplicación de estas herramientas de machine learning en la formulación y evaluación de políticas públicas. Esta carencia, evidencia falencias en los métodos disponibles, los cuales deben ir a la vanguardia de las herramientas disponibles para lograr la incorporación de mecanismos que optimicen los procesos de interpretabilidad, visualización y validación de los impactos derivados de la implementación de las políticas públicas.

Por lo cual se ha pretendido casos de uso explicables en políticas públicas para alinear mejor los métodos de machine learning con las necesidades del mundo real, identificando brechas y proponiendo direcciones de investigación para mejorar el impacto sociedad de esta herramienta en las acciones públicas (Amarasinghe et al., 2020).

Estos efectos generan tensiones en materia de transparencia y rendición de cuentas, lo que hace indispensable el desarrollo de herramientas de inteligencia artificial explicables y modelos comprensibles para el público en general (Rehill y Biddle, 2023). Adicionalmente, el uso ético y la calidad de los datos se constituyen en factores esenciales para la aplicación efectiva del machine learning en este ámbito, tal como lo señalan marcos como las Cinco Seguros (Ávila, 2024).

En este marco, se requieren enfoques híbridos que integren la capacidad adaptativa del machine learning con la solidez teórica de la econometría, a fin de lograr un adecuado equilibrio entre el desempeño predictivo y la interpretabilidad de los modelos (Silahtaroglu, 2024).

A pesar de su potencial, la integración del machine learning en la evaluación de políticas públicas requiere una atención cuidadosa a la calidad de los datos, los sesgos y la

interpretabilidad, ya que estos factores afectan directamente la confiabilidad y la legitimidad de los resultados de las políticas y (Veale y Brass, 2018, p. 20).

La evaluación de política pública constituye un eje esencial para valorar la eficacia y efectividad de la política pública de desarrollo económico y productividad para la ciudad de Bogotá 2011-2031. Esta acción gubernamental, se basa en el eje fundamental para preponderar mecanismos que permitan el crecimiento, la productividad, la innovación y competitividad de la ciudad, además de establecer condiciones en las que el desarrollo impacte a los ciudadanos en aspectos como la pobreza, el desempleo y la desigualdad

No obstante, la medición de los impactos de la implementación de la política pública enfrenta importantes retos referente con la disponibilidad y calidad de datos y de información, además la diversidad de actores involucrados y la complejidad de los fenómenos económicos que la política busca impactar.

Estos elementos se fundamentan en el entendido de que la evaluación de la política pública se ha desarrollado predominantemente desde enfoques tradicionales, lo que evidencia la necesidad de incorporar herramientas complementarias que permitan analizar con mayor precisión sus efectos. En este sentido, se busca establecer cómo la política incide sobre variables asociadas a la pobreza, la productividad, el empleo, la competitividad y la desigualdad territorial, con el fin de ampliar la comprensión de sus resultados e impactos en el territorio.

Por lo anterior, la aplicación de modelos de machine learning para evaluar la Política Pública de productividad, competitividad y Desarrollo Socioeconómico resulta pertinente y oportuno. Sin embargo, el resultado del ejercicio depende en gran medida de la calidad y disponibilidad de los datos provenientes de fuentes como el DANE, el Observatorio de

Desarrollo Económico de Bogotá, secretaria distrital de la planeación y la secretaria distrital de desarrollo económico.

Ahora bien, la evaluación de la Política Pública de Productividad, Competitividad y Desarrollo de Bogotá enfrenta el desafío de incorporar herramientas analíticas que permitan comprender con mayor precisión el comportamiento de indicadores clave como el empleo, la productividad, el dinamismo empresarial y la pobreza.

Si bien, los enfoques tradicionales han aportado elementos relevantes para el seguimiento de la acción pública, sus limitaciones frente a la complejidad territorial, la heterogeneidad de los datos y la necesidad de generar análisis en prospectiva, disponen el camino para la implementación de modelos de machine learning como apoyo metodológico.

El vacío de conocimiento que motiva este trabajo radica en la limitada evidencia aplicada sobre el uso de modelos de machine learning para apoyar el seguimiento y análisis de la Política Pública de Productividad, Competitividad y Desarrollo Socioeconómico de Bogotá.

En particular, persisten vacíos fundamentales a partir de la escasa aplicación de estas herramientas a esta política pública, junto con la comparación entre modelos supervisados y no supervisados para valorar su utilidad analítica en indicadores estratégicos de políticas. Estas metodologías pueden ofrecer mejores posibilidades de apoyo técnico para el seguimiento institucional de variables como empleo, productividad, dinamismo empresarial y pobreza multidimensional en el contexto bogotano mediante la complementariedad de modelos de machine learning

Teniendo en cuenta lo anterior, surge la necesidad de identificar qué modelos de aprendizaje automático pueden ofrecer un apoyo más útil al seguimiento y análisis de los indicadores estratégicos de esta política pública en Bogotá ¿Qué modelos de aprendizaje

automático supervisado y no supervisado ofrecen mayor capacidad predictiva y explicativa para apoyar el seguimiento y análisis de indicadores de la Política Pública de Productividad, Competitividad y Desarrollo Socioeconómico de Bogotá 2011 a 2031, en sus indicadores estratégicos de empleo, productividad, dinamismo empresarial y pobreza multidimensional desagregada por localidad?

Evaluación de Políticas Públicas de Desarrollo Económico

La economía constituye la base de los Estados modernos en una era en la que el mundo, gracias a la apertura de fronteras y a la globalización de los mercados, se encuentra particularmente condicionado por la medición del Producto Interno Bruto (PIB) de cada nación. En este contexto, las potencias actuales se caracterizan principalmente por figurar como las economías más prósperas a escala global, estableciendo comparaciones frente a los demás países.

Para autores como Paul M. Romer, el crecimiento económico se define como un aumento en el ingreso o en la producción nacional real, que ocurre cuando los recursos se utilizan de manera más efectiva (Pelsa y Balina, 2022). Lo anterior se asemeja a la definición más tradicional que se ha construido, en la cual la variable principal se concentra en la evolución nominal del Producto Interno Bruto.

En la era moderna, el crecimiento y el desarrollo económicos se encuentran estrechamente relacionados; no obstante, cada uno responde a un concepto distinto. Mientras el crecimiento económico se concentra en medidas cuantitativas, como el Producto Interno Bruto (PIB), el desarrollo económico incorpora elementos que miden las condiciones de vida de la población, así como en dimensiones sociales y ambientales del territorio. Esta distinción resulta fundamental para comprender de manera más amplia los alcances e implicaciones de las políticas económicas en los territorios.

Desde la aparición de Amartya Sen (1999) con su libro *Desarrollo y libertad*, y gracias a la evolución del mundo junto con el establecimiento de objetivos globales de la ONU, los Estados consagraron la relación entre dos términos que a mediados de los años cincuenta parecían inconcebibles: la economía y el desarrollo. Así, al tradicional enfoque productivo y sistemático del crecimiento de los mercados se articula una visión integral del desarrollo,

orientada desde una perspectiva antropocentrista, en la cual el ser humano no solo es objeto de los procesos productivos de los sistemas, sino también el elemento focal. De ahí surge la necesidad de fijar condiciones mínimas que permitan garantizar el desarrollo humano y la calidad de vida de los individuos.

En este marco, el desarrollo debe orientarse a atender los problemas estructurales, lo cual solo es posible mediante el fortalecimiento de la libertad de los individuos, la mejora de su capacidad para ayudarse a sí mismos, el incremento de su agencia y su posibilidad de influir en el mundo (Rivera, 2015, p. 110). Por lo tanto, en un sistema económico, el desarrollo sitúa a los ciudadanos como eje central de la acción.

El desarrollo económico resulta esencial para mejorar los niveles de vida en una región. Su objetivo es aumentar el ingreso real per cápita e impulsar cambios significativos en las estructuras sociales y las actitudes comunitarias. El fin último es erradicar la pobreza y reducir la desigualdad dentro del territorio (Rapanna y Sukarno, 2017, como se citó en Hidayatillah et al., 2024).

Precisamente, la relación entre crecimiento y desarrollo se configura en el propósito de utilizar la tasa de crecimiento económico como base para alcanzar un desarrollo económico equitativo, en el cual se integren dimensiones como desempleo y pobreza, con el fin de garantizar contextos de bienestar para todos los ciudadanos (Hidayatillah et al., 2024).

En concordancia con lo anterior, el desarrollo económico abarca una gama más amplia de mejoras en la calidad de vida, incluyendo factores como la alfabetización, la esperanza de vida y las consideraciones ambientales. No se limita al PIB, sino que integra métricas diversas que reflejan el nivel de vida real y la sostenibilidad ambiental (Pelsa y Balina, 2022).

Es así como el mundo mediante los objetivos de desarrollo sostenible establece en el Objetivo 8:

“Promover el crecimiento económico inclusivo y sostenible, el empleo y el trabajo decente para todos en la que se enlaza la relación que deben integrar los estados entre el crecimiento y el desarrollo en las que los y las ciudadanas tengan condiciones óptimas en sus territorios con adecuadas condiciones de vida bajo la primicia del desarrollo humano y dignidad”.

Las políticas públicas, especialmente en la etapa de evaluación, son relevantes para establecer condiciones que fomenten y garanticen mecanismos para el desarrollo económico de los ciudadanos y naturalmente mejores condiciones de vida de los ciudadanos.

Gracias a ello, en la actualidad diversos responsables de la toma de decisiones reconocen la evaluación de políticas públicas como un elemento fundamental para garantizar que estas se sustenten en evidencia sólida y de calidad. Este reconocimiento resalta la importancia de incorporar procesos de evaluación sistemática, tanto en la formulación como en el seguimiento y ajuste de las políticas públicas, con el fin de fortalecer su pertinencia, efectividad e impacto en la sociedad (Salama & Picalarga, 2024).

El éxito de las políticas públicas se refleja en diversos casos de éxito en las últimas décadas, donde la formulación e implementación de acciones públicas han impulsado el crecimiento económico y el desarrollo.

Un gran referente de estos casos de éxito es China, ya que ejemplifica como la aplicación de políticas efectivas inherentemente conllevan a un crecimiento económico y transformación social, lo que ha llevado a este país a consolidarse como un gigante asiático como la potencia económica más influyente del mundo en la actualidad (Pasco & Bendezu, 2024).

Se evidencia como al evaluar la pertinencia y efectividad de las políticas dentro de los Estados, determinando así su impacto en el desarrollo económico y en la vida de los ciudadanos.

La evaluación de las políticas públicas refleja la pertinencia de la acción estatal para atender situaciones, problemas y circunstancias específicas en un área determinada. Este proceso se evidencia en los indicadores, en la materialización de las acciones públicas en territorios definidos y naturalmente en la pertinente atención y resolución de las problemáticas o situaciones que se pretendía abordar mediante la formulación de la política pública.

Muñoz (2022) señala:

“Se debe entender que la evaluación de políticas públicas debe formar parte del propio proceso de la intervención pública; más allá de su ya obligada regulación legal, los gestores de programas públicos deben evitar que se produzcan desviaciones entre lo previsto y lo realmente ejecutado y valorar el adecuado cumplimiento de los fines y objetivos inicialmente previstos, pronunciándose sobre esta adecuación tanto en términos de eficiencia como de eficacia” (p. 3).

En efecto, resulta fundamental para responder a la pregunta clave de si las políticas implementadas han tenido el impacto esperado sobre los problemas sociales que se pretendía atender en un territorio específico (How to Evaluate Policies, 2022). De esta forma, se configura la necesidad de evaluar las políticas en las que se comprometen las finanzas de los Estados, lo cual se encuentra intrínsecamente relacionado con el crecimiento económico mediante las acciones estatales orientadas a garantizar un desarrollo continuo.

El crecimiento económico se entiende como el aumento cuantitativo del PIB de un país, que refleja el nivel de bienes y servicios producidos en un periodo determinado. En este sentido, Joseph E. Stiglitz, premio Nobel de Economía, señala que suele existir confusión entre los conceptos de desarrollo y crecimiento, afirmando que el desarrollo económico incluye metas distintas a las del crecimiento, tales como el desarrollo sostenible, equitativo y democrático (Stiglitz, 1998). En esa misma línea, Barre (1962) advertía: “El estudio del desarrollo económico

no podría confundirse con el estudio de los crecimientos equilibrados dentro de una economía capitalista evolucionada” (p. 13).

Desde una conceptualización moderna, el crecimiento económico se considera necesario, ya que permite a la comunidad consumir más bienes y servicios y garantiza un nivel cuantitativo más alto de producción, lo que a su vez contribuye a una mejora real en los estándares de vida y al crecimiento de los países (Nastu, Dumitrache & Melian, 2022, p. 1604).

En esa misma dirección, Pelsa y Balina (2022) retoman la definición de Paul M. Romer, para quien el crecimiento económico se refiere al aumento del ingreso o de la producción nacional real, basado en un uso más eficiente de los recursos. En general, este concepto está estrechamente vinculado a la variación cuantitativa del Producto Interno Bruto, que expresa el nivel de bienes, servicios y consumo de los actores económicos de un país en un periodo de tiempo específico.

En contraste, el desarrollo económico integra tanto el aspecto cuantitativo, representado por el crecimiento económico, como el aspecto cualitativo, orientado a elevar la calidad de vida de la población. Medir únicamente la dimensión cuantitativa puede llevar a interpretaciones erróneas sobre el nivel de desarrollo alcanzado (Nastu, Dumitrache y Melian, 2022, p. 1605). Como señala Blakely (1989), citado en Sinclair (1999), “la característica central del desarrollo económico radica en el énfasis en el desarrollo endógeno, utilizando el potencial de los recursos humanos y físicos locales para crear nuevas oportunidades de empleo y estimular nuevas actividades económicas de base local” (p. 2).

De acuerdo con Lauchlin Currie, el crecimiento es una condición necesaria para el desarrollo, pero no suficiente por sí solo (citado por Montenegro, 2012). Por esta razón, se trata de un proceso intrínseco: el desarrollo depende de la base que provee el crecimiento económico,

aunque este, por sí mismo, no garantiza mejoras en la calidad de vida. En este sentido, el desarrollo está condicionado por la manera en que se gestionan los recursos disponibles y cómo se distribuyen en planes, proyectos y, por supuesto, políticas.

Por lo tanto, las políticas públicas juegan un papel determinante en los procesos de construcción de las economías estatales, siendo la evaluación un instrumento vital que utiliza el gobierno para regular diversos aspectos de la vida de las personas. Las acciones estatales se evidencian en el impacto que generan, puesto que, al ser implementadas, tienen como propósito crear mejores condiciones para la sociedad (Amirullah y Natsir, 2024).

En esa misma dirección, el desarrollo y las políticas públicas se encuentran profundamente vinculados. Estas últimas constituyen los marcos que configuran los procesos de desarrollo, haciendo que su relación sea crucial para una gobernanza efectiva y para la evolución de los estados modernos, lo cual se refleja claramente en la evaluación de dichas políticas (Ghimire, 2024).

Métodos tradicionales de evaluación de políticas Públicas

En el proceso de evaluación de las políticas públicas, se han constituido herramientas para valorar el impacto de esta en los problemas o situaciones por las cuál inicialmente se formularon. Las políticas públicas se elaboran para lograr metas específicas y dirigirse a un grupo definido de posibles beneficiarios, así mismo están diseñadas para lograr ciertas metas y cubrir a un número de personas que resultan beneficiarias (Roman, 2024)

Adicionalmente las técnicas de evaluación de impacto causal permiten medir los efectos de las políticas públicas en las variables de interés de forma que pueda evaluarse hasta qué punto la política ha logrado sus objetivos (Artes y Rodríguez, 2022), Román (2024) señala que:

Las políticas públicas están diseñadas para lograr ciertas metas y cubrir a un número concreto de potenciales personas beneficiarias. En este artículo se describen los

principales métodos empleados para determinar si las intervenciones públicas realmente funcionan, así como el nivel y la naturaleza de los impactos generados sobre los destinatarios. La necesidad de evaluar el impacto generado por aquellas se justifica, fundamentalmente, por tres razones. En primer lugar, emplaza a los responsables de la gestión y formulación de políticas (“policy makers”) a contrastar si los programas están alcanzando los resultados inicialmente previstos. (párr. 1)

Así, se han configurado mecanismos que permiten establecer el impacto de las políticas públicas. Para Galtseva (2023), establece que al evaluar el desempeño de las autoridades implica, en primera instancia, analizar los costos asociados a la fase de implementación de la política. Galtseva indica que, en un segundo momento, resulta necesario establecer indicadores integrales que reflejen la relación entre costos y resultados, ajustados a cada contexto particular. Además, el autor resalta la importancia de valorar en qué medida las autoridades cumplen con sus metas y objetivos.

También señala que las evaluaciones pueden ser categorizadas en dos tipos: evaluaciones internas realizadas por las autoridades ejecutoras de la política, y evaluaciones externas realizadas por expertos o consultores independientes. Esta distinción es importante para comprender las diferentes perspectivas sobre la eficacia de las políticas (Galtseva, 2023)

Por otro lado, Muñoz (2013) plantea que la evaluación de políticas públicas puede apoyarse en enfoques de carácter económico y metodológico, orientados a valorar la eficiencia, la efectividad y el impacto de las intervenciones públicas.

Desde esta perspectiva, herramientas como el análisis costo-beneficio, el análisis costo-efectividad, los modelos econométricos, los indicadores de desempeño, las evaluaciones externas y los marcos lógicos permiten estructurar procesos de evaluación con criterios comparables, objetivos y medibles. De esta manera, la evaluación no solo busca establecer si una política

alcanza los resultados previstos, sino también determinar si los recursos utilizados guardan una relación adecuada con los efectos generados.

Según Artés y Rodríguez-Sánchez (2022), los métodos tradicionales de evaluación incluyen ensayos aleatorizados, experimentos naturales, regresiones, técnicas de emparejamiento, regresión discontinua, diferencias en diferencias y variables instrumentales. Todos ellos buscan identificar relaciones causa-efecto confiables en contextos reales.

Por otro lado, Fernández (2022) plantea la necesidad de configurar nuevos modelos de evaluación, como el idealista y el pragmático. Según Fernández, en primer lugar, se centra en alinear los cambios sociales con un estado ideal de organización social y gobernanza, mientras que el segundo evalúa el progreso social según la percepción de si la situación actual es mejor que la anterior, utilizando criterios específicos de juicio.

Bel (2017) propone la aplicación de metodologías, como estudios de caso, análisis comparativos de políticas en el marco valorar los efectos de la implementación, lo que implica naturalmente en integrar y adaptar indicadores, basarse en datos y evidencia para evaluar los resultados de las políticas. Estos métodos proporcionan una base más imparcial para el análisis, en relación a los modelos tradicionales que históricamente han privilegiado el enfoque cuantitativo.

En términos pragmáticos, la evaluación de políticas públicas en países con economías en desarrollo, como en el caso de América Latina, continúa sustentándose en elementos de análisis tradicionales, al tiempo que incorpora enfoques integrales que permiten atender la complejidad y la naturaleza particular de las políticas que se formulan en dichos contextos.

Evaluación de Política Pública en Bogotá

Bogotá es la entidad territorial del país que ha desarrollado uno de los marcos más amplios de instrumentos para la formulación, implementación, seguimiento y evaluación de políticas públicas. Por esta razón, la ciudad se ha consolidado desde hace varios años como un referente en el diseño metodológico de guías y lineamientos orientados a fortalecer los procesos de evaluación de la acción pública.

Este avance no solo responde a las necesidades propias de la gestión distrital, sino que también se articula con los lineamientos, enfoques y herramientas promovidos a nivel nacional por el Departamento Nacional de Planeación (DNP), especialmente en materia de gestión por resultados, seguimiento y evaluación.

En ese marco, Bogotá ha buscado adaptar estos referentes al contexto distrital e impulsarlos, con el propósito de contar con mecanismos más estructurados para valorar el desempeño de sus políticas y establecer una ruta vanguardista en políticas públicas.

Como resultado de este proceso, Bogotá expidió la RESOLUCIÓN No. 1634 DE 2024 “Por medio de la cual se adopta la Guía de Evaluación de Políticas Públicas en el Distrito Capital”. Según la *Guía de evaluación de políticas públicas distritales*, la evaluación de una política pública en Bogotá se desarrolla como un proceso sistemático que inicia con la selección de la intervención a evaluar, continúa con el diseño técnico de la evaluación y su ejecución, y se complementa con la socialización de resultados, el uso de recomendaciones y la gestión del conocimiento institucional.

Tabla 1

Elementos Centrales del Proceso de Evaluación de Políticas Públicas en Bogotá

Aspectos clave	Descripción
Concepto de evaluación	La evaluación es una valoración sistemática del diseño, la

Propósito de la evaluación	implementación y del grado de aporte al cumplimiento de los objetivos de una intervención pública (Mideplan, 2017). Busca generar evidencia para la toma de decisiones, la rendición de cuentas, la mejora de las intervenciones y la asignación eficiente de recursos.
Fases del proceso	Selección de la intervención, Diseño de la evaluación Ejecución.
Criterios de selección	La intervención se selecciona según su evaluabilidad e importancia. Esta depende de la disponibilidad, actualización y calidad de la información; la importancia, de su aporte al Plan de Desarrollo, magnitud y utilidad para decidir.
Diseño de la evaluación	El diseño exige identificar la cadena de valor, la teoría del cambio, los cuellos de botella, el objetivo, el alcance, las preguntas orientadoras, el tipo de evaluación y las fuentes de información.
Tipos de evaluación	Institucional, De operaciones, de resultados, de impacto costo-efectividad.
Información para Evaluar	La evaluación puede apoyarse en información primaria y secundaria, así como en enfoques cuantitativos y cualitativos; cuando es posible, se recomienda la triangulación.
Uso de los resultados	La guía enfatiza en la socialización de resultados, el uso de recomendaciones y la gestión del conocimiento para mejorar la intervención pública.

Nota. Fuente: Elaboración propia a partir de la Guía de evaluación de políticas públicas distritales.

De esta manera, el distrito define diversos tipos y alcances como parte del ejercicio de evaluación de política pública, con el propósito de orientar de forma metodológica el análisis de las intervenciones públicas. Como se observa en la siguiente tabla, la guía distrital plantea cinco tipos esenciales de evaluación: institucional, de operaciones, de resultados, de impacto y de costo-efectividad. Esta clasificación permite abordar diferentes dimensiones de la intervención, de acuerdo con el objetivo de la evaluación, el nivel de madurez de la política y la información disponible para su desarrollo.

Tabla 2*Tipos de Evaluación en Política Pública*

Tipo de Política pública	Institucional	Operaciones	Resultados	Impacto	Costo-efectividad
Descripción	Analiza y valora la capacidad institucional de las entidades públicas para operar la intervención.	Analiza los procesos de implementación de la intervención en contraste con la forma en que se planeó y verifica el cumplimiento de las metas operativas.	Evidencia los cambios en las condiciones y en el bienestar de los beneficiarios después de la entrega o consumo de los productos.	Evidencia los cambios en las condiciones y en el bienestar de los beneficiarios como consecuencia directa de los productos de la intervención.	Realiza un análisis económico que compara los costos relativos con los resultados o impactos de dos o más intervenciones.

Nota. Fuente *Guía de evaluación de políticas públicas distritales 2024*

En coherencia con ello, las metodologías establecidas por el Distrito se articulan con las herramientas y lineamientos definidos por el Departamento Nacional de Planeación (DNP), entidad orientadora del seguimiento y la evaluación de proyectos públicos en el país. En este marco, el DNP también establece distintos tipos y objetivos de evaluación de política pública, los cuales sirven como referente para estructurar ejercicios de valoración más precisos y acordes con las necesidades de las entidades territoriales.

Tabla 3*Objetivo de los Tipos de Evaluación*

Tipo	Objetivo
Ejecutiva	Establece análisis y propuestas específicas de ajuste sobre los principales aspectos del programa en revisión: diseño, resultados, insumos, manejo operativo, estructura organizacional, actividades de direccionamiento, seguimiento y control.
Operaciones	Analiza de manera sistemática la forma en que opera una intervención pública y cómo sus procesos conducen al logro de sus objetivos. Identifica las relaciones que cada una de las actividades requiere para producir un bien o servicio.
Institucional	Permite analizar y valorar un programa tomando como marco de

referencia el arreglo institucional en el que opera. Estudia las razones por las cuales falla la capacidad institucional en los programas de una intervención, para brindar los insumos para mejorar la gestión y proveer de manera más efectiva los bienes o servicios.

Resultados

Determina los efectos intencionales o no de la intervención pública, una vez se han consumido los productos. Estos efectos deben estar relacionados con la intervención pública.

Impacto

Permite identificar los efectos exclusivamente atribuibles a la intervención pública. La medición de impactos permite cuantificar y verificar la relación de causalidad entre la intervención pública y el resultado.

Nota. Fuente: Elaborado a partir de la guías del DSEPP-DNP

En conjunto, estos tipos de evaluación permiten abordar la intervención pública desde distintas dimensiones analíticas, dependiendo del propósito del ejercicio valorativo, teniendo en cuenta la naturaleza de la acción pública basada en las necesidades a atender o los problemas que se requieren dirimir.

Mientras algunas evaluaciones se concentran en aspectos de diseño, operación o capacidad institucional, otras se orientan a identificar los resultados e impactos derivados de la intervención. Bajo este enfoque, Bogotá ha fortalecido sus capacidades institucionales y metodológicas, consolidándose como un referente latinoamericano en el desarrollo de herramientas de seguimiento y evaluación de políticas públicas.

Machine Learning Aplicado a la Evaluación de Políticas Públicas

El aprendizaje automático resulta ser la herramienta más complementaria e integral en todas las labores y actividades de la sociedad actual. La generación de datos y de información, resulta elemental para la generación de toma de decisiones y acciones más informadas, usando métodos de machine learning para la predicción de eventos y reducción de la incertidumbre.

No obstante, Veale y Brass (2019) señalan que la incorporación del machine learning en las acciones gubernamentales, pese a sus avances e implementaciones, también plantea tensiones institucionales. Aunque el aprendizaje automático posee un potencial transformador para mejorar la gestión de las políticas, su adopción puede verse limitada por la resistencia al cambio dentro de las estructuras burocráticas tradicionales.

Las organizaciones del sector público suelen enfrentar dificultades para modificar prácticas, rutinas y formas de operación ya establecidas, fenómeno asociado con el conservadurismo dinámico de las burocracias públicas.

Es así como el aprendizaje automático puede mejorar significativamente los procesos de formulación de políticas y evaluación en los territorios, reduciendo la operatividad de los procesos internos.

De igual forma, el machine learning proporciona herramientas avanzadas y completas para el análisis de datos, modelado predictivo y evaluación continua, que son cruciales para la toma de decisiones informadas en la gobernanza para el direccionamiento de los Estados modernos (Hasan y Buadi, 2024)

Por supuesto el machine learning permite el procesamiento de grandes conjuntos de datos para descubrir patrones y predecir tendencias, lo cual es crucial para optimizar la asignación de recursos y ajustar las intervenciones en sectores como la salud y la educación y particularmente el desarrollo económico de los territorios (Chy y Buadi, 2024).

Además, el machine learning permite procesar grandes volúmenes de datos e identificar patrones que pueden aportar información relevante para la formulación y optimización de políticas públicas en menor tiempo. Así contribuye a la toma de decisiones más informadas,

eficientes y adaptables, aspectos que resultan esenciales para fortalecer una gobernanza efectiva (Hasan y Buadi, 2024).

El aprendizaje automático ha proporcionado diferentes técnicas y algoritmos para predecir situaciones basadas en grandes cantidades de información, ya que con un adecuado procesamiento y filtrado de datos, pueden generar predicciones muy efectivas (Corba y Negre, 2024, como se citó en Hernández, 2024).

Por lo cual las herramientas de aprendizaje automático se utilizan cada vez más en la evaluación de políticas públicas para estimar los efectos del tratamiento de información de manera flexible. Sin embargo, estos modelos a menudo funcionan como cajas negras, lo que hace que sea difícil para los gobiernos garantizar la equidad, la correcta interpretación de la evidencia y la rendición de cuentas (Rehill y Biddle, 2024).

En relación con lo anterior existen diversos desafíos para la aplicación efectiva de esta técnica en el análisis de políticas públicas. En lo anterior, los análisis representan retos relacionados con la calidad de los datos, la complejidad y heterogeneidad de la información recopilada en los métodos aplicados para la evaluación de políticas (B. Chen et al., 2019; Kumar et al., 2019).

Por supuesto que la implementación de modelos de los modelos de machine learning traen consigo vastos retos para las instituciones, ya que es imperativo se adopten lineamientos enmarcados en la responsabilidad frente a la información que se suministra a los modelos, puesto que se deben integrar esquemas éticos de recolección y uso de los datos, para que así mismo, les permitan un procesamiento más eficaz y una correcta aplicación de los modelos a la hora de realizar en análisis y evaluación de las políticas

Al utilizar herramientas de aprendizaje automático, los formuladores de políticas pueden mejorar sus procesos de toma de decisiones a través del análisis avanzado de datos, el modelado predictivo y la evaluación continua (Hasan y Buadi, 2024)

Es importante realizar avances en las metodologías causales de aprendizaje automático para garantizar que puedan ser utilizadas de manera confiable en la evaluación de políticas. Esto incluye el desarrollo de modelos que sean precisos e interpretables, que permitan evaluaciones de políticas justas y responsables (Rehill y Biddle, 2023).

Ventajas y Desafíos de Machine Learning en Evaluación de Políticas Públicas

El machine learning en la evaluación de políticas públicas constituye una línea de estudio e investigación emergente y relevante en los contextos globales que atraviesan los gobiernos de diferentes niveles. Las particularidades de las herramientas de aprendizaje automático e inteligencia artificial aplicadas al campo de la acción pública han comenzado a explorarse en años recientes, por lo que aún no existen antecedentes considerables sobre su aplicación e implementación en este ámbito.

No obstante, en la evaluación de políticas públicas, los modelos de machine learning pueden alcanzar altos niveles de precisión al incorporar índices de satisfacción ciudadana y aprovechar el big data, lo que proporciona una base más científica para la toma de decisiones (Xu et al., n.d.).

En principio, las ventajas del machine learning permiten identificar los elementos centrales de las políticas públicas, lo que facilita determinar la efectividad de la acción gubernamental y analizar los efectos de entender las problemáticas como lo pueden percibir los ciudadanos en un territorio determinado.

En los estudios revisados, se ha identificado técnicas como los bosques causales, eficientes en contextos para mejorar la inferencia causal en las evaluaciones de políticas. Estos métodos superan a los enfoques tradicionales en la recuperación de efectos causales y en la caracterización de su heterogeneidad, lo cual resulta esencial para una evaluación precisa del impacto de las políticas (Brignoli et al., 2024).

Sin embargo, a pesar de los paulatinos avances teóricos, los estudios sobre la efectividad real de la aplicación de modelos de machine learning en políticas públicas siguen siendo limitados. Esta brecha dificulta comprender cómo los marcos teóricos se traducen en la práctica y cuál es el verdadero impacto social en la evaluación de dichas políticas (Rodolfa et al., 2020).

Lo anterior refleja que las herramientas de machine learning han sido exploradas de manera incipiente por las instituciones públicas, en particular en el ciclo de las políticas públicas. No obstante, representan un elemento esencial frente a la gran cantidad de datos que producen las organizaciones estatales, los cuales en países en desarrollo suelen recibir poca atención. Del mismo modo, el escaso procesamiento de datos y el limitado uso de la información hacen que su potencial resulte subutilizado.

Con la implementación de modelos de machine learning, los sistemas de información pública pueden aplicar modelos predictivos que mejoren la eficiencia en la distribución de recursos. Como señalan Janssen et al. (2017, citado por Rodolfa et al., 2020), enmarca que las necesidades son considerables y los recursos, naturalmente limitados; por tanto, se hace indispensable mejorar la focalización para aumentar el impacto de las acciones gubernamentales con presupuestos restringidos para ser más eficiente.

En general, el desarrollo de políticas públicas considera un proceso de evaluación que hoy en día no se limita a la etapa final de implementación, sino que ha ido incorporando la

evaluación en las distintas fases de los proyectos y políticas, a través de diversas modalidades actualmente disponibles (Yanes, 2022, p. 2).

El machine learning facilita la evaluación continua de políticas, ya que posibilita el monitoreo permanente y establece relaciones de variables para evidenciar la evolución de la implementación de políticas. Esta cualidad garantiza la adaptabilidad en la implementación, asegurando que las políticas se mantengan efectivas y pertinentes frente a contextos cambiantes (Chy y Buadi, 2024).

Pese a las considerables ventajas de la implementación de los modelos de machine learning en las políticas públicas y en general en el ciclo de las acciones gubernamentales, existen naturalmente varios desafíos en la aplicación de estas herramientas.

El uso de machine learning en la evaluación de políticas públicas plantea importantes preocupaciones en torno a la privacidad de los datos, especialmente cuando se manejan conjuntos de información sensible provenientes de individuos o instituciones (Li et al., n.d.). En este contexto, la ética en el gobierno de los datos adquiere una relevancia central, ya que la recolección, procesamiento y análisis de la información deben estar debidamente regulados por los órganos competentes.

Las preocupaciones con respecto a la privacidad de los datos y transparencia son contundentes y también se destacan como cuestiones críticas que deben administrarse de manera efectiva cuando se utiliza machine learning en este dominio debido a los riesgos y situaciones que pueden emanar en una mala gestión y administración de los datos en las instituciones públicas (Hasan, Chy y Buadi, 2024)

Esta ausencia de regulaciones claras, el uso de ML podría vulnerar la confidencialidad o los derechos fundamentales de los ciudadanos (Li et al., s.f.; Santoni de Sio & Mecacci, 2021).

No obstante, uno de los principales desafíos del uso de machine learning en la toma de decisiones algorítmicas es la falta de transparencia, lo cual dificulta la rendición de cuentas. Para el autor Paul de Laat (2018) explica que:

“Casi todos los aspectos de esta toma de decisiones algorítmica siguen siendo opacos. Como regla general, apenas se proporciona explicación o aclaración. En la medida en que salen a la luz los detalles, suelen ser poco iluminadores e informativos. Incluso las cláusulas legales que abordan el problema tienen poco efecto” (p. 526).

Es esencial que estos procesos no interfieran con los principios de transparencia, legalidad y debido manejo de la información, ni vulneren los derechos de privacidad de los ciudadanos ni la confidencialidad institucional. Por lo tanto, se establece la necesidad de establecer marcos normativos claros que garanticen un equilibrio entre innovación tecnológica, protección de datos y respeto a los principios democráticos principalmente para los países que han adoptado la implementación de los modelos en sus actuaciones públicas.

Por otro lado, a pesar del evidente potencial del machine learning para mejorar la evaluación de políticas públicas, su implementación enfrenta desafíos significativos como lo es la falta de interpretabilidad de los modelos utilizados. Esto se refleja en la aplicación de los modelos debido muchas técnicas de modelos de machine learning, especialmente aquellas orientadas a estimar efectos causales, como los causal forests o el double machine learning, funcionan como cajas negras, lo que dificulta entender cómo se generan las estimaciones y limita la trazabilidad de decisiones políticas basadas en estos modelos (Rehill y Biddle, 2023).

En este sentido, la discordancia entre la información suministrada y la interpretabilidad de los resultados en la evaluación de la política pública puede generar desconfianza frente a los hallazgos obtenidos, especialmente en los procesos de rendición de cuentas y en la valoración de los efectos públicos de la acción gubernamental. Por ello, este aspecto se configura como un

desafío relevante, en la medida en que los modelos utilizados deben ser coherentes con la exposición clara de resultados a la ciudadanía y fundamentarse en principios de transparencia, publicidad y ética en el manejo de la información.

Estas descripciones ocasionan que se limite la posibilidad de auditar las decisiones algorítmicas y así mismo obstaculiza su entendimiento, lo cual es especialmente problemático cuando se aplican en contextos públicos que exigen altos estándares de transparencia y claridad en la exposición de la información a los ciudadanos.

Es así como la transparencia en la aplicación de los modelos de machine learning en la evaluación de políticas públicas resulta crucial. Al respecto, de Fine Licht y de Fine Licht (2020) en su texto establecen qué:

“La transparencia como promotora de la rendición de cuentas puede contribuir al mito de la política oculta [...]. Así, cuando se hacen transparentes el código y los datos, el público podría pensar que esta divulgación está ocultando el ‘código real’ que está en uso, o que hay algún tipo de ‘puerta trasera’ en el código para asegurar que funcione de una manera diferente a la esperada. [...] En consecuencia, la transparencia total podría, de hecho, hacer que el público se sienta menos propenso a creer que tiene más control sobre la IA que antes de tener acceso a los datos” (de Fine Licht & de Fine Licht, 2020, p. 923).

No obstante, pese a los retos y desafíos descritos, es evidente que, en los últimos años, el aprendizaje automático ha proporcionado diferentes técnicas y algoritmos para predecir situaciones basadas en grandes cantidades de información que, a través de un adecuado procesamiento y filtrado de datos, pueden generar predicciones muy efectivas en la implementación y evaluación de políticas públicas (Corba y Negre, 2024, como se citó en Hernández, 2024).

La incorporación del machine learning en el análisis de políticas públicas implica reconocer diferencias entre interpretabilidad y desempeño predictivo. Por lo que para la toma de decisiones públicas, es necesario que sus resultados sean comprensibles y coherentes con principios de transparencia y rendición de cuentas.

Modelos como la regresión lineal, la regresión logística y los árboles de decisión ofrecen mayores posibilidades de interpretación, mientras que enfoques más complejos, como Random Forest, Gradient Boosting, SVR y redes neuronales, pueden ofrecer mejores niveles de predicción.

Casos de Estudio de Machine Learning en Desarrollo Socioeconómico

Los modelos de aprendizaje automático son elementos relevantes en los procesos de valoración de políticas orientadas al desarrollo económico en los territorios y Gobiernos. Los modelos de aprendizaje de machine learning empíricamente puede mejorar significativamente la precisión de los pronósticos económicos, como las predicciones del PIB, mediante el análisis de conjuntos de datos complejos y la identificación de factores clave que contribuyen a las disparidades económicas (Gao, 2024).

En este contexto hay evidencia de ciudades que han adoptado medidas y modelos para establecer predicciones principalmente de crecimiento económico. Lo anterior debido a que el machine learning permite modelar la relación entre factores estructurales como la densidad del empleo, el capital humano, la infraestructura o el tipo de industria y variables de resultado como el ingreso, la productividad o el nivel de desigualdad (Gao et al., 2023)

Precisamente, Gao, Wang y Zhou (2023), en su texto *Is China's Urbanization Inclusive, Comparative Research Based on Machine Learning Algorithms* validan la aplicación de algoritmos de clustering de aprendizaje no supervisado con el fin de clasificar a los trabajadores

urbanos en grupos socioeconómicos bajos, medios y altos con base en 21 características individuales. Este enfoque permitió evaluar de manera más precisa el efecto de la aglomeración económica sobre los salarios, evidenciando que esta tiene un impacto positivo en los ingresos de los trabajadores de los grupos bajos y medios (Gao et al., 2023).

Adicionalmente, los autores señalan que “al combinar técnicas de aprendizaje automático con datos de alta dimensión, se obtiene una visión más objetiva de la estructura del mercado laboral urbano, superando las limitaciones de las clasificaciones tradicionales basadas únicamente en el ingreso” (Gao et al., 2023, p. 2).

Lo anterior intuye que el machine learning en la formulación de las políticas de desarrollo económico debido a que permite identificar los segmentos poblacionales con factores como la pobreza, la desigualdad, los sectores económicos e industriales, la prestación de servicios públicos, todos esto por supuesto en un territorio definido.

En el texto *Predicting Urban Growth with Machine Learning*, el autor Gao, 2024, configura elementos para establecer un protocolo de crecimiento poblacional comparativo mediante herramientas de estado unidos y china en un lustro como periodo de tiempo. El caso analizado, particularmente involucra factores directamente relacionados enmarcadas en el desarrollo económico de estas dos potencias, donde involucra elementos como crecimiento urbano futuro que son comunes tanto a Estados Unidos como a China: capital humano, inversión inmobiliaria, servicios públicos y empleo.

En esta descripción de características del caso de estudio modelos como la regresión, redes neuronales, bosques aleatorios. Estados modelos ofrecen un enfoque de amplia gama de acciones que resultan con pragmáticos con amplia gama de datos, lo que permite anticipar tendencias y orientar decisiones del crecimiento urbano y económico basadas en evidencia.

En particular modelo de redes neuronales de aprendizaje, son adecuados a capturar relaciones no lineales entre características urbanas complejas, que los modelos de regresión tradicionales a menudo pierden.

De este modo Xu, X., Chen, Z., & Chen, S en su texto Enhancing economic competitiveness analysis through machine learning: Exploring complex urban feature identifican que el modelo resulta paradigmático puesto que esta capacidad permite un análisis más matizado de la competitividad económica urbana, como lo demuestra un estudio que utiliza una red neuronal convolucional (CNN) para clasificar la competitividad económica urbana a partir de un conjunto de datos de 1008 características de ciudades chinas (Xu et al., 2023).

En consecuencia, confirma que los modelos de machine learning en la valoración de políticas de desarrollo económico enmarcado en la generalización de su aplicabilidad, configura una herramienta superior en el procesamiento de datos y registro de información para la validación de efectiva del desarrollo económico en el mundo.

Marco Teórico

Desarrollo y Crecimiento Económico: Análisis Teórico desde un Enfoque del Desarrollo

Para contextualizar los cimientos y evolución del concepto de desarrollo económico es imperativo realizar un análisis a el enfoque clásico con la teoría económica clásica, presidida principalmente por autores como Adam Smith, postula que los mercados son sistemas de autorregulación que naturalmente conducen a resultados eficientes. Esto se basa en la idea de la “mano invisible”, donde el interés propio individual beneficia inadvertidamente a la sociedad en su conjunto (Zhironkin et al., 2023) (Migunovand & Syutkina, n.d.)

Esta doctrina, fue comprendida desde finales del siglo XVIII hasta la primera mitad del siglo XIX, en lo cual, el sistema económico de los nacientes estados modernos se configurada mediante la lógica de la autorregulación, así mismo el enfoque clásico opto por limitar el papel de Estado en las decisiones del mercados, con el fin de permitir que aquella mano invisible en el mercado, esto descrito como un “vigilante nocturno” que solo interviene para proteger los derechos de propiedad y mantener el orden(Migunovand & Syutkina, n.d.)].

En concordancia con lo anterior, las teorías clásicas a menudo enfatizan etapas lineales de crecimiento, donde las economías progresan a través de una serie de etapas predecibles impulsadas por la acumulación de capital y el avance tecnológico (Dang & Pheng, 2014).

Esta escuela de pensamiento, gracias a David Ricardo con su teoría de la ventaja comparativa, explicó cómo el comercio internacional puede impulsar el desarrollo económico y Adam Smith con su libro la riqueza de las naciones estableció los principios de la economía clásica y los cimientos de los enfoques de la economía, sin consideración mayor sobre el término de desarrollo.

Por lo anterior, la noción de desarrollo estaba alejada de las consideraciones en la económica. La concentración de riqueza, acumulación de capital y por supuesto la expansión de la producción eran los elementos pioneros para garantizar el crecimiento de la economía y los nacientes Estados.

La economía neoclásica introduce el concepto de marginalismo, enfocándose en cómo los individuos toman decisiones al margen para maximizar la utilidad y las ganancias. Este enfoque asume un comportamiento racional y una información perfecta sobre los sistemas económicos (Zhironkin et al., 2023).

Gallego destaca el enfoque neoclásico destaca qué (1993):

“lo fundadores del sistema neoclásico utilizaron el análisis marginal, los conceptos de utilidad marginal y productividad marginal, para determinar la formación de los precios de los bienes, servicios y factores de producción. en mercados competitivos. Ellos enfatizaron la escasez como el elemento determinante de los precios. En particular, los economistas marginalistas estudiaron la posibilidad de que un conjunto de precios asegura la igualdad de la oferta y la demanda en los mercados” pp61

En la concepción de los clásicos, el mercado aparece como un mecanismo autorregulador que conduce a la máxima eficiencia en la producción y asignación de bienes. La intervención del Estado debe ser mínima, limitada a proteger el ‘orden natural’ y las ‘libertades’, es decir, el Estado como un ‘vigilante nocturno’ (Migunov & Syutkina, 2024, p. 2).

La concentración del enfoque clásico y neoclásico ha determinado quizás que muchos de los elementos necesarios para el progreso de un estado pasaran desapercibidos como la integralidad de elementos propios de la sociedad como aspectos de pobreza, empleo, acceso a condiciones de salubridad básica, concentrando las miradas en elementos de libre mercado, reducir costos de producción y aumentar las utilidades.

Tanto los enfoques clásicos como neoclásicos han sido criticados por su insuficiencia para explicar y abordar las complejidades del desarrollo económico, particularmente en las regiones subdesarrolladas. Los críticos argumentan que estos modelos a menudo pasan por alto factores externos y el papel de las instituciones en el desarrollo (Shaikh, 2013).

A pesar de lo anterior, las teorías neoclásicas determinaron que el mercado competitivo, sin intervención del gobierno es capaz de conducir a los mejores resultados de equilibrio y lograr el máximo bienestar de la sociedad” (Migunov & Syutkina, 2024, p. 5).

Por lo que estos enfoques son los pioneros en los procesos económicos que adoptaron los Estados nacionales, traen consigo limitaciones estructurales y considerables desafíos modernos para el desarrollo económico, entiendo que libertad de mercado y la maximización de beneficios permitió el progreso, el avance de la industrialización y expansión de las económicas locales, sin embargo, la sectorización de los beneficios fue más conveniente en algunas economías y excluyentes en otras, principalmente en los estados con economías emergente.

En consecuencia, a lo anterior El “Otro canon” y los enfoques histórico-neoinstitucionales ofrecen marcos alternativos que enfatizan el papel de la industrialización, la intervención institucional, y la reducción de los costos de transacción en el fomento del desarrollo. Estos paradigmas desafían el enfoque neoclásico de la autorregulación del mercado y abogan por políticas que promuevan activamente el crecimiento económico y reduzcan la pobreza (Bukvić, 2014)

El enfoque estructuralista pretendió aterrizar las ambigüedades heredadas de los enfoques clásico y neoclásico, precisamente integrando las limitaciones estructurales que se presentaron. En este enfoque del desarrollo reorienta los modelos económicos a elementos en los cuáles la relevancia además del crecimiento, se enfatiza en comprender las estructuras económicas y los

factores sistémicos que influyen en el crecimiento y el desarrollo, particularmente en los países en desarrollo.

El origen particular de este enfoque es crucial en la región ya que planteó herramientas para la formulación de políticas públicas económicas orientadas a las condiciones estructurales de los países en desarrollo, principalmente para Latinoamérica, resaltando por supuesto Colombia.

El origen se remonta a 1948, mientras Estados Unidos se empieza a consolidar como la nueva economía hegemónica mundial, en América Latina se configura la Comisión Económica para América Latina y el Caribe (CEPAL) bajo el liderazgo de Raúl Prebisch, Celso Furtado y de otros grandes pensadores latinoamericanos que presentan un conjunto de contribuciones originales, cuyo objetivo buscaba comprender los problemas de las economías latinoamericanas desde una perspectiva crítica (Quintero, et al, 2022, pp3).

El enfoque estructuralista se originó bajo la premisa de establecer importantes contribuciones de economistas latinoamericanos con Raúl Prebisch, quien resaltó los desequilibrios estructurales entre países desarrollados y en desarrollo, particularmente en términos de comercio e industrialización (Botta et al., 2018).

Así mismo se determinó que uno de los elementos principales del estructuralismo parte de la premisa de que los problemas del desarrollo no derivan solamente de la escasez de capital, sino de las estructuras productivas y sociales desiguales que limitan el crecimiento sostenido (Bresser-Pereira, 2019, p. 18).

Para Fernández y Ormaechea (2021), la característica particular de la propuesta estructuralista fue situar en el núcleo del análisis las trayectorias diferenciadas de desarrollo, planteando el conflicto y el poder como elementos centrales. De igual modo, Schaposnik (2003)

clasifica las corrientes principales de la teoría de la dependencia (citados en Quintero Etl al 2022).

Pese a la integridad del enfoque estructuralista en la noción de desarrollo económico surgen contras. Una crítica del enfoque estructuralista es su dependencia de la intervención estatal, que puede ser difícil de implementar de manera efectiva en países en desarrollo con capacidad administrativa limitada. Los críticos argumentan que las políticas complejas pueden ser más costosas y menos eficientes que las soluciones más simples basadas en el mercado (Chenery, 1975).

Asimismo, el Decreto 064 de 2011 establece el desarrollo para la política pública como la expansión de las libertades de las personas para llevar una vida prolongada, saludable y creativa; conseguir las metas que consideran valiosas y participar activamente para darle forma al desarrollo de manera equitativa y sostenible. Las personas son a la vez beneficiarias y agentes motivadores del desarrollo humano, como individuos colectivamente (Alcaldía Mayor de Bogotá D. C., Decreto 064, 2011, p. 3)

En consecuencia, el desarrollo económico de la ciudad propuesto en la política pública integra elementos en los cuales los efectos de su aplicación, impacten las nociones del desarrollo humano que a la luz de los avances que se presenten, impacten positivamente en las condiciones de vida de las personas que habitan en Bogotá.

Productividad y Competitividad

Las variables determinantes para catalogar e impulsar el desarrollo económico en las ciudades y países, naturalmente está relacionado con la productividad y competitividad. Las ciudades que prosperan en la economía mundial a menudo tienen sectores de exportación robustos, que crean empleos y atraen inversión internacional. Esto requiere mejoras en los sistemas educativos para

producir una mano de obra calificada y el desarrollo de infraestructura para apoyar a las empresas globalmente competitivas (Rondinelli et al., 1999).

La productividad es determinante para garantizar un sostenido crecimiento desarrollo económico de los territorios. Una mayor productividad, además de incrementar los ingresos de los distintos sectores económicos de las ciudades, también puede liberar mano de obra para otros sectores, impulsando así el crecimiento económico y el desarrollo estructural en el territorio (Rondinelli et al., 1999).

Es así como la productividad de los distintos sectores incide de manera integral en la capacidad de las ciudades para generar condiciones que favorezcan el crecimiento y el desarrollo. Por ello, resulta importante destacar aquellos elementos que son determinantes para el fortalecimiento productivo de una ciudad o región.

En primer lugar, el capital financiero constituye un factor fundamental, en la medida en que actúa como uno de los principales motores de la productividad urbana. Su importancia radica en que permite materializar recursos en inversión, infraestructura, innovación y fortalecimiento de las capacidades productivas. Bajo esta perspectiva, el Estado cumple un papel central, no solo en la definición de rubros de financiamiento público, sino también en la generación de condiciones institucionales, económicas y territoriales que favorezcan la atracción de inversión y la captación de capital.

No obstante, la inversión por sí sola no garantiza la maximización de la productividad; es trabajo el eje esencial para la cimentar la productividad. En el modelo clásico el auto Solow, 1956 planteó el trabajo junto con el capital como los factores básicos del crecimiento económico y naturalmente de productividad.

Esta relación sugiere que los factores macroeconómicos y financieros influyen conjuntamente en la productividad y el desarrollo económico, ofreciendo perspectivas de política para las economías emergentes (Cakici, 2024).

Es así como el trabajo, junto con el fortalecimiento de las capacidades humanas a través de la educación, la formación y el desarrollo de habilidades, constituye un factor fundamental para mejorar la productividad y la competitividad. Una mano de obra calificada contribuye a elevar la eficiencia de los distintos sectores productivos, y también favorece la innovación, la adaptación tecnológica y el aprovechamiento de nuevas oportunidades económicas (Flores et al., 2022).

En términos reales, la correlación entre productividad y desarrollo económico de una ciudad como Bogotá la función de producción Cobb–Douglas constituye un modelo ampliamente utilizado para explicar la relación entre los factores de producción y la productividad en una economía. Su formulación básica puede expresarse como:

$$P = A K^{\alpha} L^{\beta} T^{\gamma}$$

En la que la productividad está determinada precisamente K como el capital, L el trabajo, T la innovación y la tecnología. Sin embargo y en el caso particular de la producción en una ciudad como Bogotá A corresponde a un factor de eficiencia global, pero este caso respondería a la eficacia de la Política Pública de Desarrollo Económico debido a que son elementos intrínsecos en las variables para determinar el crecimiento económico de una ciudad.

Asimismo, la valoración de esta función en la evaluación de la productividad de una ciudad resulta pertinente para evaluar la incidencia de la inversión pública y privada, empleo formal e informal, y la adopción de tecnologías emergentes en el desarrollo económico de la

ciudad, permitiendo determinar el impacto de la Política Pública de Desarrollo Económico para la ciudad de Bogotá.

En casos de estudio similares se refleja que las políticas de ciudades inteligentes, particularmente en las ciudades orientales, tienen un efecto más pronunciado en la productividad, destacando la importancia de enfoques personalizados basados en el tamaño de la ciudad y las características regionales (Yao, 2024).

Lo anterior permite establecer la relevancia de la tecnología en la potenciación de los procesos productivos de las ciudades, en la medida en que el desarrollo tecnológico puede generar efectos positivos sobre su capacidad productiva, haciéndolas más efectivas, eficientes y competitivas.

En esta misma línea, Hommann y Lall (2019), a partir del caso de África subsahariana, destacan que para avanzar hacia ciudades más productivas y habitables es necesario abordar problemas estructurales asociados con la planificación urbana, los mercados de tierras y el financiamiento de la infraestructura. Resultan fundamentales, dado que inciden directamente en la capacidad de las ciudades para organizar su crecimiento, mejorar sus condiciones productivas y fortalecer su desarrollo urbano.

Así mismo en África, a pesar de las notorias limitaciones de la región; el crecimiento de la productividad está impulsado por la eficiencia de escala y el progreso tecnológico, aunque las limitaciones de datos plantean desafíos para el análisis integral (Shen y Valdmanis, 2022).

Adicionalmente, en la valoración de las dinámicas productivas es posible que la liberalización del comercio no siempre arroje resultados positivos de productividad, como se vio en Vietnam, donde la competencia interna y las políticas ineficaces obstaculizaron el crecimiento (Dung, 2024).

Por lo anterior, resulta pertinente establecer políticas articuladas que garanticen los medios necesarios para avanzar en el crecimiento económico y en el fortalecimiento de la productividad de las ciudades. La evaluación de la política pública de desarrollo económico en Bogotá involucra elementos esenciales para identificar si la ciudad cuenta con mecanismos adecuados para optimizar la productividad, atraer inversión y fortalecer el trabajo como factor central del desarrollo.

En coherencia con lo anterior, es claro que la maximización de los procesos productivos debe considerarse en relación directa con la competitividad. Esto, en la medida en que la competitividad de las ciudades puede entenderse como su capacidad para maximizar la productividad y el bienestar de sus ciudadanos (Vegara, 2016, p. 3).

Por lo tanto, la competitividad resulta fundamental para el desarrollo urbano, dado que no se limita únicamente al ámbito privado o empresarial, sino que integra dimensiones económicas, sociales y ambientales. En consecuencia, su propósito no debe reducirse al crecimiento económico, sino orientarse también al mejoramiento de la calidad de vida, el bienestar de la población y la generación de condiciones sostenibles para el desarrollo de las ciudades.

Así mismo la competitividad es relevante entendiendo que puede configurar la interconexión entre desarrollo social, productividad y crecimiento, enfatizando que estos temas, si bien aparentemente diferentes, son cruciales para lograr el objetivo de brindar una vida digna a la población a través de políticas públicas coherentes que se alineen con la realidad social (Manjarrés, 2016)

La capacidad general para fijar márgenes de competitividad necesariamente implica determinar la rivalidad. Se trata, también, de un concepto relativo, ya que se es competitivo si se está en condiciones de competir. La posición relativa frente al agente rival se constituye como

una de las variables determinantes del éxito o fracaso competitivo de la unidad analizada (Montoya, Sánchez & Baldenebro, s.f)

Porter, M. citado en Suñol (2006) señala que:

“La competitividad consiste en la capacidad para sostener e incrementar la participación en los mercados internacionales, con una elevación paralela del nivel de vida de la población. El único camino sólido para lograrlos se basa en el aumento de la productividad” (p.179).

El análisis comparativo de los enfoques internacionales revela diferentes estrategias para lograr la competitividad en las ciudades. Estos incluyen el desarrollo de ventajas competitivas a través de la innovación tecnológica, estrategias de interacción en red y la adquisición de nuevos conocimientos (Zhulanov y Oksman, 2024).

Por su parte Bogotá es la ciudad más competitiva de Colombia, según el Índice Departamental de Competitividad (IDC) de 2023, un estudio realizado por el y la . El IDC analiza 102 indicadores agrupados en 13 pilares, y Bogotá destacó especialmente en la adopción de tecnologías, el sistema financiero y el ecosistema de innovación (Secretaría de desarrollo económico de Bogotá)ConsejoPrivado de Competitividad (CPC)Universidad del Rosario

Para el año 2025 la universidad del rosario halló que con un puntaje de 8,13 sobre 10, Bogotá se posiciona nuevamente como el territorio más competitivo del país, liderando en 9 de los 13 pilares que componen la medición, entre ellos infraestructura, sistema financiero, sofisticación y adopción de TIC. Antioquia y Valle del Cauca se ubicaron en segundo y tercer lugar con puntajes de 6,82 y 6,30, respectivamente (Universidad del Rosario, 2025).

Sin embargo, en el ámbito de la innovación y la competitividad, Bogotá ha sido reconocida tanto a nivel nacional como internacional. En 2023, la ciudad ocupó el primer lugar

en el índice de competitividad para ciudades en Colombia, consolidándose como referente en este ámbito.

Por su parte, el Índice Mundial de Innovación (GII) evalúa la capacidad innovadora de las economías a nivel global, mientras que el Índice de Competitividad Global (ICG) se centra en la competitividad de los países. En la edición 2024 del GII, Colombia se posicionó en el puesto 61 a nivel mundial, lo que refleja los avances y retos que aún enfrenta en materia de innovación (OMPI, 2025).

La productividad y competitividad en el desarrollo urbano están influenciadas por diversos factores. Entre los más determinantes se encuentran el empleo, el tiempo aplicado en el desarrollo de actividades, la implementación de tecnologías emergentes y la innovación en los sectores productivos. Asimismo, resulta pertinente establecer una participación ciudadana significativa y un sistema legal confiable, así como considerar las variables relacionadas con los salarios, ya que la productividad laboral se ve particularmente impactada por estos aspectos (Soria Romo & López Portillo Tostado, 2018). Quizá, en materia social la pobreza y la desigualdad son las problemáticas más comunes y evidentes que enfrentan los países en desarrollo como es el caso de Latinoamérica.

En primer lugar, la pobreza existe en la gran mayoría de países del mundo a pesar de los objetivos de desarrollo sostenible fijados para 2030 para erradicarla. En común inferencia, algunos países y de acuerdo con la lógica económica, la erradicación parece lejos de la realidad del mundo, por lo cual se las acciones globales se han reorientado a la reducción de la pobreza en proporciones aceptables.

En esta dirección, la noción de pobreza ha cambiado a lo largo de los años. Una de las primeras nociones del mundo moderno fue expuesta por Amartya Sen, en su texto, fundamento

las condiciones del buen vivir el desarrollo, estableciendo así herramientas al PNUD, para lo cual establece en primer lugar que la pobreza debe entenderse como privación de capacidades humanas, cuyas causas son “la mala gobernanza, la falta de oportunidades y factores personales” (García y García, 2020)

En efecto, la noción moderna de pobreza atribuye directa responsabilidad a los Estados, entendiendo que su capacidad limitada de acciones gubernamentales, ocasionando efectos negativos en los cuales las personas carecen de bienes y servicios públicos, y altos niveles de desigualdad irremediablemente generando condiciones en las impacta directamente en la producción de pobreza como relación directamente proporcional se destaca.

En un sentido similar, se ha estudiado la pobreza y desigualdad en Costa Rica, dando una mirada más allá de la distribución de los ingresos. Estos autores se enfocan en comparar las necesidades básicas insatisfechas y las características espaciales de la pobreza, mediante análisis descriptivos, técnicas de georreferenciación y regresión ponderada. En esta investigación, se concluye: “El país necesita redefinir su estrategia de desarrollo para gestar un círculo virtuoso de crecimiento económico con mayores niveles de desarrollo humano (Arias, et al. 20220, p.24 citado por Medina et al, 2021).

La pobreza multidimensional en África Subsahariana se relaciona estrechamente con la falta de servicios básicos y educación (Devdiscourse, 2025). Asimismo, según el Banco Mundial (2024), las desigualdades territoriales en el acceso a servicios públicos de calidad en Colombia particularmente en departamentos como La Guajira, Chocó o Amazonas han contribuido a la reproducción de la pobreza intergeneracional.

A inicios de siglo, el banco mundial concibe como un fenómeno multidimensional, basado en tres ejes: “la falta de ingresos y recursos para atender las necesidades básicas

(incluyendo educación y salud), la falta de voz y poder en las instituciones estatales y en la sociedad, y la vulnerabilidad frente a choques adversos y la incapacidad de enfrentarlos” (García y García, 2020, p. 168

La pobreza puede considerarse como un estado de múltiples privaciones de individuos, grupos o comunidades respecto a su bienestar. Esto significa que los individuos carecen o son deficientes en capacidades socioeconómicas y privaciones de bienes básicos para la vida cotidiana” (Botchway, 2013,).

Los efectos de la pobreza se acompañan de exclusión social como incapacidad de los ciudadanos pobres no tienen la posibilidad de autoorganizarse, tener participación y efectiva representación política” (Kliuchnyk, 2022, p. 1). Además, se evidencia que las personas en condición de pobreza, suelen tener más obstáculos para acceder a los servicios de salud y educación debido a efectos de la incapacidad o limitada acción gubernamental.

No obstante, la limitación de los bienes y servicios públicos, es países con desmedidos niveles de desigualdad, en la distribución territorial de los espacios, la población en condición de pobreza resulta desplazada a zonas periféricas, marg finalizando la pobreza y agrupando en espacios concretos en el territorio donde paradójicamente la oferta de servicios públicos como saneamiento básico , salud, y educación son escasos, ahondando más en los efectos de la desigualdad y la pobreza.

En un estudio reciente, se destaca cómo Aliaga y Álvarez (2010) abordan la distribución espacial de la calidad de vida a través de estratos socioeconómicos para el caso particular de Bogotá (citado por Medina et al, 2021.). La segregación de la pobreza en muestra una relación compleja entre variables espaciales y variables socioeconómicas, por lo que las zonas periféricas

y marginales determinan un factor determinante una correlación para la aglomeración y concentración de personas con pobreza en la ciudad

Por su parte, Alfonso (2017) evidencia la tendencia de los hogares bogotanos a pagar más por vivienda en zonas con menor densidad poblacional y mayor dotación de activos públicos (citado por Medina et al, 2021.).

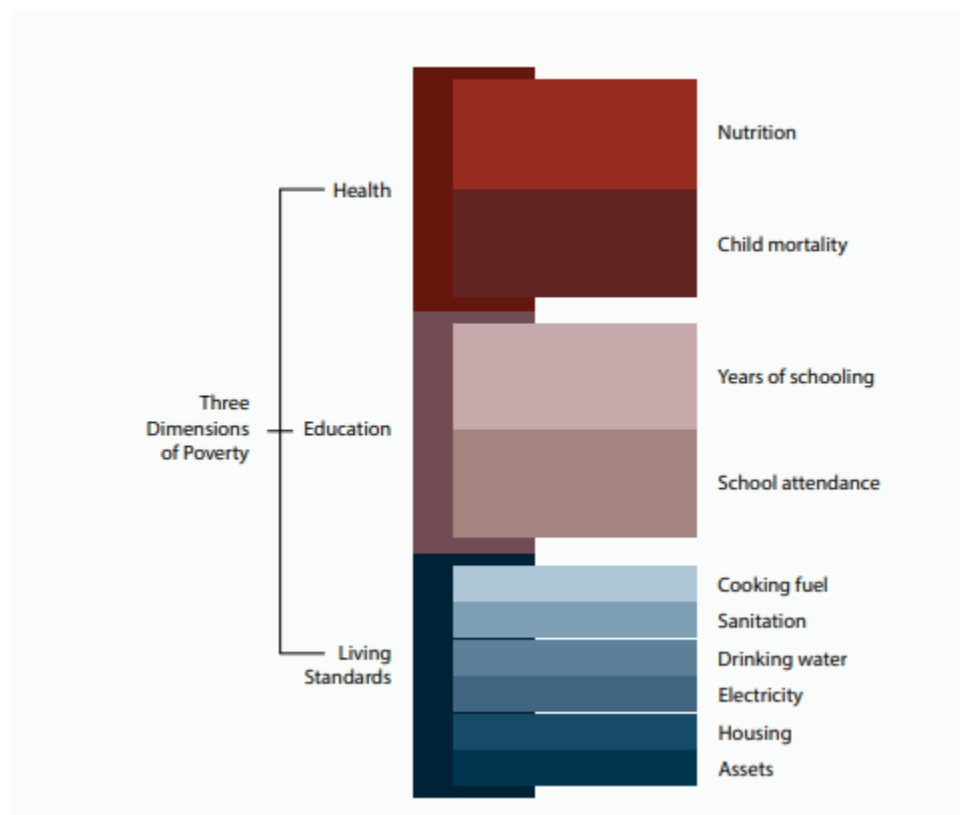
Es así como el método de necesidades Básicas Insatisfechas (NBI) establecido desde los años setenta, permitió ampliar el concepto de pobreza al incluir educación, salud, vivienda, agua potable, saneamiento y transporte, superando la visión restringida a la alimentación y la subsistencia (Rocha, 2006; Crespo & Gurovitz, 2002 Citado por santo, 2013)

Por otro lado, la pobreza multidimensional es entendida como una situación en que las personas o los hogares carecen de acceso a servicios básicos como salud, educación, vivienda y transporte” (Moreno Sandoval & De Albuquerque Jannechevitz, 2023, p. 63).

En un estudio de Multidimensional Poverty, se estima que la pobreza multidimensional refleja las diversas privaciones que enfrentan las personas en múltiples aspectos de la vida, como la salud, la educación, las condiciones de vida y las actividades del mercado laboral.

Figura 2

Índice de Pobreza Multidimensional de la Universidad de Oxford



Nota. Obtenido por Oxford Poverty and Human Development Initiative (OPHI), 2021

El Índice de Pobreza Multidimensional (IPM) aborda la pobreza a través de tres dimensiones centrales: salud, educación y nivel de vida. En la primera, se consideran indicadores como la nutrición y la mortalidad infantil; en la segunda, los años de escolaridad y la asistencia escolar; y en la tercera, aspectos vinculados al acceso a combustibles de cocina seguros, saneamiento, agua potable, electricidad, vivienda y activos básicos.

De este modo, la pobreza se entiende como un fenómeno que no solo depende del ingreso monetario, sino de un conjunto de privaciones simultáneas que afectan la calidad de vida de los hogares, ofreciendo así una visión más integral para orientar las políticas públicas (Programa de las Naciones Unidas para el Desarrollo, 2020).

En principio parece una problemática principalmente visible en países con economías endebles y en desarrollo, lo cual, dado las capacidades institucionales, complejiza la garantía de acceso a dimensiones de salud, educación y entorno de infraestructura más adecuado en las ciudades.

Estudios empíricos muestran que la pobreza multidimensional se manifiesta con mayor intensidad en zonas rurales que en las urbanas, debido a privaciones en saneamiento, servicios básicos y educación (Helfand et al., 2009; Kageyama & Hoffmann, 2006 citado por Santos, 2013).

Modelos de Machine Learning en Política Pública

Particularmente las políticas públicas no han involucrado directamente los modelos de machine learning en los ciclos de la política pública. Lo anterior se ahonda en la limitada producción literaria sobre la integración de estas herramientas para validar la acción gubernamental, para lo cual se reitera que los modelos de machine learning y las políticas públicas, es una relación prematura, pero con prospectiva emergente.

Según Riascos (2025), el machine learning permite que los gobiernos pasen de decisiones basadas en intuición o juicios cualitativos a modelos predictivos y causales fundamentados en datos masivos, lo cual incrementa la eficiencia institucional.

En los últimos años, el aprendizaje automático ha proporcionado diferentes técnicas y algoritmos para predecir situaciones basadas en grandes cantidades de información que, a través de un adecuado procesamiento y filtrado de datos, pueden generar predicciones muy efectivas (Corba & Negre, 2024, p. 210 citado por Hernández, 2024 p26).

En la fase de diseño, los modelos de machine learning posibilitan identificar patrones ocultos y segmentar poblaciones mediante algoritmos supervisados y no supervisados. Esto permite caracterizar con mayor precisión a los beneficiarios y definir prioridades territoriales. Por ejemplo, en el sector salud colombiano, los modelos de gradient boosting y random forest mejoraron el ajuste de riesgo en el sistema de aseguramiento, optimizando la asignación de recursos públicos y reduciendo incentivos a la selección adversa (Riascos, 2025, p. 7).

En casos relacionados, los modelos de machine learning, además de su implementación en temas asociados a la política de salud en el país, también se han aplicado en diferentes ciudades de Colombia para apoyar procesos de decisiones en materia de seguridad. En estos casos, se evidenció que la aplicación de modelos estructurales y de aprendizaje por refuerzo

permitió distribuir de forma óptima el personal policial, logrando reducir hasta en un 8 % los delitos en Bogotá, sin aumentar el presupuesto operativo (Riascos, 2025, p. 19).

En concordancia, en este trabajo se desarrolla bajo la pretensión de identificar más óptimos los modelos de machine learning en la evaluación de la política pública, para lo cual se valoran modelos de aprendizaje supervisado y no supervisado.

Los algoritmos de ML pueden clasificarse en supervisados, no supervisados y parcialmente supervisados. La principal diferencia entre estos tres tipos de algoritmos radica en la presencia o ausencia de una variable de resultado o dependiente, a la cual se le denomina marca o etiqueta (Pedrero et al., 2021, p. 12).

Estos enfoques permiten comprender que el desarrollo socioeconómico requiere una lectura articulada de múltiples dimensiones económicas y sociales. Desde esta perspectiva teórica, el machine learning complementa los enfoques de desarrollo económico

Para el caso de Amartya Sen, el desarrollo permite justificar la inclusión de variables como pobreza multidimensional, empleo y desigualdad territorial, cuya observación requiere métodos capaces de identificar patrones complejos y heterogéneos como sucede en la ciudad de Bogotá.

En línea con lo anterior, para Romer, la centralidad del conocimiento, la innovación y el uso eficiente de los recursos se relaciona con modelos que permiten analizar dinámicas de productividad, competitividad y crecimiento a partir de grandes volúmenes de información.

Por su parte, Stiglitz advierte que el PIB ignora factores vitales para la sociedad, lo que exige incorporar análisis multidimensionales para comprender el desarrollo socioeconómico

De esta forma, el machine learning aporta herramientas pertinentes para procesar diversos indicadores para detectar relaciones no lineales y reconocer diferencias territoriales que fortalecen el seguimiento y la evaluación de la política pública.

Modelos de Machine Learning para Evaluar Política

En la revisión literaria, destaca modelos de aprendizaje supervisado como regresión lineal, support vector regression, random forest y árbol de decisiones como los referentes en la implementación de acciones gubernamentales. Así mismo, en modelos no supervisados K-Means clustering y redes neuronales, siendo la última la más destacada en la implementación de modelos de machine learning para variabilidad y adaptabilidad para predecir eventos.

Particularmente, el aprendizaje supervisado generalmente dispone previamente conjuntos de datos de entrenamiento que son elementos etiquetados. Este modelo se entrena con un conjunto de datos etiquetados, lo que significa que para cada entrada en el conjunto de entrenamiento hay una salida esperada o etiqueta asociada. A través de este proceso, el modelo aprende a asociar las entradas con sus respectivas salidas, de modo que pueda predecir correctamente la etiqueta de nuevas entradas (Hernández,2024,p. 26).

Por otro lado, el aprendizaje no supervisado los modelos se entrena con datos no etiquetados, lo que significa que no tiene una salida esperada en el conjunto de entrenamiento. El objetivo del modelo es identificar patrones o estructuras en los datos por sí mismo, sin guía explícita sobre lo que se pretende hallar con la base de datos utilizada (Torrealba, 2023 citado por Hernández, 2024, p.26).Tabla 4Modelos de machine learning en políticas públicas

Tabla 4

Modelos de Machine Learning en Políticas Públicas

Política evaluada	País/Sector	Principales hallazgos	Título del documento	Bibliografía
--------------------------	--------------------	------------------------------	-----------------------------	---------------------

Política de vivienda, acceso y vivienda	Brasil – vivienda social	Random Forest	Rent vouchers y ayudas monetarias más efectivas que adquisición de casas (Furtado & Andreão, 2022)	Furtado & Andreão (2022)
Política pública de la pobreza	México/ Reducción de la pobreza	No se especifica. sin embargo, el documento establece las técnicas de aprendizaje automático son comúnmente utilizadas para conciliar los beneficios estimados con los pesos de bienestar asignados a diferentes personas, los efectos heterogéneos del tratamiento de la intervención de la política	La Evaluación de la política refleja la priorización de la implementación en grupos Etnico	Björkegren, D., Blumenstock, J. E., & Knight, S. (2022). (Machine) Learning What Policies Value. arXiv.Org, abs/2206.00727 . https://doi.org/10.48550/arXiv.2206.00727
Educación, salud y servicios sociales (evaluación de equidad en ML)	EE. UU / políticas públicas sociales	No especifica el modelo empleado, pero precisa métodos como el posprocesamiento que implican elegir umbrales de puntuación específicos de grupo, eliminan consistentemente las disparidades.(L	El hallazgo tiene encontrado por el estudio tiene implicaciones importantes principalmente para la por las contribuciones en las investigación de ML y su despliegue ML en las	Lamba, H., Rodolfa, K. T., & Ghani, R. (2021). An Empirical Comparison of Bias Reduction Methods on Real-World Problems in High-Stakes Policy Settings. Sigkdd Explorations,

crecimiento en PYMES	Crecimiento económico/ Italia	La evaluación de la política de socio crecimiento en el caso de estudio se implementaron técnica de IA como modelos predictivos	Se destaca la importancia de modelos y herramientas para el análisis de datos y cómo la IA puede integrarse eficazmente en los procesos de formulación de políticas para apoyar el desarrollo de las PYME	23(1), 69–85. https://doi.org/10.1145/3468507.3468518 Perboli, G., Tronzano, A., Rosano, M., Tarantino, L., & Velardocchia, F. (2021). Using machine learning to assess public policies: a real case study for supporting SMEs development in Italy. 1–6. https://doi.org/10.1109/TEMSCON-EUR52034.2021.9488581
Políticas de salud pública (mortalidad infantil, atención de urgencias, gasto público, etc.)	Sector Salud/N/A	Random Forest, Causal Forest, BART, Gradient Boosting, SVM, regresión logística	El estudio señala machine learning permite establecer un balance el as variables y el correcto procesamiento de los datos implementados con los modelos reduce los sesgos en los análisis de las políticas. Sin embargo, pocos estudios reportan métricas de desempeño, y	Lourenço, L., Weber, L., Garcia, L., Ramos, V., & Souza, J. (2024). Machine Learning Algorithms to Estimate Propensity Scores in Health Policy Evaluation: A Scoping Review. International Journal of Environmental Research and Public Health,

políticas de seguridad ciudadana	Colombia / Seguridad ciudadana	Máquinas de soporte vectorial para regresión (SVR)	<p>en sistemas simples no supera a regresión logística.</p> <p>Contribuye a una optima predicción del modelo reflejando una disminución de hurtos en las principales ciudades y adicionalmente, permitió configurar elementos para mejorar la toma de decisiones en materia de la política pública a nivel sectorial.</p>	<p>21(11), 1484. https://doi.org/10.3390/ijerph21111484</p> <p>Ordóñez, H., Cobos, C., & Bucheli, V. (2020). Modelo de machine learning para la predicción de las tendencias de hurto en Colombia. <i>Revista Ibérica de Sistemas e Tecnologías de Informação</i>, (E29), 494-506.</p>
---	--------------------------------	--	---	--

Nota. Fuente: Elaboración propia partir de la revisión bibliográfica en bases de datos abiertas.

Los antecedentes revisados evidencian que la integración de modelos de machine learning en el ámbito de las políticas públicas, tanto a nivel regional como global, aún se encuentra en una fase incipiente de consolidación y exploración.

En otras palabras, apenas se están estableciendo las bases metodológicas e instrumentales para la incorporación de estas herramientas en los contextos de aplicación de acciones gubernamental. Adicionalmente, la literatura revisada muestra que su uso en procesos estrictamente orientados a la evaluación de políticas públicas continúa siendo limitado y desigual.

Metodología y Aplicación Metodológica

En el contexto del análisis económico y de desarrollo en la evaluación de políticas públicas, el ML puede identificar los factores que contribuyen a la desigualdad económica y ayudar a cerrar la brecha digital, apoyando así intervenciones específicas para reducir las disparidades (Gao, 2024).

La persistencia de estas brechas en los indicadores de desigualdad y en el crecimiento heterogéneo entre localidades impone retos significativos a la acción gubernamental, que debe diseñar e implementar políticas capaces de abordar de manera simultánea la exclusión económica, la informalidad y la segregación territorial.

El uso de Machine Learning en la evaluación de políticas públicas plantea importantes desafíos relacionados con la interpretación de los mecanismos y herramientas empleadas en la predicción de eventos, así como con la presentación de los resultados a la ciudadanía en general.

Bajo esta perspectiva, para la construcción del estado del arte y del marco teórico se realizó una revisión bibliográfica orientada a identificar estudios sobre evaluación de políticas públicas, desarrollo económico y aplicación de modelos de machine learning en contextos de aplicación de política públicas. La búsqueda se concentró en literatura académica en español e inglés, priorizando artículos científicos, revisiones, documentos de trabajo, tesis y documentos institucionales con pertinencia metodológica y temática para el problema de investigación.

El presente estudio adopta un enfoque mixto secuencial explicativo, siguiendo el marco metodológico propuesto por Creswell y Plano Clark 2018, en el cual una fase cualitativa precede y orienta a una fase cuantitativa, y los resultados de ambas se integran para responder de manera complementaria a la pregunta de investigación. Esta elección obedece a la naturaleza del problema abordado, que combina la necesidad de comprender el marco normativo y conceptual

de la Política Pública de Productividad, Competitividad y Desarrollo Socioeconómico de Bogotá con la necesidad de modelar empíricamente sus indicadores estratégicos mediante aprendizaje automático.

La primera fase, de carácter cualitativo documental, consistió en una revisión sistemática de literatura académica nacional e internacional sobre evaluación de políticas públicas y aplicación de aprendizaje automático en el sector público, complementada con el análisis de los documentos oficiales que estructuran la política, en particular el Decreto 064 de 2011 y el documento CONPES Distrital 25 de 2023. Esta fase permitió delimitar los conceptos clave, identificar los indicadores estratégicos que serían modelados y seleccionar las técnicas analíticas pertinentes con base en la evidencia internacional.

La segunda fase, de carácter cuantitativo aplicado, consistió en la construcción de bases de datos a partir de fuentes oficiales del Departamento Administrativo Nacional de Estadística, la Secretaría Distrital de Desarrollo Económico, la Secretaría Distrital de Planeación, la Cámara de Comercio de Bogotá y el Observatorio de Desarrollo Económico de Bogotá, y en la implementación, validación y comparación de ocho modelos de aprendizaje automático supervisado y no supervisado. Los hallazgos de la fase cuantitativa se interpretan a la luz del marco construido en la fase cualitativa, lo cual permite que la lectura técnica de los resultados se ancle al contexto institucional y conceptual de la política evaluada.

La integración entre las dos fases se realiza en la discusión, donde los resultados cuantitativos se contrastan con la literatura revisada y con los objetivos institucionales declarados por la política pública. Este diseño es coherente con la naturaleza del estudio, en tanto reconoce que la evaluación de políticas públicas requiere tanto comprensión contextual como evidencia

empírica, y permite que la monografía aporte simultáneamente al conocimiento aplicado del campo y al fortalecimiento de la práctica institucional de evaluación en el Distrito Capital.

La búsqueda documental se realizó en buscadores y bases académicas como Google Scholar, Scopus, ScienceDirect y repositorios institucionales de bases de textos de uso abierto, complementada con la revisión de referencias de artículos identificados y Para la localización de los documentos, se realiza mediante tesauros combinaciones de palabras clave en español e inglés, como “machine learning”, “public policy”, “machine learning” and policy evaluation”, “public policy”, “evaluación de políticas públicas con machine learning”, “desarrollo económico”, “machine learning” y “policy analysis” y “predictive models”.

En la selección e inclusión de los documentos bibliográficos, se discriminaron mediante la relación directa con la evaluación de políticas públicas, el análisis de acciones gubernamentales y la aplicación de modelos de machine learning con relación directa en la formulación, implementación y evaluación de políticas públicas.

Naturalmente, se priorizaron artículos científicos, revisiones, documentos de trabajo y estudios de caso con desarrollo metodológico claro y modelos.

Ahora bien, como criterios de exclusión se descartaron documentos duplicados, textos sin relación directa con políticas públicas, trabajos centrados exclusivamente en desarrollos técnicos sin aplicación a las acciones gubernamentales

La búsqueda inicial permitió identificar un conjunto amplio de documentos.

Posteriormente, se efectuó un proceso de depuración sucesiva con base en criterios de pertinencia temática, calidad académica, relación con el campo de las políticas públicas y utilidad metodológica para el desarrollo del estudio. Como resultado, se seleccionaron 14 textos

que ofrecen soporte conceptual y metodológico para el análisis de modelos de *machine learning* aplicados a la evaluación de políticas públicas.

Sin embargo, es necesario precisar que, durante la revisión bibliográfica realizada a partir de los tesauros previamente establecidos, no se hallaron suficientes antecedentes que expusieran de manera explícita modelos de machine learning implementados específicamente en procesos de evaluación de política pública. En términos generales, la literatura encontrada se concentra más en el uso de estas herramientas para apoyar la toma de decisiones, la predicción de fenómenos sociales o económicos, estableciendo un marco limitado dentro de esquemas formales de evaluación de políticas con modelos de machine learning.

El estudio utiliza exclusivamente fuentes secundarias oficiales. Los datos provienen de las siguientes instituciones: el Departamento Administrativo Nacional de Estadística (DANE), a través de la Gran Encuesta Integrada de Hogares (GEIH) y las cuentas nacionales departamentales; la Secretaría Distrital de Desarrollo Económico (SDDE), mediante los informes de seguimiento del CONPES D.C. 25; la Secretaría Distrital de Planeación (SDP), a través de la Encuesta Multipropósito; la Cámara de Comercio de Bogotá, con los registros de creación, cancelación y vigencia empresarial por localidad; y el Observatorio de Desarrollo Económico de Bogotá, con indicadores de productividad, competitividad e informalidad laboral desagregados por localidad para el período 2011- 2025.

Las variables utilizadas en la presente documento, corresponden exclusivamente a bases de datos oficiales disponibles en los anexos del trabajo, las cuales fueron recopiladas a partir de fuentes institucionales reconocidas, entre ellas el Departamento Administrativo Nacional de Estadística (DANE), la Secretaría Distrital de Desarrollo Económico, la Secretaría Distrital de

Planeación, la Cámara de Comercio de Bogotá, entre otras entidades públicas y privadas con funciones estadísticas y de seguimiento económico para la ciudad de Bogotá.

En coherencia con el objetivo general de evaluar comparativamente los modelos de Machine learning para el apoyo al seguimiento de la política pública de desarrollo económico en Bogotá, la presente monografía parte, en primer lugar, del análisis de la política pública estudiada y de sus principales indicadores. En segundo lugar, y en correspondencia con el objetivo específico de identificar modelos de *machine learning* pertinentes para su evaluación, se seleccionaron técnicas supervisadas como la regresión lineal, la regresión logística, el árbol de decisión, Random Forest y SVR.

De manera complementaria, se incorpora K-means como modelo de aprendizaje automático no supervisado, esto para explorar agrupamientos territoriales o socioeconómicos, con la determinación de identificar heterogeneidades en la dinámica del desarrollo económico en Bogotá. De igual forma, se integran modelos de redes neuronales, ya que su naturaleza de procesamiento permite capturar relaciones complejas que posibilitan un análisis óptimo de los datos.

Tabla 5

Modelos de Machine Learning para Evaluar Política Pública

Modelo	Tipo	Variable dependiente (Y)	Variables independientes (X)
Regresión lineal	Supervisado – Regresión	PIB de Bogotá	Periodo / año
Regresión logística	Supervisado –	Desempeño empresarial (alto/medio/bajo)	Creación neta, rotación empresarial
SVR	Supervisado –	Ocupados, desocupados, Gini	Crecimiento PIB, IED, variación anual IED
Árbol de Decisión	Supervisado	Desempleo, pobreza o desempeño empresarial	Informalidad, Gini, pobreza, desempleo, Año, creación neta,

Random Forest	Supervisado	Desempleo o pobreza	rotación Informalidad, Gini, pobreza, desempleo, Año
Gradient Boosting	Supervisado	Desempleo o pobreza	Informalidad, Gini, pobreza, desempleo, Año
K-Means	No supervisado – Agrupamiento	No aplica	Producción, ventas, empleo y variaciones anuales del sector manufacturero
Red neuronal MLP	Supervisado	PIB trimestral de Bogotá desde el 2011 a 2025	Sectores CIIU, rezagos del PIB y variables temporales

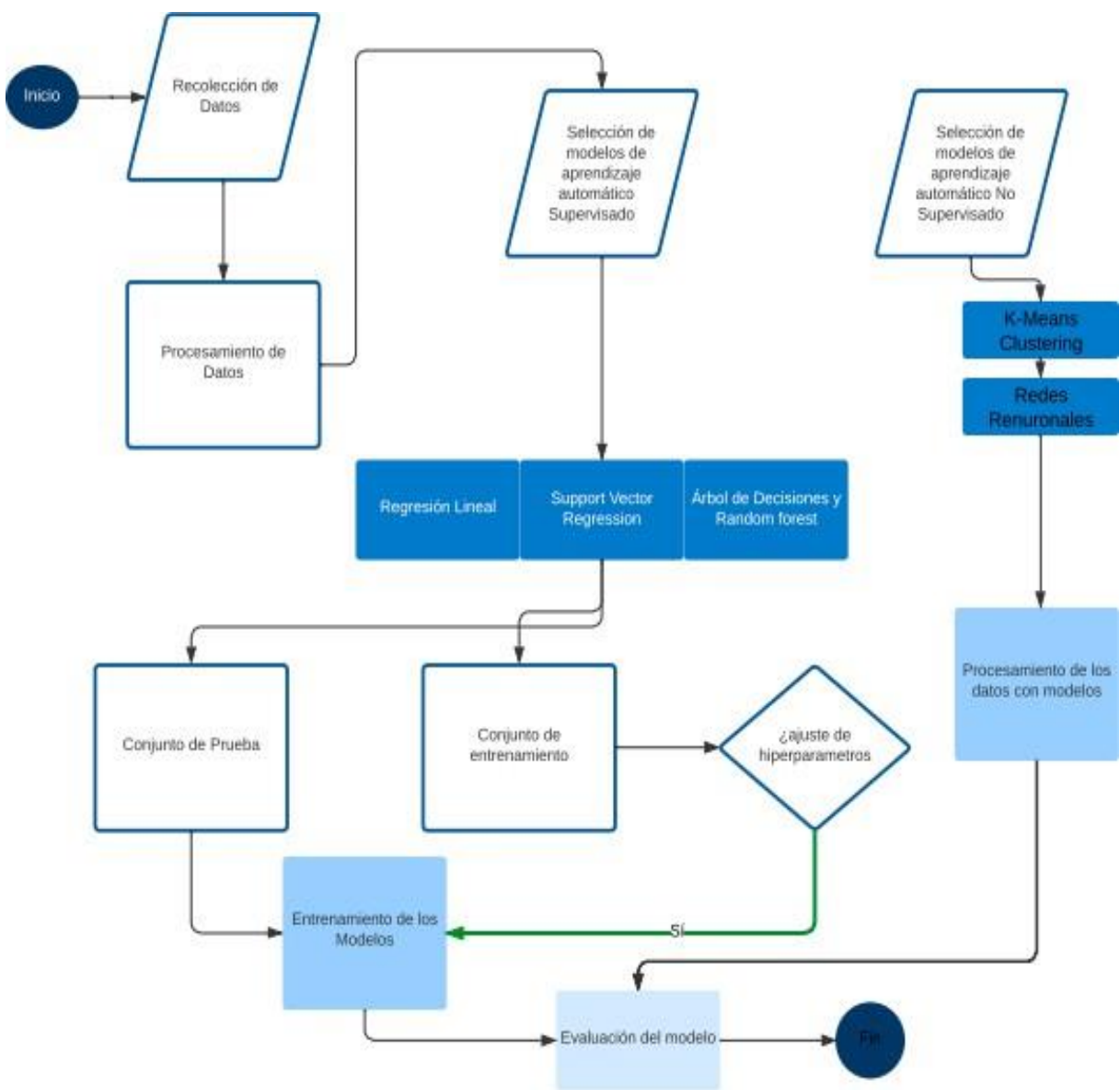
Nota. : Fuente: elaboración propia

El documento se orienta principalmente a contrastar técnicas como SVR, árbol de decisión, random Forest, gradient boosting, K-means y redes neuronales. Sin embargo, para obtener mayor profundidad e interpretabilidad, no todos los modelos operaron bajo condiciones idénticas ni sobre variables objetivo equivalentes. Adicionalmente, para el modelo de redes neuronales MLP se empleó un esquema de validación temporal tipo TimeSeriesSplit, preservando el orden cronológico de la serie.

La selección de modelos se basa en la revisión bibliográfica de estudios relacionados con la aplicación de estas herramientas en el análisis y la evaluación de políticas públicas. En este sentido, los estudios revisados evidencian que Random Forest ha sido aplicado en políticas de vivienda social en Brasil y en evaluaciones del sector salud; las máquinas de soporte vectorial para regresión (SVR) han sido utilizadas en análisis de seguridad ciudadana en Colombia; y revisiones en salud pública identifican también el uso de modelos como gradient boosting, SVM y regresión logística. Aunque en algunos casos la literatura no precisa de manera uniforme el modelo implementado, sí permite identificar un conjunto de técnicas recurrentes en el análisis de problemáticas públicas.

Figura 3

Procedimiento para el Procesamiento de Datos



Nota. Fuente: Elaboración propia

La comparación entre los modelos implementados como una valoración comparativa de su utilidad analítica frente a distintos indicadores y estructuras de datos, de acuerdo con la naturaleza de cada problema abordado.

Política Pública de Productividad, Competitividad y Desarrollo Socioeconómico de Bogotá

La política pública formulada mediante el Decreto 064 de 2011 para la Productividad, Competitividad y Desarrollo Socioeconómico de Bogotá “tiene como fines principales la promoción del crecimiento económico y el posicionamiento competitivo de Bogotá, la creación y desarrollo de alternativas productivas, el fortalecimiento empresarial, la generación de ingresos y oportunidades de empleo y la promoción de las capacidades y potencialidades de las personas, para el mejoramiento del ambiente competitivo y la consolidación de Bogotá y la Región Capital como foco de desarrollo socioeconómico en el contexto nacional e internacional, ampliando la base de generación y distribución de la riqueza a todos los estratos de la sociedad (Alcaldía Mayor de Bogotá D. C., 2011, p3).

Bajo el anterior, la política pública delimita el alcance de tiempo desde e 2011 hasta 2023 para dar cuenta el crecimiento económico, el fortalecimiento empresarial, la generación de ingresos y oportunidades de empleo, y el desarrollo de capacidades de las personas en la ciudad que dinamicen los enfoques productivos y sociales.

De igual forma el propósito objetivo pretende Promover el crecimiento económico y el posicionamiento competitivo de Bogotá, la creación y desarrollo de alternativas productivas, el fortalecimiento empresarial, la generación de ingresos y oportunidades de empleo, y la promoción de las capacidades y potencialidades de las personas, para el mejoramiento del ambiente competitivo y la consolidación de Bogotá y la Región Capital.”(Decreto 064 de 2011 y CONPES D.C. 2023 pp. 52-53).

En el año 2023, la secretaria distrital de desarrollo y comprendiendo la transformación global y local de las necesidades, y la dinámica de la ciudad posterior a la pandemia del covid, se

plasma la necesidad de actualizar el plan de acción de la política pública de productividad, competitividad y desarrollo socioeconómico 2011-2038

La actualización de la política busca comprender las ciudades como motores de productividad, orientando sus intervenciones hacia acciones transversales a las distintas actividades económicas. Es decir, se enfoca en medidas no exclusivamente sectoriales, dirigidas a promover las unidades productivas, potenciar el capital humano, aprovechar las ventajas derivadas de las economías de aglomeración, fortalecer las relaciones productivas entre los diferentes actores y promover, de manera integral, el desarrollo competitivo desde los territorios, involucrando y beneficiando tanto a las personas como a la sociedad en su conjunto (CONPES D.C., 2023, p. 9).

Tabla 6

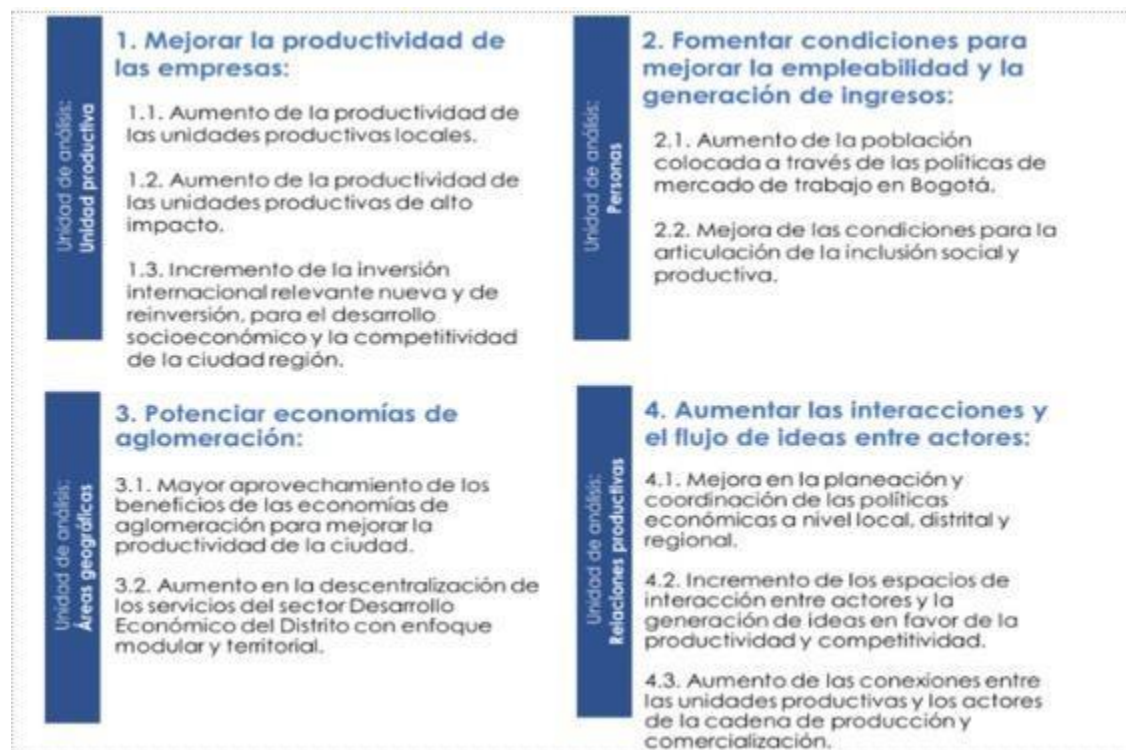
Marco Conceptual de la Actualización de la Política Pública

Determinante	Descripción	Autores / Fuentes	Relación con la política pública
Eficiencia de las firmas	Capacidad de las unidades productivas para transformar insumos en productos. Se logra mediante acceso a financiamiento, cambio tecnológico, innovación y aprendizaje organizacional ('aprender haciendo').	Cusolito & Maloney (2018); Hausmann (2013); Helpman (2004)	Promueve el fortalecimiento empresarial, la innovación y la productividad, objetivos centrales de la Política Pública de Productividad, Competitividad y Desarrollo Socioeconómico de Bogotá.
Conocimiento y capital humano	Resultado de la acumulación de capital y talento humano; incluye la innovación y el desarrollo tecnológico como factores de productividad. El conocimiento es un bien	Romer (1990); Helpman (2004)	Refuerza la necesidad de invertir en educación, formación y talento humano como base para el desarrollo productivo distrital.

	intangible que impulsa la creación de nuevos y más sofisticados bienes.		
Instituciones	Conjunto de normas, organizaciones y mecanismos que permiten la coordinación entre el sector público, privado y académico dentro del ecosistema de productividad.	Cusolito & Maloney (2018); Helpman (2004)	Favorece la gobernanza económica y la articulación interinstitucional (Secretarías, empresas, academia), principio estructural de la política distrital.
Inter instituciones	Relaciones funcionales y la integración de las unidades productivas en mercados regionales e internacionales.	Helpman (2004); López Ghio et al. (2018)	Integración económica Bogotá–Región y la participación en cadenas globales de valor. Alianzas de mercados regionales e internacionales para favorecer el comercio

Nota. Fuente: Elaboración propia con base en Cusolito & Maloney (2018); Helpman (2004); y conpes 25 de Política Económica y Social del Distrito Capital (CONPES D.C., 2023).

En la actualización de la política pública, y ante los nuevos retos derivados de la pandemia por COVID-19 que impactaron a la capital del país, se reorientaron los objetivos y la acción gubernamental hacia aquellos aspectos que se busca atender mediante su implementación. Esta revisión permitió estructurar cuatro objetivos fundamentales: mejorar la productividad de las empresas; fomentar condiciones para fortalecer la empleabilidad y la generación de ingresos de las personas; potenciar las economías de aglomeración; y aumentar las interacciones y el flujo de ideas entre los diferentes actores productivos de Bogotá

Figura 4*Objetivos de la Política Pública de Productividad, Competitividad y Desarrollo Socioeconómico*

Fuente: Tomado de Consejo Distrital de Política Económica y Social del Distrito Capital 25 “política pública de Productividad, competitividad y desarrollo económico.

Conforme los objetivos establecidos, el distrito ha adoptado elementos de seguimiento anual para validar el efectivo avance de la política pública en comparación con los objetivos principales. Para el año de 2023, la secretaria de planeación indico que el corte a diciembre 31 de 2023, el avance acumulado de la política alcanzó un 48,98%.

A 31 de diciembre de 2023, el informe de Seguimiento Plan de Acción Política Pública de Productividad, Competitividad y Desarrollo Socioeconómico 2011-2038 de la secretaria de planeación se resaltan diferentes estrategias y actividades que aportan al ejercicio real y plural en la promoción del crecimiento y desarrollo económico junto con el posicionamiento competitivo de Bogotá, con programas como “Bogotá Produce” llevando de 539 unidades productivas, a

2.592 Unidades Productivas del Distrito(*Secretaría Distrital de Desarrollo Económico, 2024, Informe de Seguimiento CONPES D.C. 25*)

De igual manera, el informe resalta en el marco de la implementación de la política pública realizaron capacitaciones empresariales a ciudadanas que participan de las manzanas del cuidado. “Con la Academia Financiera se fortalecieron 229 unidades productivas en proceso de formación. Se realizó articulación interinstitucional con 3 localidades: Barrios Unidos, Antonio Nariño y Bosa. Se hizo fortalecimiento empresarial a emprendimientos, con el curso de higiene y manipulación de alimentos a través del SENA(*Secretaría Distrital de Desarrollo Económico, 2024, Informe de Seguimiento CONPES D.C. 25*).

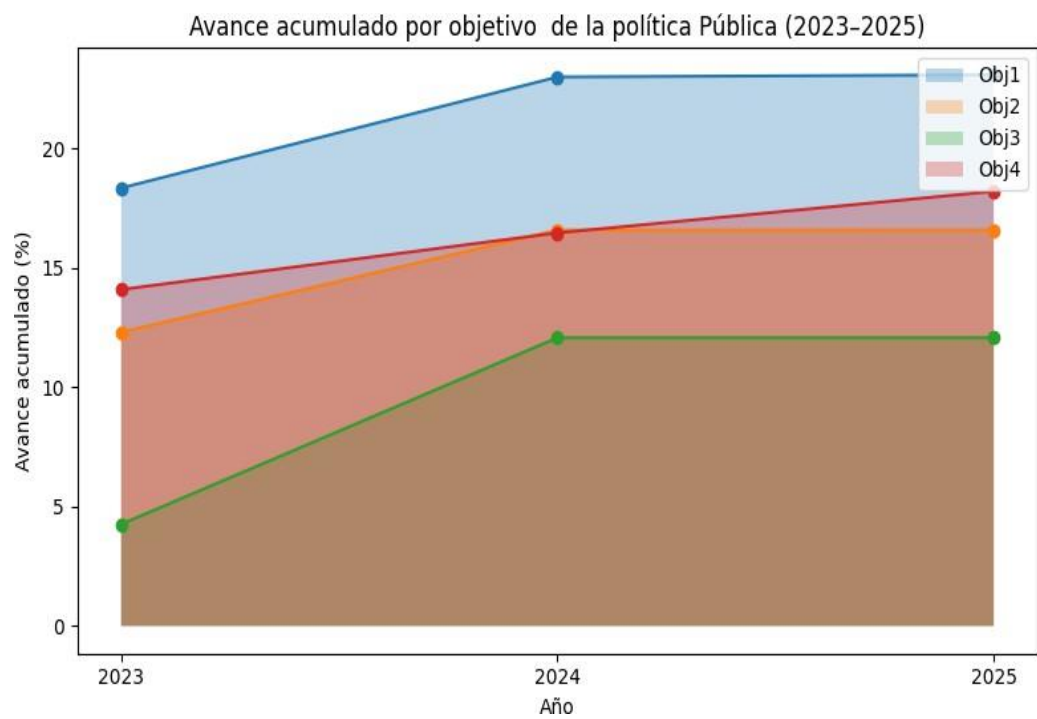
Para el año 2024, según el informe de implementación de ese periodo, la Política Pública de Productividad fortaleció de manera significativa el tejido empresarial de Bogotá mediante asistencia técnica a 4.912 unidades productivas, acceso a financiamiento para 6.387 negocios y procesos de formalización para 1.284 unidades. Un aumento considerable a lo reportado en el año 2023.

En paralelo, se avanzó en empleabilidad con 14.879 personas formadas, 69.383 colocaciones laborales y 22.399 orientaciones realizadas por la Agencia Distrital de Empleo. Para el distrito, estas acciones evidencian el impacto integral de la política en competitividad, generación de oportunidades y dinamización económica del territorio (*Secretaría Distrital de Desarrollo Económico, 2024*).

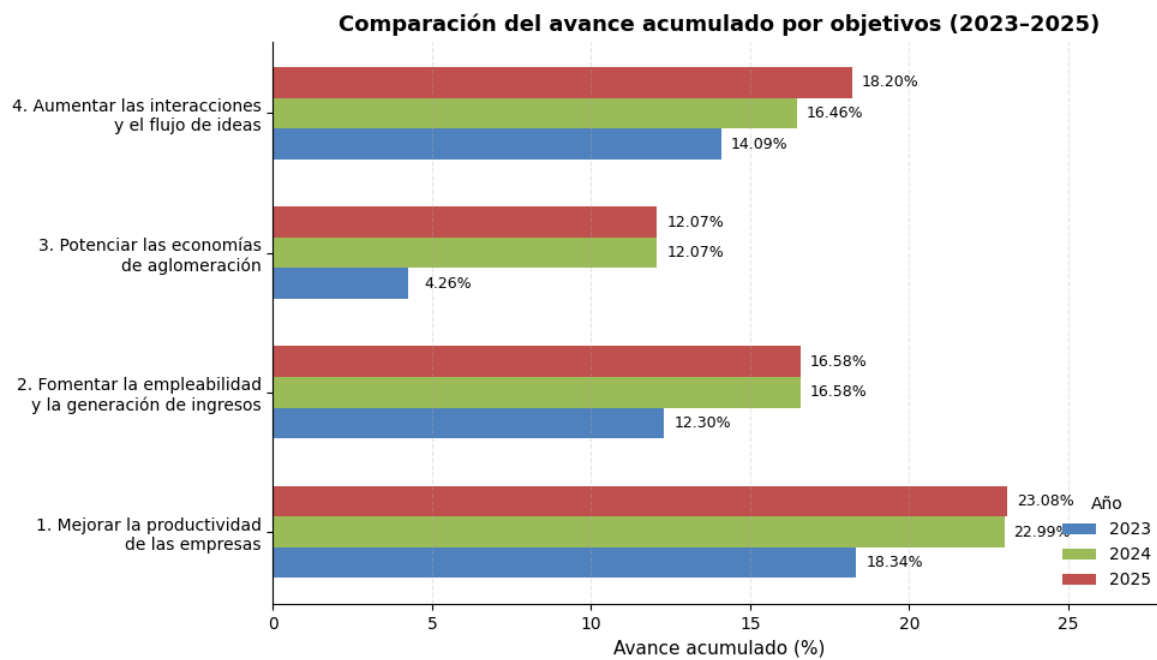
Ya La concreción de 21 proyectos de inversión en la región de Bogotá, que se espera genere más de 3,000 empleos directos y una inversión estimada de más de 56 millones de dólares, representa un aporte significativo al desarrollo económico de la ciudad.

Figura 5

Avance Acumulado de la Política Pública 2023–2025



Fuente: Elaboración Propia realizado a partir del plan de acción Política pública de Productividad, Competitividad y Desarrollo Socioeconómico de Bogotá

Figura 6*Comparación del Avance Acumulado por Objetivos Estratégicos**Nota.* Fuente: Elaboración propia

Resultados

Aprendizaje Supervisado

Regresión Lineal

En el aprendizaje automático supervisado comúnmente se abordan complejidades en diversos campos para resolver problemas complejos que requiere esquemas para abordarlos de forma efectiva. La regresión lineal es uno de los algoritmos de aprendizaje automático más simples y utilizados, siendo un método matemático empleado para realizar análisis predictivos” (Maulud & Abdulazeez, 2020, p. 140).

Naturalmente, el modelo comprende la relación entre una variable dependiente y variables independientes.

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Según Maulud y Abdulazeez (2020), la regresión lineal es uno de los algoritmos de aprendizaje automático más simples y utilizados, siendo un método matemático empleado para realizar análisis predictivos (p. 140). Así mismo, en aplicaciones reales, el modelo de regresión lineal se ha implementado en valoración 230 de políticas de seguridad para el caso colombiano.

Ordóñez, H., Cobos, C., & Bucheli, V. (2020) en su texto Modelo de machine learning para la predicción de las tendencias de hurto en Colombia, destacan los elementos esenciales de la implementación del modelo de regresión lineal para determinar las acciones públicas en materia de hurtos e inseguridad con variables definidas para Colombia en sus principales ciudades, configurando herramientas que contribuyen a la mejora en las decisiones públicas.

El análisis de regresión resulta pertinente en el entendido que se emplea principalmente para la predicción y la estimación causal entre variables independientes y dependientes. Sin

embargo, el modelo de regresión lineal por sí solo muestra únicamente relaciones estadísticas y no necesariamente relaciones causales (Maulud & Abdulazeez, 2020, p. 141).

La pretensión de la implementación del modelo de regresión lineal, el crecimiento económico en la ciudad, y tener predicciones sobre conforme la evolución de los últimos años, también determinar como factor predictivo cuál será su alcance en los próximos 3 años.

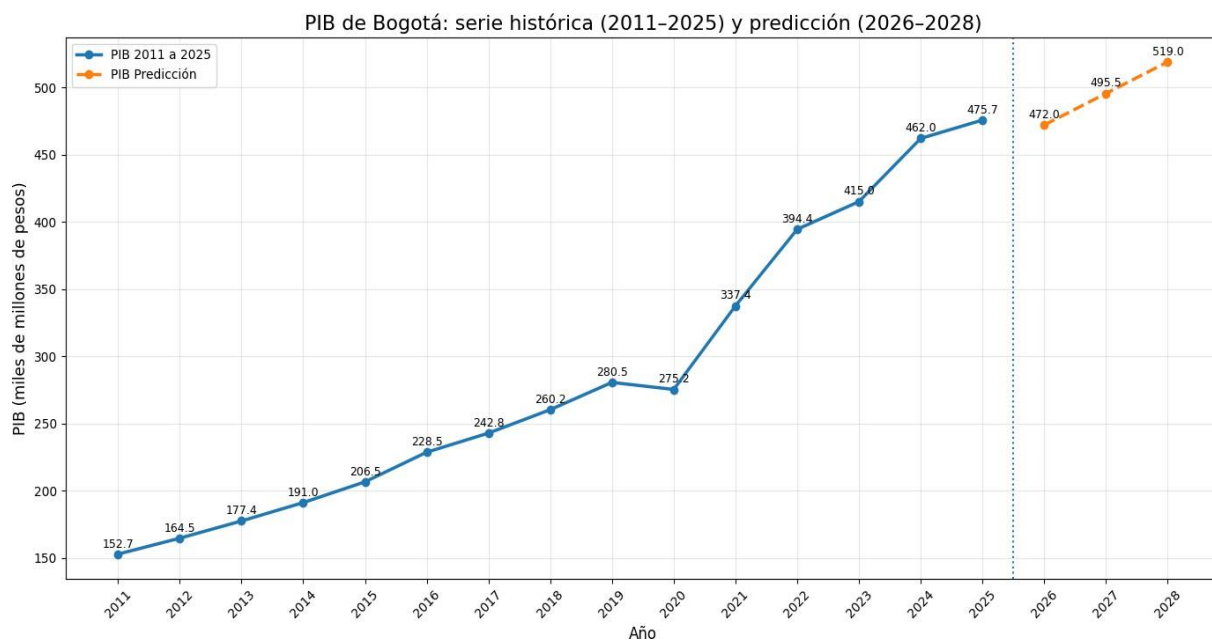
Para ello se implementó tomando como variable dependiente el PIB de Bogotá y como variable independiente el periodo 2011–2025

Tabla 7

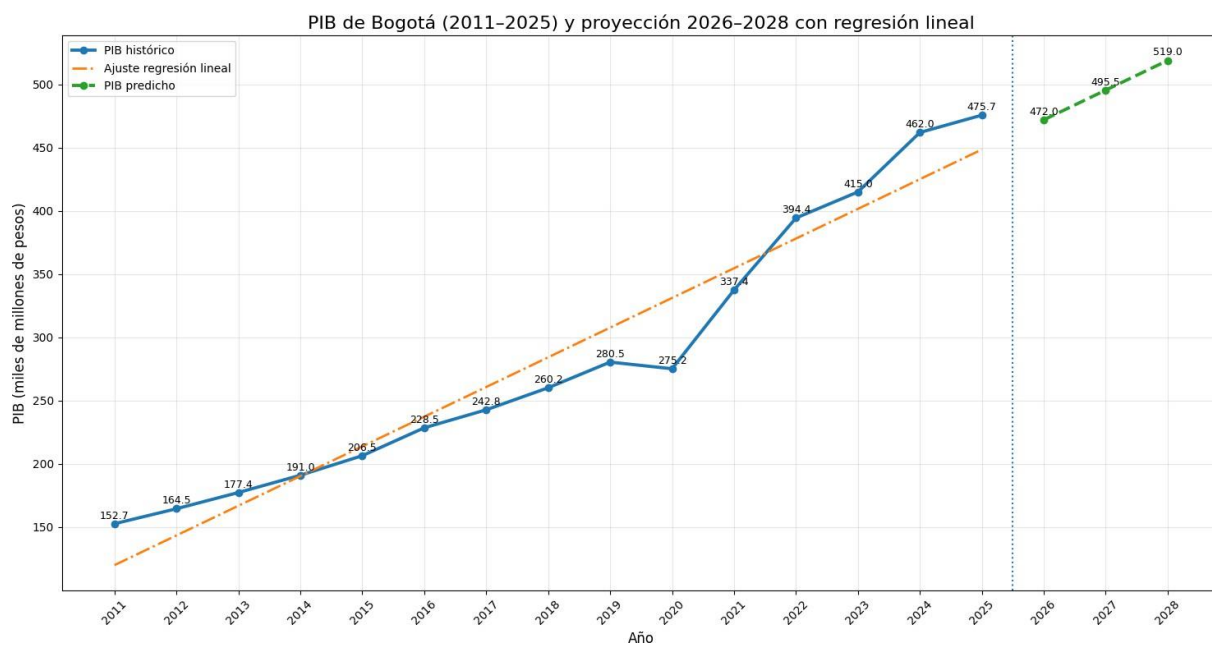
PIB, ocupación y tasa de producción per cápita 2005 -2025

Periodo	PIB Valor (miles de millones de \$)	Ocupación (miles de Personas)	Tasa de producción per capita
2005	89.832	-	
2006	100.462	-	
2007	112.441	-	
2008	122.101	-	
2009	130.672	-	
2010	139.740	-	
2011	152.691	3.675	149.016
2012	164.542	3.768	160.774
2013	177.361	3.814	173.547
2014	191.026	3.885	187.141
2015	206.478	3.880	202.598
2016	228.525	3.841	224.684
2017	242.755	3.758	238.997
2018	260.177	3.791	256.386
2019	280.473	3.870	276.603
2020	275.222	3.380	271.842
2021	337.429	3.596	333.833
2022	394.414	3.812	390.602
2023p	415.022	4.050	410.972
2024pr	462.030	4.218	457.812
2025 IIT	475.718	4.282	471.436

Fuente: Elaboración propia con datos recopilados y suministrados por el departamento administrativo de estadística DANE . Cifras en miles de millones

Figura 7*PIB Bogotá: Serie Histórica y Predicción (regresión Lineal)*

Fuente: Elaboración propia

Figura 8*PIB de Bogotá y Proyección 2026–2028*

Fuente: Elaboración propia

Función:

$$Y = -47080.65 + 23.47X$$

Por cada unidad adicional en la variable independiente, la variable dependiente aumenta en promedio 23.47 miles de millones de pesos el PIB, evidenciando una relación lineal positiva entre ambas variables.

R2: 0.9426506264506774

MAE: 21.16583111111183

RMSE: 25.012542375511398

Como variables independientes (X) se definieron la inversión extranjera directa, personas desocupadas, y personas ocupadas, lo anterior para determinar la relación de la productividad obtenida por las personas ocupadas y su relación con los márgenes de empleo, desempleo e inversión extranjera con el fin de establecer el impacto de la productividad en la ciudad de Bogotá desde 2011 hasta 2025

Tabla 8

Datos

Periodo	Monto Inversión Extranjer a	Empleos directos	Personas ocupadas	Personas desocupadas
2010	726	6.414	-	-
2011	1.041	6.832	3.675	392
2012	1.274	6.774	3.768	407
2013	2.674	13.025	3.814	385
2014	2.161	13.678	3.885	379
2015	2.110	11.642	3.880	379
2016	1.725	11.806	3.841	397
2017	2.131	17.383	3.758	450
2018	2.674	17.450	3.791	449
2019	3.229	21.254	3.870	472
2020	870	9.905	3.380	767
2021	1.497	16.868	3.596	693

2022	1.553	15.863	3.812	491
2023p	628	6.549	4.050	470
2024p	2.515	14.005	4.218	454
III-Trim 2025	631	8.292	4.282	417

Fuente: Elaboración propia a partir de la recopilación de datos de Secretaría distrital de planeación y departamento administrativo de estadística DANE

Para efectos del procesamiento de datos, se implementaron los datos hasta el 2024. En este caso se modeló el comportamiento de las personas ocupadas y desocupadas a partir del monto de inversión extranjera y los empleos directos. Para el caso de las personas ocupadas, el ajuste alcanzó un coeficiente de determinación de 0.256, lo que muestra una capacidad explicativa limitada. En el caso de las personas desocupadas, el modelo obtuvo un R^2 de 0.514, lo que indica un ajuste moderado, pero es preciso señalar que el modelo para el análisis de relación de variables, es

$$\text{Ocupados} = 3562.67 + 0.1952(\text{Inversión}) - 0.0133(\text{Empleos directos})$$

$$\text{Desocupados} = 482.00 - 0.1902(\text{Inversión}) + 0.0261(\text{Empleos directos})$$

Tabla 9

Métricas Regresión Lineal

Modelo	Variable dependiente	R²	MAE	RMSE
Regresión lineal	Personas ocupadas	0.256	117.17	161.87
Regresión lineal	Personas desocupadas	0.514	61.20	78.75

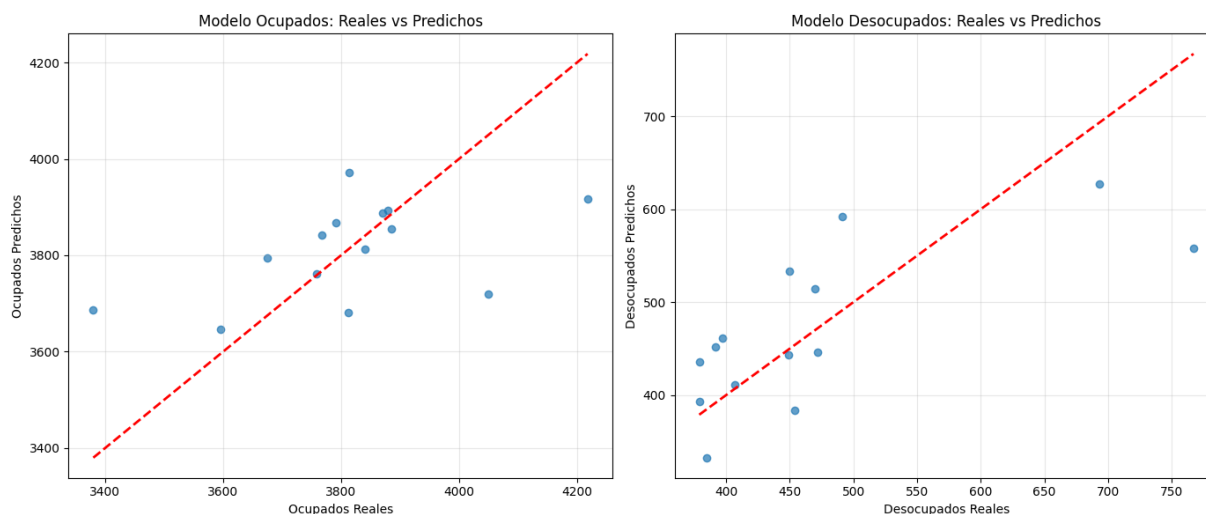
Fuente: Elaboración propia

Asimismo, el error absoluto medio MAE evidencia que, para la variable personas ocupadas, las predicciones del modelo presentan una desviación promedio de 117,17 unidades

respecto a los valores observados. En el caso de las personas desocupadas, el MAE es de 61,20, lo que indica una menor distancia promedio entre los valores reales y los estimados.

Figura 9

Regresión Lineal: Valores Reales vs. Predichos



Fuente: Elaboración propia

En la representación de los valores reales y predichos se observa dispersión en ambos casos, tanto para personas ocupadas como para personas desocupadas. No obstante, el modelo logra captar la tendencia general de la relación entre las variables analizadas. En particular, el modelo de personas ocupadas muestra una mayor alineación y correspondencia entre los valores observados y los estimados. En contraste, para las personas desocupadas se identifica una dispersión más marcada, lo que refleja limitaciones más evidentes en la precisión del ajuste.

La regresión lineal permite identificar una tendencia creciente del PIB en el periodo analizado y ofrecer una aproximación exploratoria de proyección de corto plazo. Sin embargo, debido al tamaño muestral anual reducido, sus resultados aportan complementariamente al análisis del comportamiento futuro del PIB en la ciudad de Bogotá.

Regresión Logística

El modelo de regresión logística es una herramienta de clasificación integrada de un modelo estadístico bien conocido que se utiliza comúnmente cuando el resultado es una variable (Bercu et al., 2019). En sentido pragmático, el modelo permite realizar la simulación para determinar la relación de variables económicas y su incidencia en el desempeño de los factores clave de la Política Pública de Desarrollo Económico en Bogotá. Se ha estructurado como variable dependiente los valores recopilados.

Para ello se categoriza el desempeño de empresas que se construye a partir de un índice compuesto que integra variables de dinamismo empresarial, constituido entre la creación neta de empresas y la cantidad de rotación el margen de empresas para las localidades de la ciudad de Bogotá

Para obtener la creación neta de empresas, se realiza la diferencia entre las empresas creadas y canceladas por periodo de año analizado puesto que refleja el crecimiento o contracción empresarial del territorio.

Adicionalmente, estas variables se complementan con la pobreza multidimensional, que captura privaciones en educación, salud, vivienda, trabajo y servicios públicos, y con el coeficiente de Gini, cuando se encuentra disponible como indicador en la base, el cual mide el grado de desigualdad en la distribución del ingreso en la ciudad.

La categoría de desempeño de empresas se construye a partir de da a partir de un índice compuesto que integra variables de dinamismo empresarial, constituido entre la creación neta de empresas y la cantidad de rotación el margen de empresas para de la ciudad de Bogotá

Tabla 10*Creación Neta de Empresas: Creadas-canceladas*

Localidad	Creadas							Canceladas						
	2019	2020	2021	2022	2023	2024	2025*	2019	2020	2021	2022	2023	2024	2025*
Antonio Nariño	1.576	1.137	1.282	1.344	1.247	1.340	834	803	915	859	1.029	1.098	1.488	847
Barrios Unidos	3.031	2.335	2.616	2.593	2.457	2.331	1.637	1.347	1.562	1.604	1.684	1.875	2.276	1.454
Bosa	3.855	2.797	3.864	4.314	3.854	3.780	2.506	2.713	2.493	2.555	2.765	2.909	3.862	2.554
Candelaria	684	472	522	691	638	646	448	295	361	342	341	390	596	380
Chapinero	5.339	3.801	4.405	5.018	4.556	4.476	3.020	1.779	2.036	2.222	2.182	2.423	3.003	1.800
Ciudad Bolívar	3.185	2.260	3.112	3.346	3.169	3.093	2.056	1.565	1.910	1.989	2.141	2.300	2.955	1.959
Engativá	7.282	5.675	6.998	7.019	6.530	6.216	4.344	3.615	4.104	4.256	4.662	4.917	6.604	4.213
Fontibón	3.469	2.667	3.287	3.374	3.157	3.170	2.309	1.600	1.793	1.918	2.021	2.153	2.889	1.862
Kennedy	8.974	7.060	8.240	8.297	7.588	7.365	5.177	4.690	4.948	5.172	5.381	5.796	7.548	5.124
Los Martires	2.739	2.011	2.306	2.664	2.529	2.435	1.784	1.266	1.474	1.390	1.561	1.745	2.340	1.433
Puente Aranda	3.482	2.824	3.260	3.150	3.081	2.762	1.930	1.673	1.870	1.988	2.146	2.293	2.912	1.877
Rafael Uribe Uribe	2.746	1.912	2.445	2.632	2.481	2.235	1.545	1.434	1.578	1.608	1.647	1.785	2.428	1.583
San Cristóbal	2.261	1.549	1.882	1.952	1.930	1.790	1.217	1.157	1.270	1.290	1.416	1.479	1.959	1.257
Santa Fe	2.515	1.827	2.167	2.399	2.408	2.391	1.544	1.079	1.201	1.257	1.335	1.459	2.087	1.266
Suba	10.571	7.839	9.038	9.384	8.481	8.425	5.913	4.599	5.252	5.475	5.658	5.961	8.352	5.386
Teusaquillo	2.398	1.782	2.237	2.259	2.218	2.051	1.443	1.050	1.273	1.298	1.329	1.458	2.075	1.287
Tunjuelito	1.608	1.182	1.388	1.544	1.349	1.309	959	744	916	896	998	1.016	1.325	890
Usaquén	6.879	5.381	5.828	5.901	5.632	5.701	3.984	2.800	3.105	3.256	3.439	3.629	4.811	2.967
Usme	1.677	1.403	1.761	1.744	1.629	1.525	1.089	585	814	790	994	1.108	1.425	1.027
Sin localidad	4.633	6.303	3.816	4.674	4.570	3.920	14.343	2.703	2.653	2.148	2.455	2.683	4.232	3.201
Total general	78.904	62.217	70.454	74.299	69.504	66.961	58.082	37.497	41.528	42.313	45.184	48.477	65.167	42.367

Fuente: Datos recopilados de la secretaria distrital para el desarrollo económico y secretaria de planeación de Bogotá

Tabla 11*Empresas Vigentes*

Localidad	2023	2024	2025
Antonio Nariño	8.367	8310	8278

Barrios Unidos	19.965	19724	19773
Bosa	18.186	18499	18336
Candelaria	4.127	4266	4278
Chapinero	39.850	40590	40628
Ciudad Bolívar	14.138	14028	14211
Engativa	37.916	37998	37578
Fontibón	20.775	21482	21675
Kennedy	42.425	41744	41945
Los Martires	17.823	18007	18314
Puente Aranda	20.444	20280	20397
Rafael Uribe Uribe	12.304	12063	12094
San Cristóbal	9.235	8970	9055
Santa Fe	15.165	15597	15435
Sin Localidad	19.844	20582	20171
Suba	52.145	52316	52521
Teusaquillo	14.907	14865	14554
Tunjuelito	7.422	7346	7592
Usaquén	42.398	43025	42889
Usme	7.274	7189	7349

Fuente: elaboración propia a partir de los datos suministrados por la secretaria de desarrollo económico de Bogotá. Asimismo, se consolidó el registro del número de empresas creadas y canceladas desde el 2019 hasta los datos obtenidos para septiembre de 2025 por localidad

Rotación empresarial = (creadas + canceladas) / vigentes

Posterior a ello, las variables de rotación empresarial en Bogotá y creación neta de empresas fueron normalizadas mediante el método Min–Max, aplicando la transformación de forma anual para cada una de las variables, con el fin de llevar los indicadores a una escala entre 0 y 1 y garantizar la comparabilidad relativa entre localidades desde el periodo 2023 hasta el III trimestre del periodo 2025.

En procesamiento de los datos mencionado, naturalmente fueron normalizados y después se establecieron las escalas para establecer las escalas del valor de creación neta de empresas.

$$= \text{Índice de dinamismo} = \text{Creación netanorm} + \text{Rotaciónnorm} / 2$$

Para optimizar el procesamiento de los datos, se clasificaron los datos obtenidos en la denominación del índice de dinamismo, ya que como variable dependiente (Y) la pretensión es determinar los valores de percentiles clasificados como Altos, medianos y bajos.

Bajo: $\leq P33$

Medio: P33–P66

Alto: $> P66$

La clasificación de las localidades en tres niveles de desempeño empresarial de bajo, medio y alto, se realizó mediante los percentiles 33 y 66 calculados de forma independiente para cada año analizado. Esta decisión metodológica responde a un criterio estadístico de distribución equiprobable de terciles, que divide la distribución en tres grupos de tamaño similar puesto que no hay umbrales dentro de la política definidos para establecer el índice de dinamismo empresarial. Este criterio de terciles es coherente con prácticas de análisis territorial aplicadas institucionalmente por el Departamento Nacional de Planeación (DNP) y la Secretaría Distrital de Planeación para clasificar localidades por niveles de desempeño socioeconómico (DNP, 2019).

La categorización del índice en tres niveles mediante percentiles 33 y 66 se adoptó como una estrategia operativa de segmentación relativa y no como una clasificación teórica cerrada del desempeño empresarial.

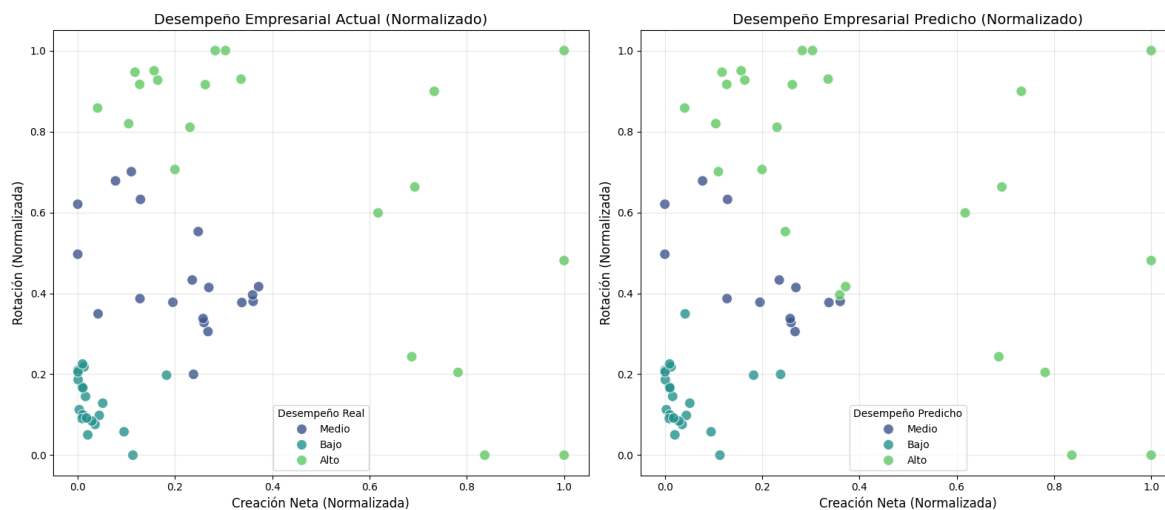
De esta forma, la aplicación de percentiles anuales, en lugar de percentiles sobre el conjunto agregado, garantiza la comparabilidad relativa entre localidades dentro de cada periodo, evitando que la clasificación quede condicionada por la distribución total de la serie.

Esto se debe a que los umbrales de clasificación son arbitrarios para una mejor explicación del índice de dinamismo empresarial en las localidades. Además pueden influir en la dificultad relativa de clasificación entre clases..

Esta decisión operativa implica que el accuracy reportado debe interpretarse teniendo en cuenta la limitación de observaciones y la distribución en tres grupos como se manifiesta anteriormente, ya que los umbrales de clasificación inciden en la dificultad relativa de cada clase. Para esto, se empleó información de empresas creadas, canceladas y vigentes para las localidades de Bogotá en los años 2023, 2024 y 2025

Figura 10

Desempeño Empresarial Actual-predicho (regresión Logística)



Fuente: Elaboración propia

El gráfico muestra una correspondencia alta entre la distribución real y la distribución predicha del desempeño empresarial. La similitud visual entre ambos resultados respalda el resultado el modelo reproduce adecuadamente los patrones observados del índice empresarial en Bogotá en las localidades en las categorías bajo, medio y alto, lo que es consistente con el accuracy de 0,90 alcanzado por la regresión logística.

Tabla 12*Métricas Regresión Logística*

Categoría	Precisión	Recall	F1-score	Soporte
Alto	0,84	1	0,91	21
Bajo	0,91	1	0,95	20
Medio	1	0,68	0,81	19

Fuente: Elaboración propia

A partir del reporte de clasificación, el modelo en la categoría alto, mediano el modelo obtiene una presión muy favorable, lo que confirma la adecuada clasificación del modelo en el índice de desempeño empresarial. En el modelo en categoría media, obtuvo precisión perfecta lo sugiere que el modelo para los años 2023,2024.2025 en las localidades con rendimiento de desempeño empresarial medio.

En la categoría alto, el modelo logró identificar correctamente todos los casos reales de desempeño empresarial alto, lo que refleja una alta capacidad para reconocer las localidades con mejores condiciones relativas en términos de dinamismo empresarial.

Ahora bien, la precisión inferior a 1,00 sugiere que algunas observaciones pertenecientes a otras categorías fueron clasificadas como alto.

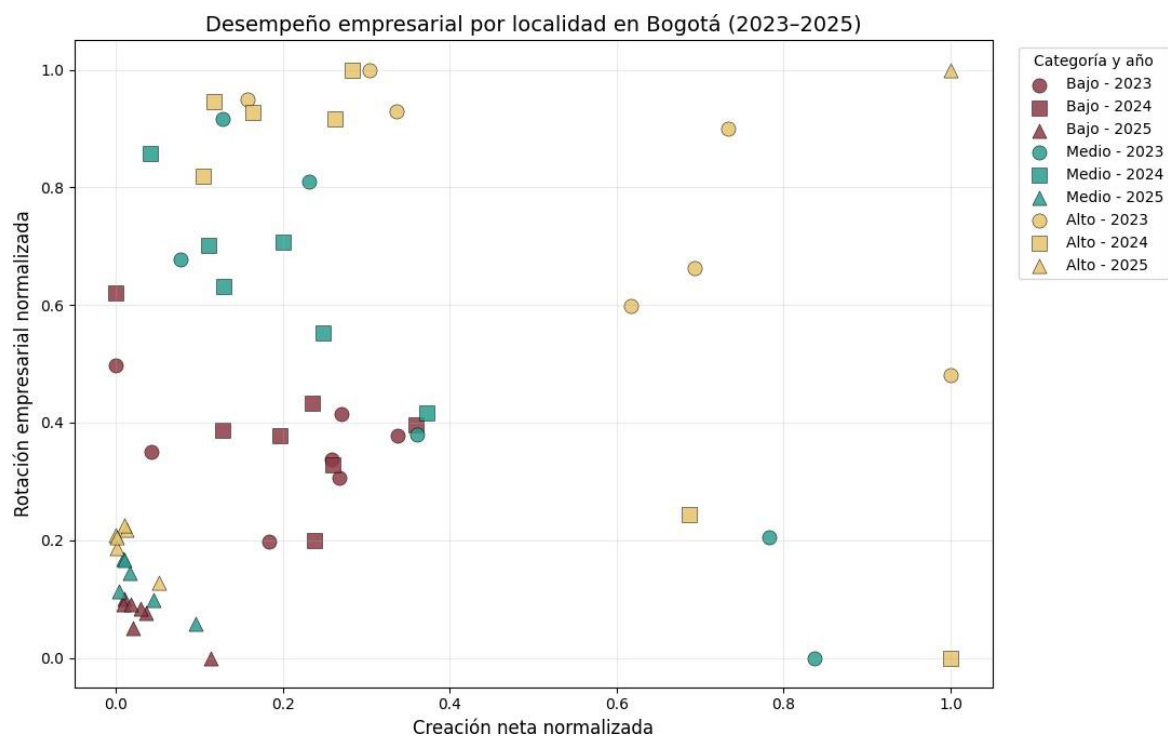
Por su parte, en la categoría bajo, el comportamiento del modelo en las métricas tuvo resultados más consistentes con una precisión de 0,91, un recall de 1,00 y un F1-score de 0,95, sobre 20 observaciones sugiriendo que los casos reales de bajo desempeño, la mayoría de las observaciones clasificadas en esta categoría correspondían efectivamente a ella.

Respecto a la categoría media, esta presentó una precisión de 1,00, un recall de 0,68 y un F1-score de 0,81, con un soporte de 19 observaciones. En conjunto, el resultado de las métricas refleja que el modelo las observaciones que efectivamente correspondían a un desempeño medio fueron asignadas a las categorías alto o bajo. Esto se debe a que la categoría medio porque

comparte características con las clasificaciones alto y medio en el desempeño empresarial en las localidades de Bogotá.

Figura 11

Desempeño Empresarial por Localidad en Bogotá 2023–2025



Fuente: Elaboración propia

Tabla 13

Clasificación de Rendimiento: alto, medio, bajo

Localidad	2023	2024	2025
Antonio Nariño	Bajo	Medio	Medio
Barrios Unidos	Bajo	Bajo	Bajo
Bosa	Alto	Alto	Alto
Candelaria	Bajo	Bajo	Bajo
Chapinero	Medio	Alto	Bajo
Ciudad Bolívar	Alto	Alto	Alto
Engativá	Alto	Bajo	Medio
Fontibón	Medio	Bajo	Medio
Kennedy	Alto	Medio	Medio
Los Mártires	Bajo	Bajo	Bajo
Puente Aranda	Bajo	Bajo	Bajo

Rafael Uribe Uribe	Medio	Alto	Alto
San Cristóbal	Medio	Alto	Alto
Santa Fe	Bajo	Medio	Bajo
Sin Localidad	Alto	Medio	Alto
Suba	Alto	Medio	Alto
Teusaquillo	Bajo	Bajo	Bajo
Tunjuelito	Medio	Medio	Medio
Usaquén	Medio	Alto	Medio
Usme	Alto	Alto	Alto

Fuente: Elaboración propia

Como resultado del procesamiento del desempeño empresarial en las localidades de Bogotá, se refleja que, para el periodo 2023–2025, el comportamiento fue diferenciado en el tiempo. Bosa, Ciudad Bolívar y Usme obtienen un desempeño alto durante los tres años, por lo que estas localidades, en los tres periodos analizados, mantienen elementos de consolidación empresarial y presentan condiciones favorables en la competitividad de las empresas ubicadas en estos territorios.

De igual forma, resulta importante resaltar que, geográficamente, estas localidades se ubican en el sur de Bogotá, por lo que el resultado permite inferir que en esta zona se concentran dinámicas empresariales constantes.

Support Vector Regression SVR

Por su parte el modelo SVR, en el proceso de evaluación de la Política Pública, el uso de modelos de machine learning basados en regresión permite comparar el comportamiento de diferentes algoritmos para identificar aquel que proporcione mejores resultados predictivos.

En este sentido, la literatura evidencia que la combinación de modelos de regresión lineal con técnicas más avanzadas, como la support vector regression (SVR), mejora la precisión en la estimación de variables dependientes relacionadas con fenómenos económicos o sociales.

Ordóñez, Cobos y Bucheli (2020) explican que “se evaluó el modelo con un modelo de regresión lineal estándar frente a un modelo de SVR estándar y al modelo ideal establecido por los datos reales. El modelo SVR minimiza la distancia entre cada punto de la variable dependiente al margen de tolerancia ϵ , logrando una menor desviación estándar de los residuos y mejorando la precisión de la predicción” (p. 501).

El modelo de SVR resulta pertinente, debido a que los permite analizar relaciones no lineales entre variables económicas y sociales relevante para la política pública de desarrollo económico en Bogotá. La aplicación de SVR permite además modelar el comportamiento y determinar las relaciones de las variables cuándo no necesariamente hay una relación lineal.

En este caso, el modelo se estructuró con tres variables predictoras (X) mediante el porcentaje de crecimiento del PIB, el monto de inversión extranjera directa (IED) y su variación anual; mientras que las variables objetivo (y) fueron las personas ocupadas, las personas desocupadas y el índice de Gini. Esta selección permite analizar cómo variables asociadas al dinamismo económico y productivo de la ciudad se relacionan con resultados del mercado laboral y con la desigualdad en Bogotá.

De este modo, el análisis contribuye a comprender cómo el desempeño económico se traduce en mejoras en las condiciones socioeconómicas de la población, aportando elementos para determinar el avance de la Política Pública de Desarrollo Económico en Bogotá.

Tabla 14

Datos

Periodo	Monto Inversión Extranjera	variación anual	IB % de crecimiento	Personas ocupadas	Personas desocupadas	personas desocupadas	de personas desocupadas	Pobreza Multidimensional	Indice de Gini	Pobreza Monetaria
2010	726	-	-	-	-	-	-	-	-	-
2011	1.041	43,3	4	3.675	67,6	392	9,6	-	0,522	13.1
2012	1.274	22,4	3,9	3.768	68,2	407	9,8	-	0,497	11,60%

2013	2.674	109,9	4,4	3.814	68,1	385	9,2	-	0,504	10.2
2014	2.161	-19,2	5,2	3.885	68,5	379	8,9	-	0,502	10.1
2015	2.110	-2,3	2,4	3.880	67,7	379	8,9	-	0,498	10.4
2016	1.725	-18,3	2,7	3.841	66,2	397	9,4	-	0,499	11.6
2017	2.131	23,5	1,9	3.758	64,1	450	10,7	-	0,498	12.4
2018	2.674	25,5	3,5	3.791	63,6	449	10,6	4,1	0,504	12.4
2019	3.229	20,7	3,8	3.870	63,3	472	10,9	7,1	0,513	-
2020	870	-73,1	-4,1	3.380	54	767	18,5	7,5	0,599	-
2021	1.497	72,2	12,8	3.596	56,6	693	16,2	5,7	0,565	-
2022	1.553	3,7	3,7	3.812	59,3	491	11,4	3,8	0,549	-
2023p	628	-59,6	1,2	4.050	62,3	470	10,4	3,6	0,532	-
2024p	2.515	300,6	3,1	4.218	64,1	454	9,7	5,4	0,523	-
III-Trim 2025	631	-6,2	3,9	4.282	65,1	417	8,9	-	-	-

Fuente: Elaboración propia a partir de los datos suministrados por el departamento administrativo de estadística, cámara de comercio de bogotá y Bogotá cómo vamos y la secretaría de desarrollo económico de Bogotá

El sistema de información para la medición de la pobreza en Colombia, administrado por el departamento administrativo nacional de estadística (DANE), ha integrado históricamente la pobreza monetaria, asociada al nivel de ingresos de los hogares, y la pobreza multidimensional, relacionada con las condiciones de vida de la población. Sin embargo, a partir de 2018 se fortalece el uso de la medición multidimensional con un mayor énfasis territorial, lo que permite analizar con más detalle las condiciones sociales en departamentos y ciudades principales

Este enfoque permite una valoración más integral del desarrollo, al incorporar dimensiones adicionales como educación, salud, condiciones de la vivienda, acceso a servicios públicos y condiciones laborales, superando las limitaciones de una medición centrada exclusivamente en desde una perspectiva de rango.

Por lo anterior, resulta pertinente implementar el modelo SVR de machine learning supervisado, con los datos de pobreza multidimensional, por lo que el análisis se basa a partir de 2018 hasta 2024, debido a que aún no se encuentran disponibles los valores para los cuatro trimestres del año 2025.

El índice de Gini se ha utilizado ampliamente como herramienta principal en varios análisis para investigar las variables que pueden afectar a la distribución de los ingresos consecuentemente agudice la desigualdad en el territorio (Sandoval, 2025).

Para Medina (2001), este indicador:

“Se clasifica entre las medidas estadísticas para el análisis de la distribución del ingreso, no utiliza como parámetro de referencia el ingreso medio de la distribución —a diferencia de la desviación media, la varianza y el coeficiente de variación—, dado que su construcción se deriva a partir de la curva de Lorenz”. (p. 7)

En general, la Curva de Lorenz representa gráficamente el nivel en cualquier ámbito en el que se quiera medir la desigualdad en la distribución de algún recurso o bienestar en una población territorial definida. A partir de esta curva se pueden calcular indicadores, como el coeficiente de Gini, que cuantifica el grado de la desigualdad, en una escala de 0 a 1 (donde 0 representa la igualdad perfecta y 1 la desigualdad total) y tiene relación con el área que queda encerrada entre la curva de igualdad perfecta y la curva de Lorenz (Vietri, & Del Duca, 2023).

Ya en la aplicación del modelo. previamente al ajuste de SVR, las variables independientes fueron estandarizadas mediante el método standard scaler, con el fin de llevarlas a una escala comparable y evitar que diferencias de magnitud entre los indicadores influyeran de manera desproporcionada en el aprendizaje del modelo.

Una vez limpiado y normalizado la base, se implementa el modelo de SVR, se obtiene como resultado de la aplicación del modelo para las variables objetivo

Personas ocupadas

Personas desocupadas

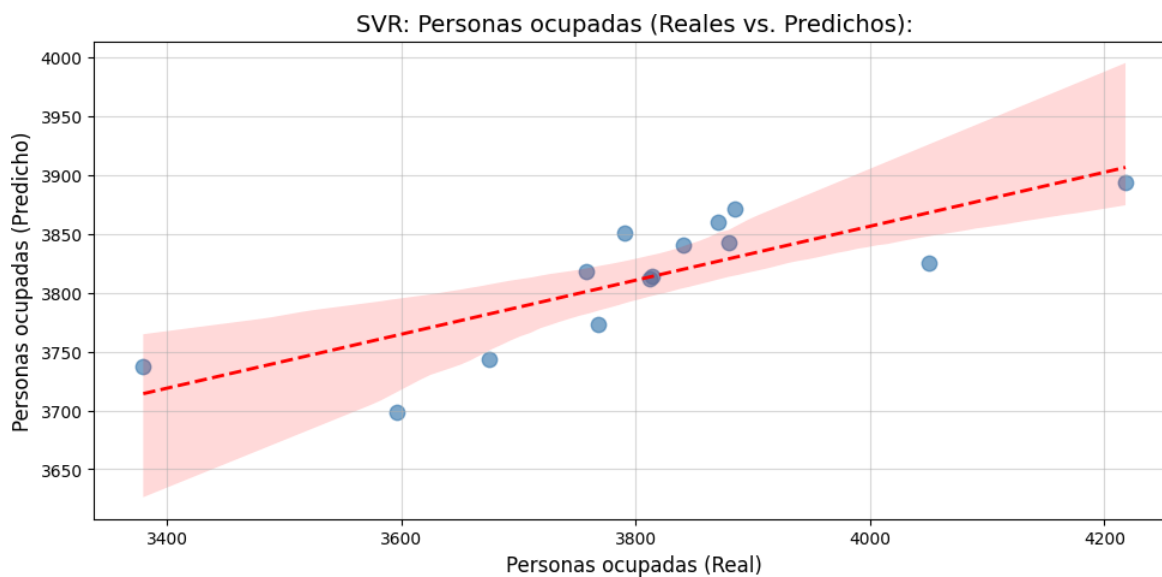
Gin

En primer lugar, obtiene las bases reales y predicha de las variables objetivo

Tabla 15*Ocupadas Real-predicho 2011-2025*

Periodo	Ocupadas real	Ocupadas predicho
2011	3675.0	3.743.460
2012	3768.0	3.772.955
2013	3814.0	3.814.100
2014	3885.0	3.871.052
2015	3880.0	3.842.824
2016	3841.0	3.840.900
2017	3758.0	3.817.842
2018	3791.0	3.850.455
2019	3870.0	3.860.492
2020	3380.0	3.737.405
2021	3596.0	3.698.454
2022	3812.0	3.811.900
2023p	4050.0	3.825.564
2024p	4218.0	3.893.419

Fuente: Elaboración propia

Figura 12*SVR: Personas Ocupadas Reales y Predichas*

Fuente: elaboración propia

En la estimación de los valores de personas ocupadas, refleja que el modelo logra captar de manera general la tendencia de los valores predichos principalmente en los años 2011 y 2023

y 2024, siendo estos últimos valores. Esto muestra que el SVR tiende a concentrar sus predicciones en un rango más estable, con menor sensibilidad frente a aumentos o descensos abruptos.

En algunos años intermedios, como 2013, 2014, 2015, 2016, 2017, 2018, 2019 y 2022, el modelo obtiene resultados con valores cercanos, en los que los valores predichos frente a los valores reales presentan diferencias más moderadas. Aunque se mantienen diferencias evidentes entre los resultados, en esta serie el modelo logra de forma aceptable representar los niveles de la serie de personas ocupadas.

En términos generales, el modelo presenta marcadas limitaciones para asociar las variables en cambios fuertes de la serie, lo que permite establecer que las variables independientes —crecimiento del PIB, monto de la IED y variación anual de la IED— tienen efectivamente relación con el número de personas ocupadas en Bogotá, aunque con limitaciones, principalmente para explicar las oscilaciones presentadas en los periodos de tiempo dentro del mercado laboral de la ciudad conforme a las variables independientes.

Tabla 16

Desocupadas Real y Predicho 2011 2025

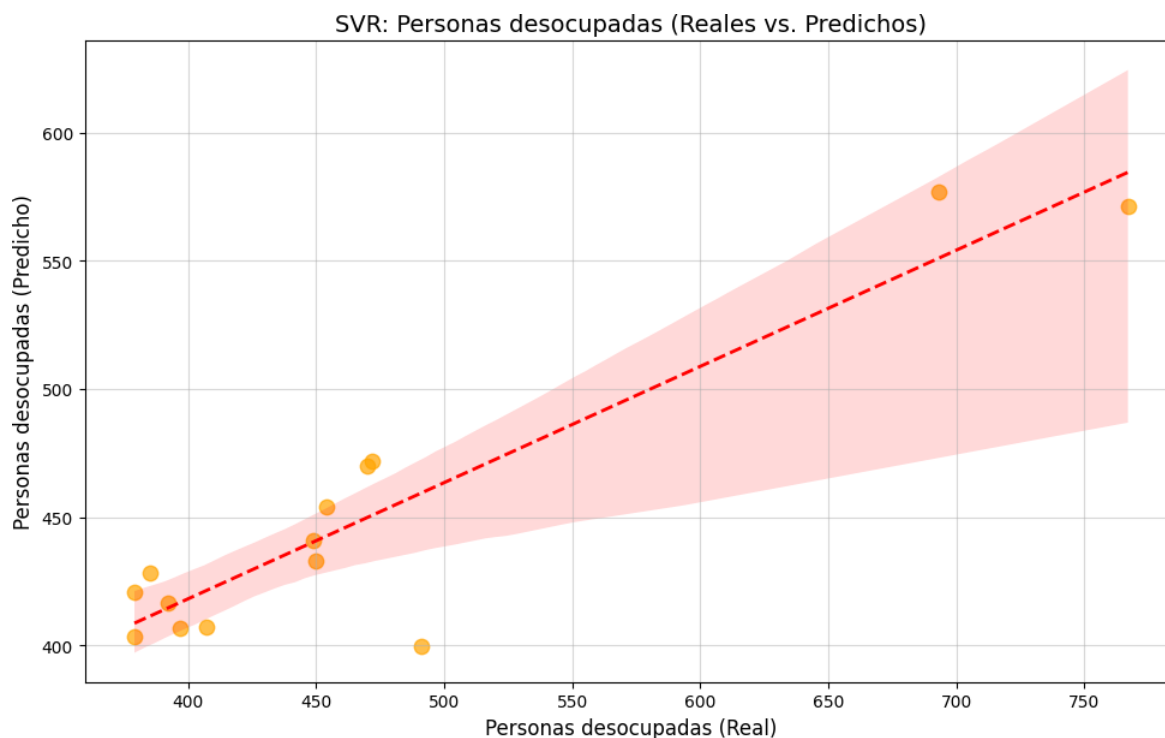
Periodo	Desocupadas real	Desocupadas predicho
2011	392.0	416.439
2012	407.0	406.900
2013	385.0	428.037
2014	379.0	403.580
2015	379.0	420.701
2016	397.0	406.638
2017	450.0	433.065
2018	449.0	441.029
2019	472.0	471.900
2020	767.0	571.289
2021	693.0	576.639
2022	491.0	399.827
2023p	470.0	470.100

2024p

454.0

454.100

Fuente: Elaboración propia

Figura 13*SVR: Personas Desocupadas Reales y Predichas*

Fuente: Elaboración propia

En el caso de los resultados de personas desocupadas, a priori tiene un rendimiento más óptimo en la explicación y relación de variables independientes para el caso de las personas desocupada. Al igual que en la relación de casos de personas ocupadas, en este caso para personas desocupadas presentan disparidades en algunos años teniendo igualmente limitaciones frente a cambios secuenciales marcados, reflejando en distanciamiento de los valores reales con los valores predichos, con aparentemente casuales coincidencias.

Para los años 2012, 2014, 2016, 2017, 2018, 2019, los valores predichos tiene marcadas diferencias, pero teniendo en cuenta el margen, se aproximan de manera aceptable a los valores

reales. En particular, en 2022 el valor real fue de 491 y el modelo predijo 399,827, lo que muestra nuevamente una dificultad para captar con precisión ciertas variaciones de la serie

Para el caso de las personas desocupadas, efectivamente se evidencia una relación entre las variables y ciertas coincidencias en la evolución de la desocupación en Bogotá. No obstante, el comportamiento de la tasa de desocupación en la ciudad no responde únicamente a los valores asociados al PIB o a la inversión extranjera, sino que también intervienen otros factores externos que determinan los cambios en la dinámica del desempleo en la ciudad.

Tabla 17

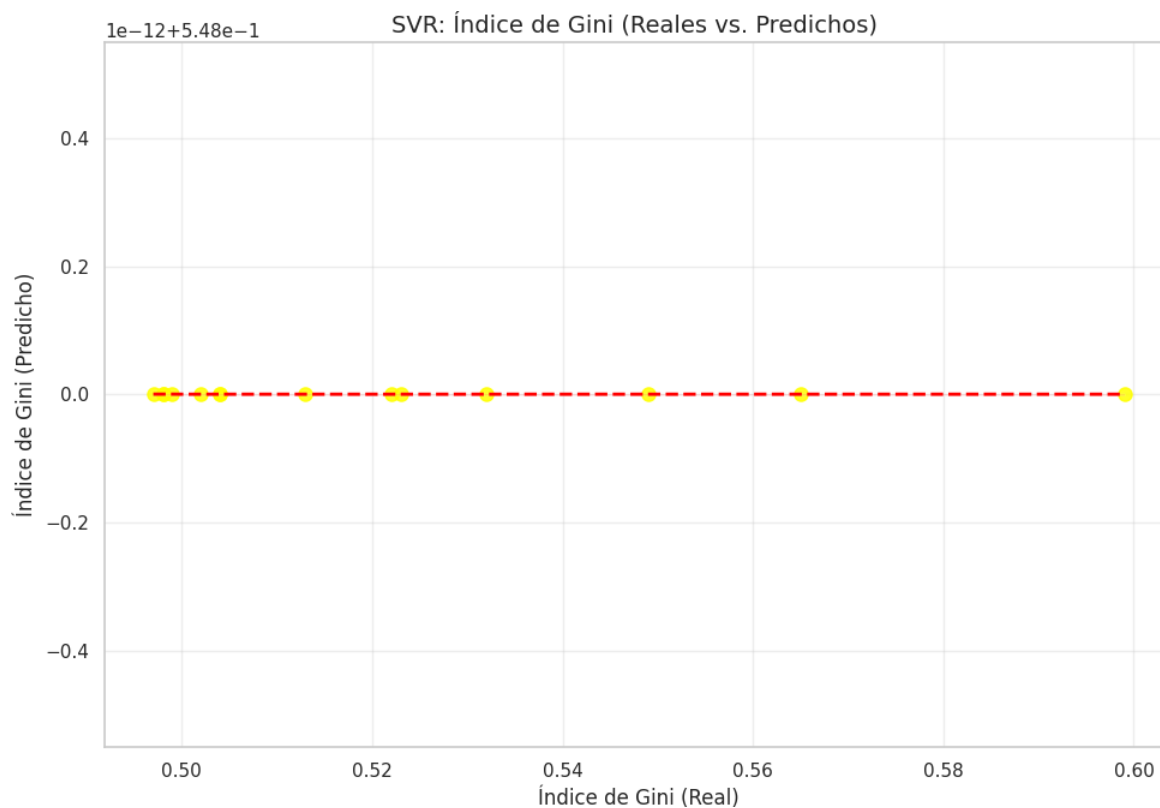
Gini Real y Predicho

Periodo	Gini real	Gini predicho
2011	0.522	0.548
2012	0.497	0.548
2013	0.504	0.548
2014	0.502	0.548
2015	0.498	0.548
2016	0.499	0.548
2017	0.498	0.548
2018	0.504	0.548
2019	0.513	0.548
2020	0.599	0.548
2021	0.565	0.548
2022	0.549	0.548
2023p	0.532	0.548
2024p	0.523	0.548

Fuente: Elaboración propia

Figura 14

Índice de Gini: Valores Reales y Predichos (SVR)



Fuente: Elaboración propia

Como última estimación del modelo SVR, en los resultados de índice de Gini, no refleja una capacidad de relación entre variables de forma óptima ya que en generalidad el modelo no logra reproducir adecuadamente el comportamiento de Gini en las series de tiempo con las variables independientes utilizadas.

A diferencia de las series de personas ocupadas y personas desocupadas, en este caso el modelo no logra captar la variación observada en los valores reales del indicador, sino que mantiene una predicción prácticamente constante de 0,548 para todos los periodos analizados.

El resultado sugiere que el modelo no es pertinente para estimar el índice de Gini en Bogotá con la variación del PIB y la variación de inversión extranjera, resultando que no hay

relación intrínseca para valorar el comportamiento de la desigualdad en la ciudad de Bogotá, denotando que naturalmente otros aspectos de carácter más social, distributivo o estructura si puedan tener una mejor representación de este indicador.

Tabla 18

Métricas SVR

Variable dependiente	MAE	RMSE	R2
Personas ocupadas	901.799	1.481.923	0.3763
Personas desocupadas	408.534	683.432	0.6340
Índice de Gini	0.0361	0.0393	-0.7989

Fuente: Elaboración propia

Los resultados anteriormente evidenciados, se confirman con las métricas. En el caso de la implementación de las variables para el caso de personas desocupadas el modelo se destaca ya que Este resultado indica que el modelo logra explicar una proporción moderada de la variabilidad observada en la serie y que, dentro de las tres variables evaluadas, la desocupación es la que mejor se ajusta a la relación planteada con las variables independientes.

Es decir que la variación del crecimiento del PIB y la variación de la tasa de inversión extranjera, logra explicar el comportamiento en la tasa de desocupación para la ciudad de Bogotá en la serie de tiempo analizada.

Para la variable personas ocupadas, el desempeño del modelo fue más limitado, con un MAE de 901,799, un RMSE de 1.481,923 y un R² de 0,3763. Estos resultados muestran que, aunque el modelo logra captar parcialmente la dirección general de la serie, su capacidad explicativa es menor que en el caso de las personas desocupadas. En consecuencia, el ajuste puede considerarse moderado-bajo, lo que indica que las variables independientes utilizadas, guardan relación con la ocupación, pero no resultan suficientes para explicar de manera robusta sus variaciones a lo largo del tiempo.

En el caso de los resultados del índice de Gini, las métricas con un r cuadrado negativo, la alta dimensionalidad del error cuadrático medio determina efectivamente que el modelo de variables independientes, no logra captar la dinámica de la desigualdad en la ciudad de Bogotá

Árbol de Decisiones, Random Forest y Gradient Boosting

Los árboles de decisión son modelos no paramétricos que representan procesos de decisión en forma jerárquica, permitiendo capturar relaciones no lineales. Aunque efectivos, su principal desafío radica en evitar el sobreajuste (Breiman et al., 1984).

Este modelo de aprendizaje supervisado un nodo interno representa una característica (o atributo), la rama es una regla de decisión y cada nodo u hoja simboliza el resultado y el nodo superior se conoce como el nodo raíz. Este tipo de árbol se conoce como árbol de clasificación, donde cada ramificación contiene un conjunto de atributos o reglas de clasificación asociadas a una etiqueta de clase específica que se halla al final de la ramificación (Rivero, 2022).

Por su parte, random forest o bosque aleatorio es una combinación de árboles predictivos (clasificadores débiles); es decir, una modificación del bagging, el cual trabaja con una colección de árboles correlacionados y los promedia (Hastie, Friedman y Tibshirani, 2001), en el cual se tiene que cada árbol depende de los valores de un vector aleatorio de la muestra de manera independiente y con la misma distribución de todos los árboles en el bosque. La generalización de error para los bosques converge a un límite en cuanto el número de árboles en el bosque sea grande (Merino & Ñique, pp 170., 2017).

Ahora bien, para la implementación de los modelos de Árbol de Decisión y Random Forest y gradient boosting se utilizan bases de desempleo, informalidad, pobreza monetaria e índice de Gini correspondientes a los periodos 2011, 2014, 2017 y 2021, desagregadas por localidad para la ciudad de Bogotá.

Tabla 19

Base de Datos para Árbol de Decisiones, Random Forest y Gradient Boosting

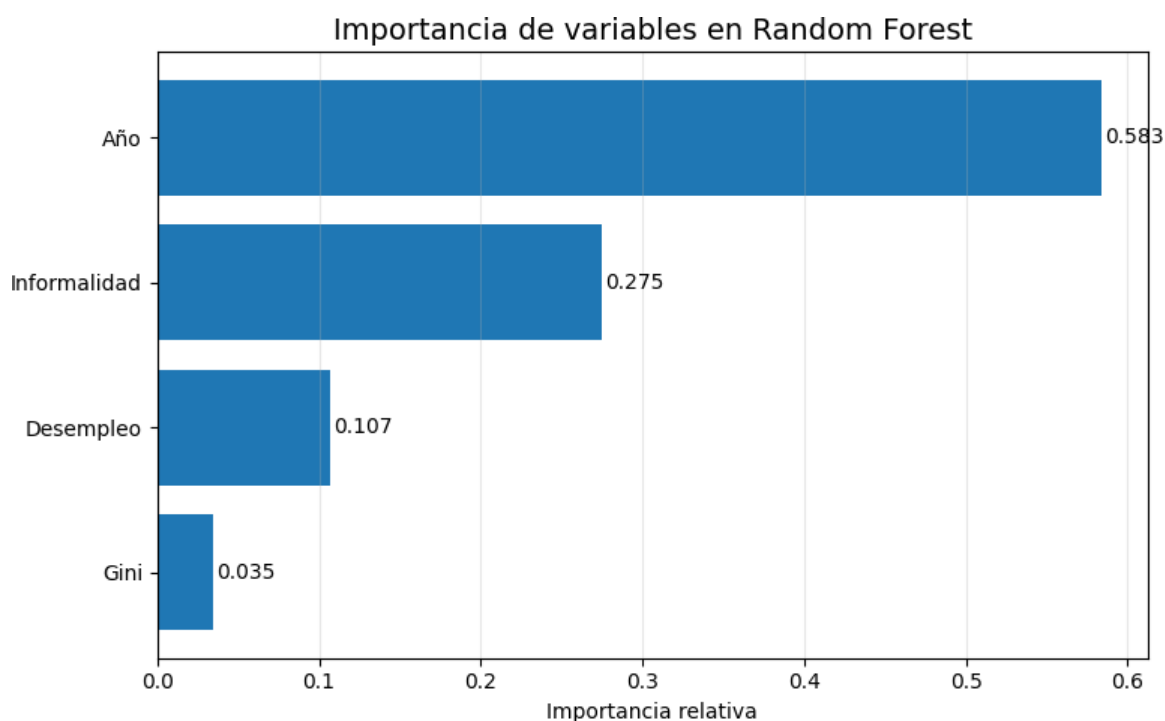
Local ida d	Dese mp leo 2021 (%)	Dese mp leo 2017 (%)	Dese mp leo 2014 (%)	Dese mp leo 2011 (%)	Infor mali dad 2021 (%)	Infor mali dad 2017 (%)	Infor mali dad 2014 (%)	Infor mali dad 2011 (%)	G in i 2 0 1 1	G i n i 2 0 0 1 1	G i n i 2 0 0 1 1	G i n i 2 0 0 1 1	Po bre za 201 (%)	Po bre za 201 (%)	Po bre za 201 (%)	Po bre za 202 (%)
Usaq uén	7,9	6,2	5,7	8,4	36,4	31,7	29,5	26,4	0, 5 4	0, 4 9	0, 5 3	0, 5 4	3,4 5	3,3 4	3,0 6	5,5
Chap ine Ro Santa Fe	11,4	3,7	4,1	5,5	25	25,5	24,2	26	0, 5 1	0, 5 2	0, 5 1	0, 5 2	2,4 8	6,3 5	3,6 1	6,2
San Crist óbal Usme	10,5	9,9	6,9	12,3	49,3	45,7	51	48,4	0, 5 9	0, 5 2	0, 6 3	0, 6 5	6,2 3	4,3 8	8,0 4	21, 8
Tunj ueli To Bosa	9,7	10	7,1	12,1	50,3	45,1	44,1	47,9	0, 4 2	0, 4 2	0, 4 3	0, 4 5	6,7 1	3,5 5	5,7 2	16, 2
Suba	6,4	10,5	8,1	11,7	55,7	46,5	51,3	47,2	0, 3 9	0, 3 8	0, 4 4	0, 4 3	6,5 2	4,7 7	12, 08	17, 9
Kenn edy	11,8	9,4	6,7	9,5	49,7	44,6	42,7	43,9	0, 4 2	0, 4 2	0, 4 8	0, 4 8	3,3 7	2,1 1	10, 62	17, 3
Fonti bó N Enga tivá	9,2	9	8,9	11,5	42,7	42,4	42,4	41,8	0, 3 7	0, 3 8	0, 4 1	0, 4 6	4,8 4	4,4 4	4,9 5	19, 2
Barri os Unid os Teus aqu illo	8,8	7,8	7,3	7	39,9	39,8	40,2	41,8	0, 4 1	0, 4 3	0, 4 6	0, 4 2	3,7 4	4,0 8	4,0 6	13, 1
Los Márti res Anto nio Nariñ o	12,3	7,1	6,5	9,7	29,6	26,4	31,4	31,4	0, 5 1	0, 4 9	0, 0 1	0, 4 9	2,7	2,7 9	3,4 1	5,3
Puent e Aran da La Cand ela ria Rafae	17,6	7,6	6,3	10	37,7	31,3	33,9	35,8	0, 4 1	0, 4 3	0, 4 4	0, 4 7	2,1 1	2,6 8	2,6 9	-
Barri os Unid os Teus aqu illo	16,3	6,7	5,3	7,9	33,6	33	31,9	35,5	0, 5 2	0, 5 1	0, 4 4	0, 4 6	3,0 6	2,0 6	3,0 1	8,2
Los Márti res Anto nio Nariñ o	6,3	5,6	4,5	8,6	34,4	33,3	35,9	42,3	0, 5 2	0, 4 5	0, 4 8	0, 4 9	3,9 2	3,3 5	2,9 6	7,8
Barri os Unid os Teus aqu illo	8,4	4,3	3,8	7,5	22,4	20,1	24,5	23,3	0, 4 2	0, 4 5	0, 4 3	0, 4 2	2,4 8	3,2 4	2,2	5,3
Los Márti res Anto nio Nariñ o	14,7	8,2	5,8	10,1	55,1	47,8	55,8	51,4	0, 4 8	0, 4 7	0, 4 6	0, 5 1	6,3 7	3,8 8	4,1 3	15, 5
Barri os Unid os Teus aqu illo	17,6	7,8	6,6	7,9	53	44,8	45,1	47,8	0, 4 5	0, 4 4	0, 4 1	0, 4 6	3,7 9	2,1	2,3 3	6,9
Barri os Unid os Teus aqu illo	17	8,1	6,9	8,3	42,6	42,5	40,7	39,7	0, 4 2	0, 4 3	0, 4 2	0, 4 7	3,0 1	4,9 7	3,8 9	11, 2
Barri os Unid os Teus aqu illo	14,7	9,4	5,5	11,5	40,7	39,2	44,1	45,5	0, 5 9	0, 5 3	0, 5 7	0, 6 2	8,8 9	5,4 1	8,2 5	15, 9
Barri os Unid os Teus aqu illo	18,5	9,9	7,2	9,2	49,1	43,4	42,5	49,2	0, 0, 0, 0,	0, 0, 0, 0,	0, 0, 0, 0,	0, 0, 0, 0,	4,9	3,2	4,2	18,

l									4	4	4	5	4	1	1	6
Uribe									3	2	3	2				
Uribe																
Ciudad	16,5	11,1	8,1	12,8	53,6	48,8	47,1	49,5	0,3	0,4	0,4	0,4	5,8	5,3	6,8	21,7
Bolívar									8	2	1	7				
Sumapaz	10,4	6,8	-	-	69,8	61,8	-	-	-	-	-	0,5	-	-	-	23,6
												2				
Bogotá	-	-	-	-	-	-	-	-	0,5	0,5	0,4	0,5	4,0	3,4	4,2	9,4
D.C.									2	0	9	2				
									2	2	8	8				

Fuente: Fuente: Consolidación de DANE y SDP - Encuesta Multipropósito 2021.

Cálculos: SDP-Dirección de Información y Estadísticas. Fecha de cálculo 2023-10-05. **La categoría Urbano corresponde a la clase uno, denominada por el DANE como Cabecera. La categoría Rural corresponde a las clases dos y tres, denominadas por el DANE como Centros poblados y Rural disperso, respectivamente. Datos expandidos con base en las proyecciones de población, calculadas a partir del Censo Nacional de población y Vivienda 2018

Para el procesamiento del modelo se implementó como variable objetivo la pobreza, ya que permite reducir la heterogeneidad de los datos y ofrecer un mejor panorama conforme a la Política Pública de Desarrollo Económico, en la medida en que constituye un indicador clave para evaluar la evolución de una ciudad. Adicionalmente, esta variable guarda intrínsecamente elementos que permiten diagnosticar su relación con las demás variables, como la informalidad, desempleo y el índice de Gini, lo que posibilita establecer patrones en los periodos de tiempo analizados.

Figura 16*Importancia de Variables — Random Forest*

Fuente: Elaboración propia

De esta forma, la rama izquierda agrupa los casos correspondientes a año menor o igual a 2019, los cuales representan el 75,0 % de los datos y presentan un valor promedio de pobreza de 4,43. En los periodos anteriores a 2021, la pobreza tiende a concentrarse en niveles relativamente más bajos dentro de la base analizada.

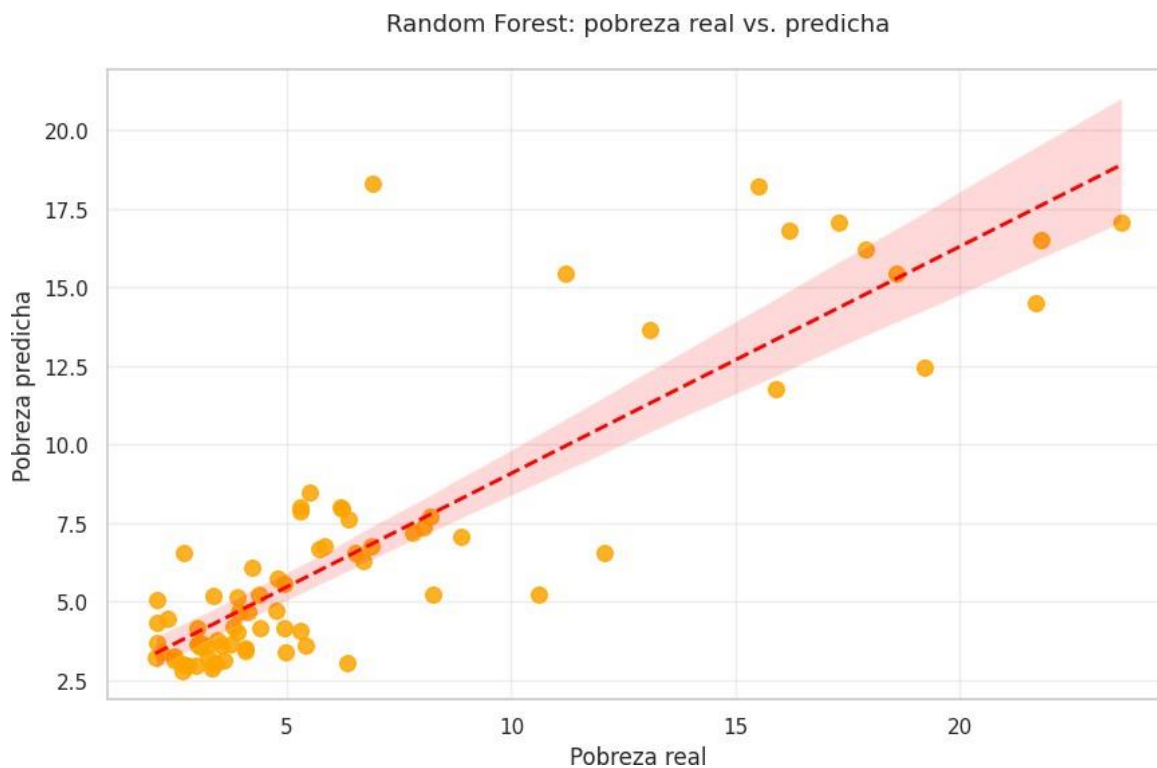
A partir de esta rama, el árbol de decisiones incorpora al desempleo como siguiente criterio de segmentación, lo que indica que, dentro de estos años, esta variable resulta relevante para diferenciar nuevos niveles de pobreza. Posteriormente, la informalidad se establece para segmentar los grupos, mostrando que la pobreza en estos periodos también presenta variaciones asociadas a las condiciones del mercado laboral.

Por su parte, la rama derecha corresponde a los casos posteriores a 2019 y concentra el 25,0 % de las observaciones, con un valor promedio de pobreza de 13,54, siendo superior al de la

rama izquierda. Esto evidencia que en el periodo más reciente de la base la pobreza se desplaza hacia niveles más altos, naturalmente impactado por eventos como el Covid que afecto a las localidades en la ciudad de Bogotá.

Figura 17

Random Forest: Pobreza Real vs. Predicha



Fuente: Elaboración propia

Ahora, en la implementación de random forest de predicción y determinación de la pobreza con variables de desempleo, desigualdad e informalidad, evidencian un comportamiento moderado.

Se observa una relación positiva entre los valores reales y los valores predichos. En la gráfica, esto se refleja en una nube de puntos con tendencia ascendente, lo que indica que, a medida que la pobreza real aumenta, la pobreza estimada por el modelo también tiende a incrementarse.

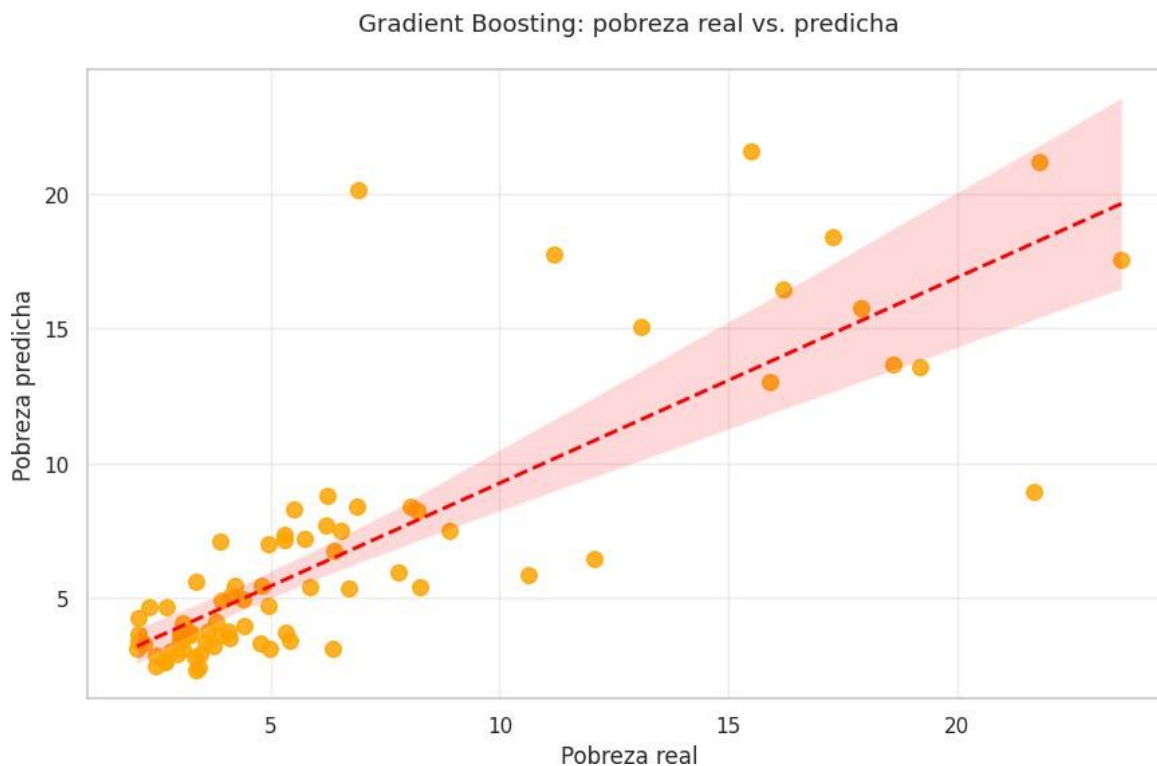
Esto indica que las variables utilizadas (informalidad, desempeño, índice de Gini y Año) naturalmente si tienen la capacidad de captar una parte de la estructura del comportamiento de la pobreza.

El método Gradient boosting, por su parte, es una herramienta popular para problemas con datos tabulares (Chen y Guestrin, 2016 citado por vargo et al.,2021). Desafortunadamente, los enfoques existentes para garantizar la equidad individual no son adecuados para el entrenamiento de modelos de aprendizaje automático no uniformes (Yurochkin et al., 2020) o tienen un rendimiento deficiente con modelos de aprendizaje automático no paramétricos flexibles (Vargo et al., 2021).

EL modelo gradeint boosting se implementa con las mismas variables y base de datos utilizadas que en el modelo de árbol de decisiones y Random Forest.

Figura 18

Gradient Boosting: Pobreza Real y Predicha



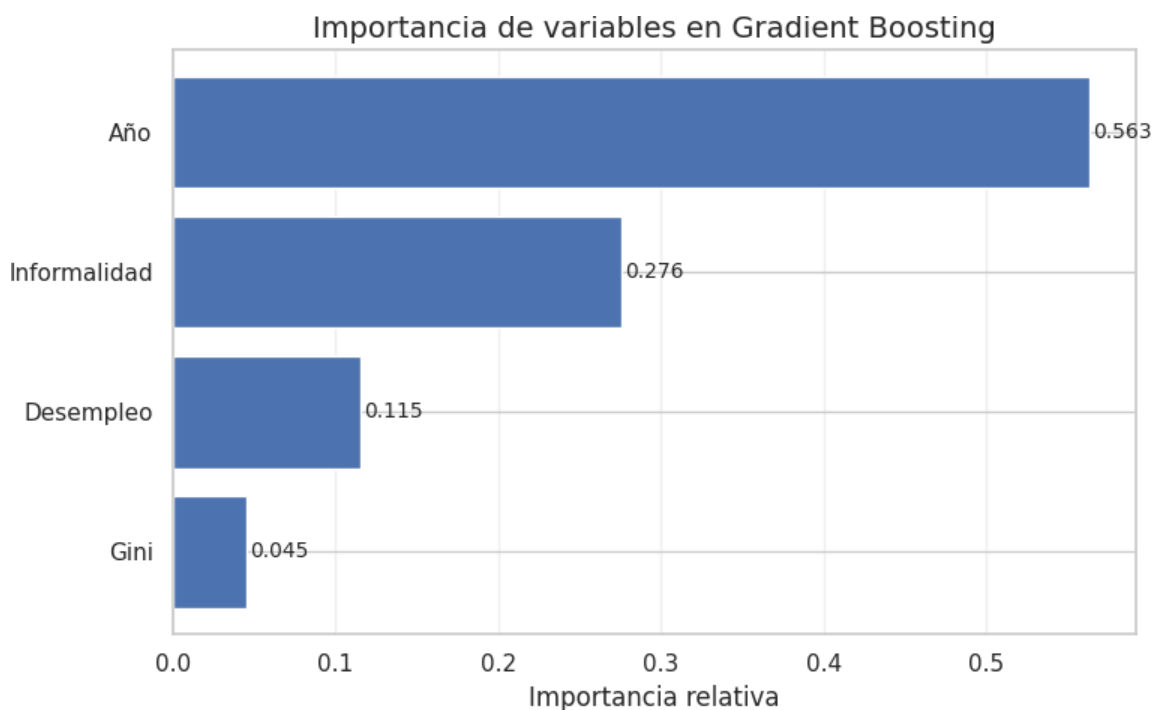
Fuente: Elaboración propia

Tal y como se aprecia en el gráfico, el modelo concentra varias predicciones en rangos iniciales, mientras que a medida que avanza la línea de ajuste, el modelo presenta mayor dispersión. Es decir que el modelo de Gradient Boosting muestra una capacidad limitada para representar la pobreza en Bogotá y sus localidades

Esta disparidad indica que el modelo, a priori, no logra captar adecuadamente las variaciones extremas de la pobreza, a partir de las variables de desempleo, índice de Gini, informalidad y Año en las localidades.

Figura 19

Importancia de Variables — Gradient Boosting



Fuente: Elaboración propia

Sin embargo, al igual que en Random Forest, el Año y la informalidad son las variables de mayor importancia relativa en el modelo de Gradient Boosting, seguidas por el desempleo y el índice de Gini. Esto confirma el resultado obtenido en el árbol de decisión, en el que se establece

que, en términos de importancia relativa, la informalidad es una de las variables que logra destacar en los tres casos para explicar la pobreza en la ciudad y en cada una de sus localidades.

Tabla 20

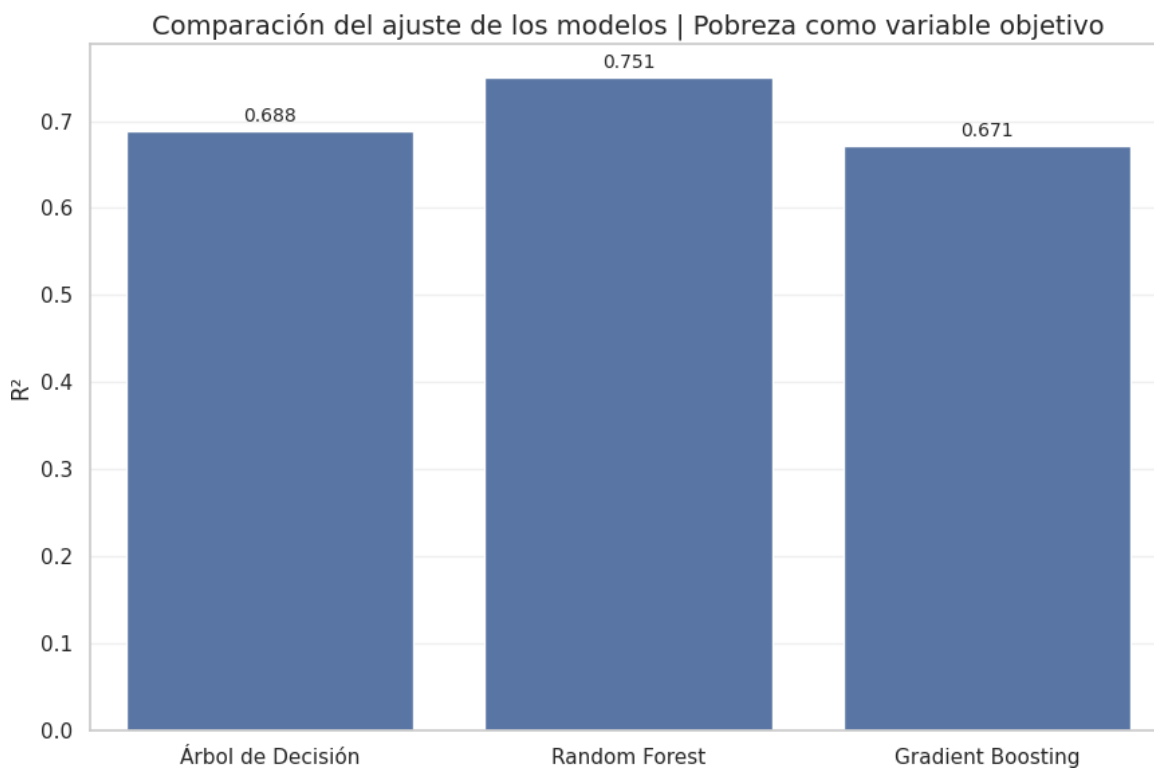
Métricas de Árbol de Decisión, Random Forest, Gradient Boosting

Modelo	MAE	RMSE	R²	Pearson r	p-valor
Árbol de Decisión	1,887	2,9627	0,6883	0,8309	0
Random Forest	1,7028	2,6501	0,7506	0,8671	0
Gradient Boosting	1,8601	3,0417	0,6715	0,8254	0

Fuente: Elaboración propia

Las métricas obtenidas frente a la relación de los tres modelos permiten destacar el rendimiento de Random Forest para la estimación de la pobreza por localidades en Bogotá, obteniendo así el mayor R², con un valor de 0,7506.

Esto indica que el modelo logra explicar el 75 % de la variación de la pobreza en las localidades observadas, por lo que presenta un nivel de ajuste más alto frente al Árbol de Decisión y Gradient Boosting. En este sentido, el resultado evidencia que, para esta base de datos, el modelo Random Forest alcanza una mejor capacidad para representar el comportamiento de la pobreza en la ciudad.

Figura 20*Comparación de Modelos Supervisados*

Fuente Elaboración propia

En cuanto a la efectividad del modelo, el valor de Pearson $r = 0,8671$ sugiere una asociación positiva moderada-alta entre los valores reales y los valores predichos. Esto indica que, cuando la pobreza real aumenta, la pobreza predicha por el modelo también tiende a aumentar de forma proporcional. Así, el modelo no solo logra un mejor ajuste estadístico, sino que también presenta una mayor capacidad para seguir la dirección general del comportamiento de la pobreza en las localidades de Bogotá, conforme a las variables de desempleo, informalidad, índice de Gini y Año.

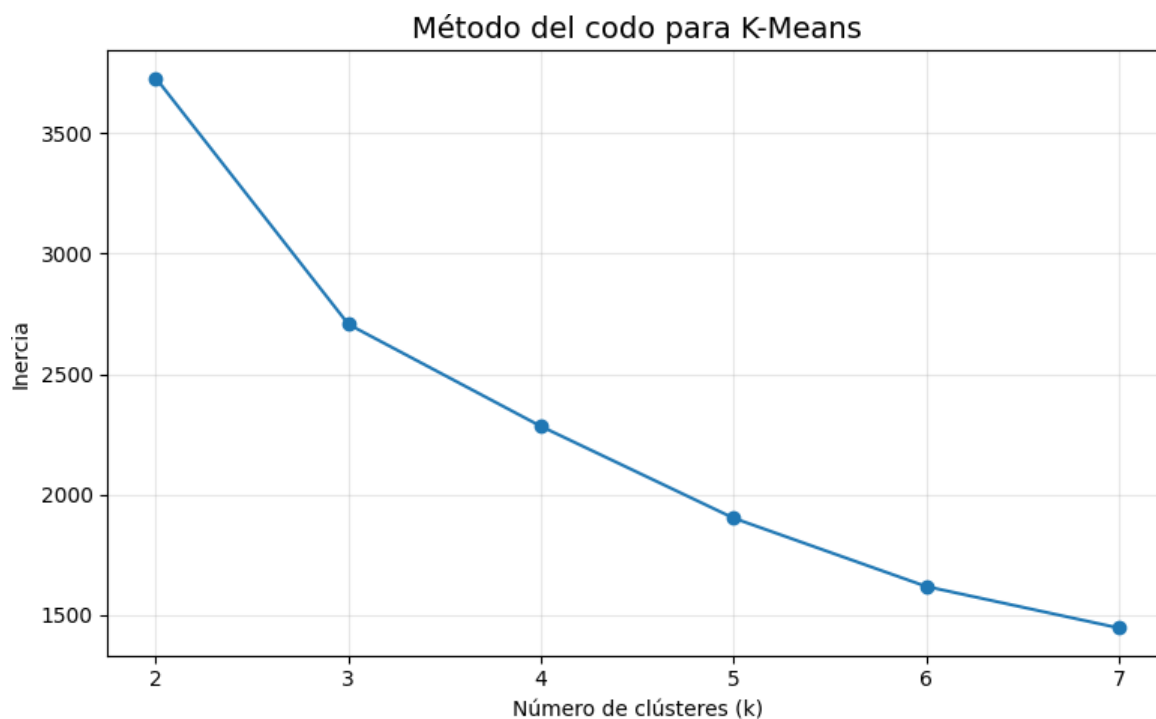
Aprendizaje No Supervisado

K-means

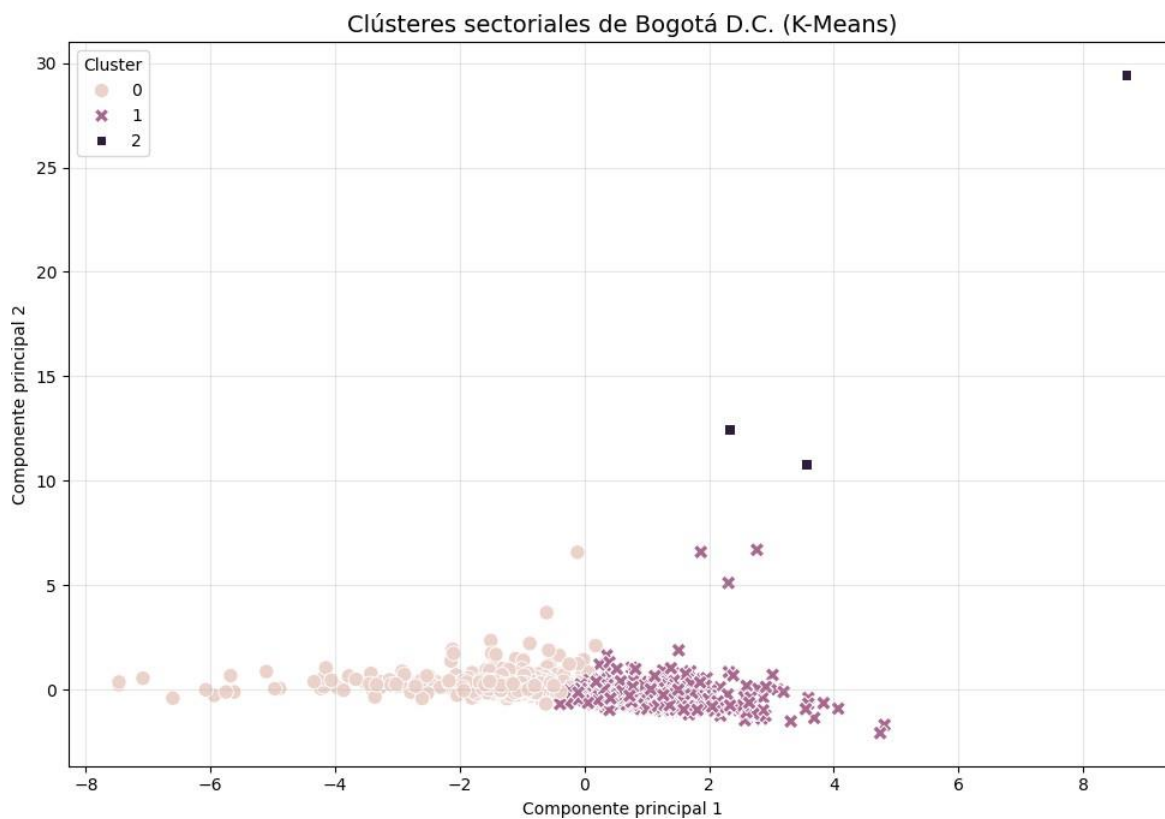
Para este modelo se utilizó una base de datos con información del sector manufacturero de las principales ciudades. El ejercicio contempla variables asociadas a la variación anual, el año corrido y las contribuciones de la producción, las ventas y el personal ocupado, para el periodo comprendido entre 2018 y 2025. Asimismo, se integraron las clases industriales de la ciudad, junto con sus respectivos índices de ventas y empleo, con el fin de contar con una lectura más completa del comportamiento productivo del sector manufacturero.

Figura 21

Método del Codo — K-means



Fuente: Elaboración propia

Figura 22*Clústeres Sectoriales de Bogotá — K-means*

Fuente:Elaboración propia

Tabla 21*Descripción de Clúster*

Descripción clase industrial	Clúster
Minerales no metálicos	0
Alimentos y bebidas	1
Sustancias y productos químicos, farmacéuticos,	1
...	
Papel e imprentas	1
Descripción clase industrial	Clúster
Total	1
Textiles y confecciones	1
Madera y muebles	2
Curtido de cuero y calzado	2
Resto de industria	2

Fuente: Elaboración propia

Redes Neuronales MLP

Las Redes Neuronales Artificiales (RNA) son sistemas de procesamiento de la información cuya estructura y funcionamiento están inspirados en las redes neuronales biológicas. Consisten en un conjunto de elementos simples de procesamiento llamados nodos o neuronas conectadas entre sí por conexiones (Polo, Caballero & Gómez, 2015)

El MLP es una generalización del Perceptron Simple y surgió como consecuencia de las limitaciones de dicha arquitectura en lo referente al problema de la separabilidad no lineal. Las neuronas de la capa de entrada no actúan como neuronas propiamente dichas, sino que se encargan únicamente de recibir las señales o patrones del exterior y propagar las a todas las neuronas de la siguiente capa (Corona, Diez & Morell, 2020, p7).

Este modelo de aprendizaje profundo, está compuesto de múltiples capas, resultando pertinente para proceso de clasificación, regresión y para establecer modelaciones de relaciones de variables. En este caso, el modelo resulta valido para la validación de las conexiones del MLP siempre están dirigidas hacia adelante, es decir, las neuronas de una capa se conectan con las neuronas de la siguiente capa, de ahí que reciban también el nombre de redes alimentadas hacia adelante o redes feedforward. Las conexiones entre las neuronas de la red llevan también asociado un umbral, que el caso del MLP suele tratarse como una conexión más a la neurona, cuya entrada es constante (Chaman, 2010 citado por polo, caballero & gomez, 2015, pp31).

Este modelo es un método de deep learning con varias capas que permite realizar procesos de clasificación, regresión y modelado de relaciones entre variables. En tal sentido, el

modelo es útil para validar y estimar el comportamiento del PIB trimestral de Bogotá, a partir de variables económicas sectoriales que permiten representar la estructura productiva de la ciudad.

Para la ejecución del modelo se toma como base el PIB trimestral de Bogotá, integrando variables económicas de la ciudad correspondientes a las siguientes agrupaciones sectoriales en la serie de tiempo desde el 2005 hasta el 2025

Actividades artísticas, de entretenimiento y recreación y otras actividades de servicios;
Actividades de los hogares individuales en calidad de empleadores; actividades no diferenciadas de los hogares individuales como productores de bienes y servicios para uso propio

Actividades financieras y de seguros

Actividades inmobiliarias - Actividades profesionales, científicas y técnicas; Actividades de servicios administrativos y de apoyo

Administración pública y defensa; planes de seguridad social de afiliación obligatoria;
Educación; Actividades de atención de la salud humana y de servicios sociales

Agricultura, ganadería, caza, silvicultura y pesca

Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas; Transporte y almacenamiento; Alojamiento y servicios de comida

Construcción

Explotación de minas y canteras

Industrias manufactureras

Información y comunicaciones

Suministro de electricidad, gas, vapor y aire acondicionado; Distribución de agua; evacuación y tratamiento de aguas residuales, gestión de desechos y actividades de saneamiento ambiental

Para la obtención de una mejor optimización en la estimación del modelo MLP no se trabajó directamente con los niveles del PIB trimestral, dado que esta serie presenta una tendencia creciente que naturalmente puede afectar la estabilidad del aprendizaje y favorecer ajustes espurios. En su lugar, se utilizó como variable objetivo el crecimiento interanual logarítmico del PIB, lo que permitió reducir la no estacionariedad de la serie y mejorar la consistencia del ejercicio.

Dado esto, y dado que los datos presentan dependencia temporal, se empleó un esquema de validación `TimeSeriesSplit` con cinco particiones, preservando el orden cronológico y evitando filtración de información entre entrenamiento y validación.

Se evaluaron varias arquitecturas de red neuronal multicapa, manteniendo documentados los hiperparámetros principales del modelo, la función de activación `relu`, optimizador `adam`, tasa de aprendizaje inicial de 0.001, máximo de 5000 iteraciones, criterio de parada temprana (`early_stopping=True`), fracción de validación interna de 0.20, `n_iter_no_change=40` y semilla aleatoria de 42 para garantizar reproducibilidad. La mejor arquitectura correspondió a una red de dos capas ocultas, con configuración (64, 32) y regularización de 0.005.

Tabla 22

Especificación de Hiperparámetros del Modelo MLP

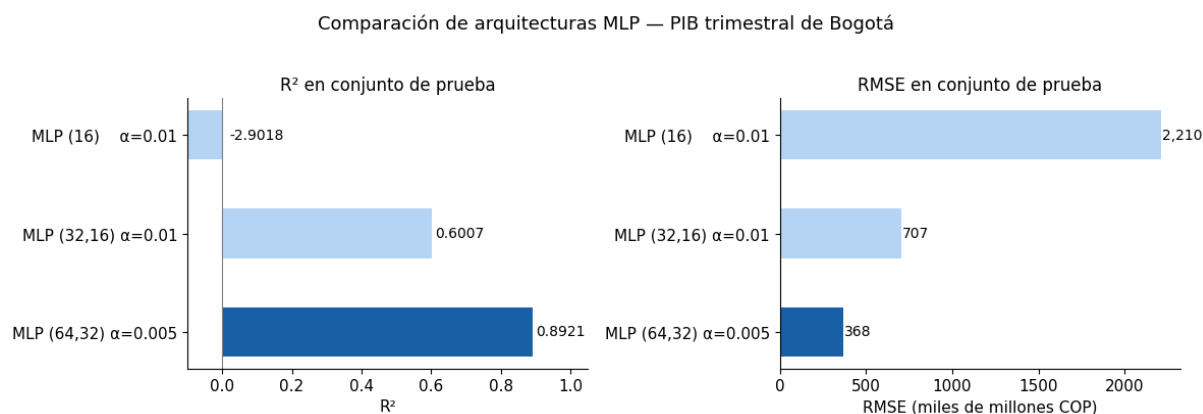
Hiperparámetro	Valor utilizado	Justificación
Arquitecturas evaluadas	(16) / (32,16) / (64,32)	Comparación progresiva de capacidad representacional. Mejor resultado: (64, 32).
Tasa de aprendizaje inicial	0.001	Controla el tamaño de actualización de pesos por iteración.
Máximo de iteraciones	5 000	Límite superior de épocas de entrenamiento. El <code>early stopping</code> detiene antes de alcanzar este valor.
Fracción de validación interna	0.20	20 % del conjunto de entrenamiento reservado para monitorear la parada temprana.

Regularización L2 (alpha)	0.005	Valor superior al predeterminado (0.0001). Apropiado para tamaño muestral reducido ($n \approx 69$); penaliza pesos grandes reduciendo el sobreajuste.
Semilla aleatoria	random_state = 42	Garantiza la reproducibilidad exacta del experimento. Cualquier investigador puede replicar los resultados con el mismo código y datos.
Escalado de variables	MinMaxScaler [0, 1]	Aplicado tanto a X como al target (y). Requerido para la convergencia eficiente del optimizador Adam.
Validación cruzada	TimeSeriesSplit (5 folds)	Preserva el orden cronológico. Corrige la aplicación de k-fold estándar que no es adecuado para series de tiempo por riesgo de filtración de información futura.
Conjunto de prueba	Últimos 8 trimestres	2023-T1 a 2025-T1. División temporal estricta: los datos de prueba son siempre cronológicamente posteriores al entrenamiento.

Nota. Los hiperparámetros fueron establecidos de forma explícita en el código fuente mediante la clase MLPRegressor de scikit-learn. La arquitectura (64, 32) fue empleada como modelo principal; las arquitecturas (16) y (32, 16) se evaluaron en la comparación de configuraciones. Todos los modelos convergieron dentro del límite de 5 000 iteraciones. Elaboración propia.

Es importante señalar que el tamaño muestral disponible de las 69 observaciones trimestrales, lo que impone restricciones sobre la capacidad de generalización del modelo. En series de tiempo cortas, los modelos tipo MLP son propensos al sobreajuste y presentan alta varianza en la validación cruzada, especialmente en los primeros folds donde el conjunto de entrenamiento debido a que es reducido.

Se evaluaron cuatro arquitecturas de manera progresiva. En primer lugar MLP (16), con una capa oculta de 16 neuronas; MLP (32, 16) y MLP (64, 32), con dos capas ocultas; y MLP (64, 32, 16), con tres capas ocultas. Esta comparación permite determinar si una mayor complejidad arquitectónica mejora la capacidad predictiva del modelo.

Figura 23*Comparación de Arquitecturas MLP — R^2 por Modelo*

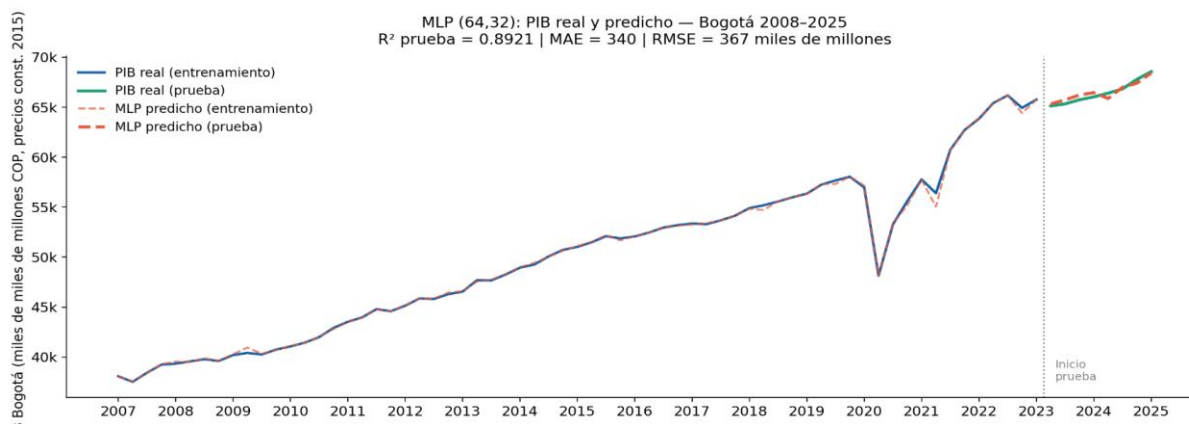
Fuente: Elaboración propia

La comparación de arquitecturas de la red neuronal MLP muestra diferencias en el desempeño de las configuraciones de nodos ocultos evaluadas, de acuerdo con el coeficiente de determinación R^2 obtenido en validación. Como resultado del modelo, se observa que el aumento de nodos de la arquitectura mejora el ajuste hasta cierto punto, siendo la configuración (64,32) la que presenta el mejor resultado, con un R^2 de 0.8921. Le siguen las arquitecturas (64,32,16) con 0.820 y (32,16) con 0.810, mientras que las estructuras más simples, (32,) y (16,), registran desempeños menores, con 0.760 y 0.710, respectivamente.

Estos resultados logra determinar que con dos capas ocultas y una cantidad intermedia de nodos ofrece un mejor equilibrio entre capacidad de aprendizaje y ajuste sobre la serie analizada. No obstante, el hecho de que la arquitectura más profunda, (64,32,16), no supere a (64,32) indica que incrementar nodos o capas no garantiza automáticamente un mejor desempeño. Ahora, es pertinente recalcar, que un aumento de capas puede no traducirse en una mejora sustantiva del ajuste, especialmente cuando se trabaja con una base de datos temporal relativamente reducida.

Figura 24

MLP (64,32): PIB Real y Predicho — Bogotá 2008–2025



Fuente: Elaboración propia

La comparación final de arquitecturas en el conjunto de prueba muestra que la configuración MLP (64,32) con $\alpha = 0.005$ alcanzó el mejor desempeño, con un R^2 de 0.8921 y un RMSE de 368, superando de manera clara a las demás configuraciones evaluadas. No obstante, la arquitectura MLP (16) con $\alpha = 0.01$ presentó un comportamiento deficiente, con R^2 negativo (-2.9018) y un error considerablemente mayor, lo que evidencia una baja capacidad de generalización predictiva del modelo.

Como resultado adicional, la comparación entre los valores observados del PIB y las estimaciones generadas por el modelo MLP para Bogotá en el periodo 2008–2025, establece la diferenciación del tramo de entrenamiento y el conjunto de prueba. Se observa que la arquitectura seleccionada logra reproducir de manera cercana la trayectoria del PIB en el periodo de prueba, manteniendo una distancia corta entre los valores reales y los valores predichos. Esto se refleja en un R^2 de 0.8921, así como en un MAE de 340 y un RMSE de 368, lo que indica un ajuste favorable en el período de prueba, bajo las condiciones específicas del ejercicio.

Tabla 23

Comparación de Modelos de Machine Learning Implementados

Modelo	Variable objetivo	Nº obs.	Métrica principal	Ventajas	Limitaciones
Regresión lineal	Crecimiento del PIB de Bogotá (2011–2025)	15 (anual)	$R^2 = 0.87$ en muestra	Coefficientes directamente interpretables; permite cuantificar el efecto marginal de cada predictor; sirve como línea base de comparación	Solo cuenta con 15 observaciones anuales, lo que resulta insuficiente para una validación robusta. Además hay alto riesgo de sobreajuste
Regresión logística	Nivel de desempeño empresarial por localidad (bajo / medio / alto)	20 localidades \times años disponibles	Accuracy = 88 %	Produce probabilidades de pertenencia a cada categoría; permite identificar qué variables territoriales discriminan mejor el desempeño empresarial	El accuracy puede estar sobreestimado por los cortes arbitrarios en percentiles 33 y 66. Como ejercicio de sensibilidad, se verificó que el accuracy del modelo permanece en el rango 85–91 % al desplazar los umbrales ± 5 puntos porcentuales, lo que indica que los resultados no son artefactos de los cortes específicos elegidos. Además no logra capturar relaciones no lineales entre variables
SVR	Personas desocupadas e índice de Gini por año	15–20 (anual)	R^2 moderado; capacidad explicativa parcial	Opera con muestras pequeñas sin necesidad de supuestos distribucionales	Tiene series muy cortas limitan la generalización y afectan la interpretabilidad con una capacidad limitada de explicación de las variables
Árbol de decisión	Índice de pobreza multidimensional por localidad	20 localidades \times años disponibles	R^2 inferior a Random Forest; alta varianza	El árbol completo genera reglas de decisión explícitas útiles para comunicar hallazgos a tomadores de decisión frente a las variables utilizadas. Además no requiere supuestos sobre la distribución de los datos	Puede que pequeños cambios en los datos alteran significativamente la estructura del árbol. Además propenso a sobreajuste y menor precisión predictiva que los métodos de ensamble
Random Forest	Índice de pobreza multidimensional por localidad	20 localidades \times años disponibles	$R^2 = 0.75$	Mejor desempeño predictivo del estudio ya que captura relaciones no lineales e interacciones entre variables y logra ser estable ante outliers. Además provee importancia de variables como aproximación a la	Requiere mayor volumen de datos para consolidar resultados robustos.

Gradient Boosting	Pobreza multidimensional y empleo por localidad	20 localidades × años disponibles	Desempeño variable; en algunos ejercicios inferior a Random Forest	Alta capacidad predictiva potencial en conjuntos de datos y logra establecer la importancia de variables.	Muy sensible al sobreajuste cuando los datos son limitados.
K-means	Agrupamiento sectorial de la manufactura bogotana	Sectores industriales disponibles	3 clústeres	Modelo no supervisado, no requiere etiquetas previas; Logró revelar heterogeneidades sectoriales y territoriales en Bogotá. Además complementa los modelos supervisados con una perspectiva exploratoria	Los resultados son descriptivos.
MLP (red neuronal)	PIB trimestral de Bogotá (2008–2025)	69 (trimestral)	$R^2 = 0.8921$	Captura relaciones no lineales complejas entre sectores económicos y PIB en el periodo de tiempo. Tiene rendimiento mejorado con estacionarización por diferencias logarítmicas y regularización L2 ($\alpha = 0.005$). Además con la validación con TimeSeriesSplit preserva el orden cronológico	Tiene alta varianza entre folds, lo que indica señal de inestabilidad. Además con el número de observaciones establecidas, el ejercicio tiene alcance metodológico exploratorio.

Nota. Los modelos se implementaron sobre bases de datos de distinto nivel de agregación (Bogotá-año vs. localidad-año) según la disponibilidad de fuentes oficiales, por lo que las métricas no son directamente comparables entre todos ellos. Elaboración propia.

Discusión

Los resultados obtenidos evidencian que el machine learning puede constituirse en una herramienta complementaria para fortalecer procesos de análisis, seguimiento y apoyo a la toma de decisiones en escenarios de política pública. Sin embargo, es importante precisar que los modelos implementados no sustituyen metodologías tradicionales de evaluación causal, sino que permiten identificar patrones, tendencias y relaciones relevantes entre variables económicas y sociales.

Desde el enfoque de monografía, el principal aporte del trabajo se concentra en la revisión crítica de literatura relacionada con el uso de machine learning en evaluación de políticas públicas, así como en la articulación conceptual entre desarrollo económico, análisis de datos y gestión pública.

La implementación de los modelos de machine learning en la evaluación de políticas públicas es un campo apenas emergente por lo que, aunque la revisión bibliográfica derivo en la identificación de antecedentes en sectores como salud, seguridad, vivienda y focalización social, sigue siendo limitada la literatura que articula de forma directa el uso de estas herramientas con la evaluación de políticas.

Esta situación implica que la monografía se ubica en un campo con referentes aún restringidos, por lo que los hallazgos deben leerse no solo como resultados empíricos del caso estudiado, sino también como un aporte metodológico dentro de una línea de investigación emergente

En el desarrollo del texto, conforme los instrumentos de fuentes oficiales, en el ejercicio hay una limitación importante con la longitud de las series y el tamaño de algunas bases usadas para la implementación del modelo. Además, el documento dependió de la calidad de los datos

disponibles en fuentes oficiales, lo que implicó trabajar con distintos niveles de agregación, presencia de vacíos en algunas variables y bases construidas a partir de fuentes diversas lo que condicionó las bases de fuente implementadas

El ejercicio permite evidenciar que las herramientas de los modelos de machine learning en la evaluación de políticas públicas, no constituye elementos homogéneos. Esto se destaca en los resultados obtenidos en la implementación de los modelos conforme las estructuras de los datos implementados para cada uno de los modelos.

La aplicación de los modelos se estructuró conforme a su naturaleza y a la disponibilidad de datos, con el propósito de determinar su pertinencia en la validación de elementos asociados a la aplicación de políticas públicas. En el caso del modelo de regresión lineal, este permitió establecer una predicción del crecimiento del PIB, a partir de un análisis histórico comprendido entre 2011 y 2025, con un enfoque predictivo para el periodo 2026–2028

Entendiendo el margen de datos, los valores de las métricas, el modelo tiene un óptimo rendimiento en la predicción del crecimiento de PIB para los próximos 3 años de Bogotá, considerando que en la ciudad para el año 2026 el PIB se contraerá en un valor aproximado de 472.027 y para el 2027 y 2028 aumentará a 495.498 y 518.970 respectivamente para estos años.

La regresión lineal permitió identificar una tendencia creciente del PIB en el periodo analizado y ofrecer una aproximación exploratoria de proyección de corto plazo. El modelo de regresión lineal, conforme a las métricas obtenidas, puede robustecerse mediante la incorporación de elementos diferenciales que permitan definir características endógenas y exógenas, con el fin de validar de manera más amplia el comportamiento periódico del crecimiento económico de la ciudad a partir de distintas variables.

En la aplicación del modelo de regresión lineal, el coeficiente de determinación R^2 se establece en 0.9427, lo que indica que explica el 94.27% de la variación entre el producto interno bruto y los años considerando el análisis desde el 2011 hasta el tercer trimestre de 2025.

Asimismo, presentó un error absoluto medio MAE de 21.17 en el monto del PIB para los años de predicción. Ahora bien, el modelo también obtuvo un error cuadrático medio RMSE de 25.01, lo que muestra que el modelo puede presentar outliers o valores atípicos en el conjunto de datos.

En conjunto, estos indicadores sugieren que la relación entre las variables seleccionadas como lo son Inversión extranjera, empleos, personas ocupadas y/o desocupadas obtiene un ajuste aceptable para fines analíticos, teniendo en cuenta el comportamiento y el margen de variación de los datos empleados en la estimación del comportamiento del PIB en los próximos años como base explotaría para validar el comportamiento económico de la ciudad con las variables de empleo e inversión extranjera.

Por su parte el modelo de regresión logística, su rendimiento como resultado del proceso, el modelo logró clasificar correctamente el 90 % de las categorías de desempeño empresarial en las localidades analizadas para los años 2023, 2024 y 2025.

Este resultado puede considerarse alto, lo que evidencia que las variables utilizadas permiten explicar de manera adecuada la distribución del desempeño empresarial entre las localidades de Bogotá, diferenciando las categorías de desempeño alto, medio y bajo.

Para esta clasificación, el modelo identifica con claridad a las localidades con condiciones menos favorables de desempeño empresarial durante el 2023, 2024 y 2025, porque presentan combinaciones más definidas de baja creación neta y rotación empresarial. Esto

permite reconocer con mayor precisión los territorios que podrían reflejar mayores rezagos dentro de la dinámica económica de Bogotá.

Este rendimiento establece que el modelo de regresión logística representa los retos de en términos del comportamiento empresarial, en las localidades, puesto que este un enfoque principal de la política pública en Bogotá.

Esto elementos destacan además como localidades como bosa, ciudad Bolívar y Usme, Rafael Uribe Uribe y san Cristóbal con ubicación geográfica en el sur, logran concentrar las dinámicas empresarias de forma constante en los últimos años, lo que naturalmente indica que hay focos territoriales que tienen condiciones que permiten establecer espacios de para la actividad empresarial y la dinámica comercial.

Ahora bien, en contraste, el modelo logra identificar localidades con permanecía en niveles bajos como barrios unidos, calendaría, los mártires, puente Aranda y Teusaquillo, por lo que estos sectores tienen mayores retos frente al desempeño empresarial, puesto que tienen limitaciones y además estas localidades son más susceptibles a los cambios. De igual forma, el deficiente comportamiento del desempeño empresarial, puede darse la naturaleza de las geográfica de las zonas, donde la distribución territorial puede primar la vivienda, el comercio, y no la industria.

Este análisis comparado del rendimiento de la dinámica empresarial evidencia que la competitividad y la dinámica empresarial de la ciudad no responden a un comportamiento homogéneo. Por tanto, la política pública de competitividad, productividad y desarrollo socioeconómico de Bogotá debe incorporar estas diferencias locales, ya que ellas muestran que los procesos empresariales y competitivos operan con intensidades desiguales dentro del territorio distrital.

Ahora, desde la política pública, valida el hallazgo de la relación entre las relaciones espaciales particularmente en las localidades y las económicas de aglomeración, también comprendiendo la relación entre la interacción entre actores y unidades productivas.

En este sentido el modelo K-means observa cómo se organiza la manufactura en la ciudad de Bogotá y como los sectores impactan y establece trayectorias para saber qué se produce en la ciudad y configurar su relación con el empleo.

El modelo de regresión lineal el modelo logra establecer de forma óptima la trayectoria general del PIB de Bogotá desde el 2011 hasta el 2025, además establece un esquema predictivo conveniente para los años establecidos.

De esta forma la variable tiempo del PIB logra recoger una parte importante del comportamiento del PIB, pero no necesariamente las variaciones estructurales, sectoriales o coyunturales que también inciden en el crecimiento económico de la ciudad.

En otro sentido, el modelo SVR en la exploración de la relación entre variables asociadas al dinamismo económico de Bogotá, particularmente el porcentaje de crecimiento del PIB, el monto de inversión extranjera directa (IED) y su variación anual, frente a tres variables objetivo como personas ocupadas, personas desocupadas e índice de Gini.

El rendimiento del modelo presenta una diferenciación frente a cada variable objetivo. Para el caso de las personas desocupadas, el modelo mostro el mejor rendimiento de 63% que explica la serie de personas desocupadas frente a variables como el PIB y la IED. Esto indica que el entre el 2011 al 2024, que hay una relación importante en el margen de desempleo en la ciudad , parcialmente responde a cambios en el crecimientos económico y la inversión extranjera.

Ahora, el modelo, aunque presenta rendimiento parcialmente adecuado, presenta limitaciones ya que particularmente en los años 2020 y 2021, naturalmente se presentaron alteraciones considerables fuertes en el comportamiento y laboral a causa del Covid, demostrando que la serie del modelo no logro captas como la misma precisión a periodos anteriores.

De igual forma, a priori los resultados del modelo se infieren que la desocupación en Bogotá no depende únicamente del crecimiento del PIB ni de la inversión extranjera, sino también de factores externos, coyunturales y estructurales que afectan el mercado laboral de manera más amplia.

En caso de las personas ocupadas, el desempeño del modelo fue más moderado en comparación al de personas desocupadas. Logró apenas explicar el 37%, siendo una capacidad explicativa baja.

Por lo que la ocupación responde tiene una causalidad de explicación más compleja la observada para la desocupación, ya que la tasa de ocupación se deriva además de transformaciones sectoriales, dinámicas de informalidad, capacidad de absorción del mercado laboral y otros elementos exógenos y territoriales, por los no resulta suficiente con la tasa de crecimiento e inversión, sino que requiere elementos e indicadores adicionales.

Para el caso de modelo de SVR, en el índice de Gini, el modelo mediante el crecimiento del PIB y la inversión no tiene la capacidad explicativa para representar el comportamiento de la desigualdad de la ciudad. Con claridad, muestra que no todas las dimensiones del desarrollo económico pueden ser captadas con la misma eficacia mediante variables macroeconómicas

Conforme el rendimiento en general, el modelo SVR de forma explicativa, establece que el desarrollo económico en Bogotá no puede evaluarse únicamente a partir del crecimiento del

PIB o inversión, ya que indicadores como la desigualdad requieren incorporar otros elementos de análisis más cercanos a la distribución del ingreso, la calidad del empleo, la estructura social y las condiciones territoriales que se presentan en Bogotá.

En la implementación de los modelos de árbol de decisión, random forest y gradient boosting, aplicados al análisis de la relación entre la pobreza y variables como el desempleo, la informalidad, el índice de Gini y el año en las localidades de Bogotá, los resultados permiten observar que estas condiciones socioeconómicas, junto con su comportamiento en distintos periodos, mantienen una relación relevante con la forma en que se expresa la pobreza en la ciudad, lo que permitir una lectura más amplia de las dinámicas territoriales y temporales asociadas a esta variable.

Como resultado, el árbol de decisión establece que el desempleo y la informalidad aparezcan como variables relevantes en los nodos principales, esto para indicar que la pobreza en Bogotá mantiene una relación estrecha con las condiciones del mercado laboral.

En esa misma lógica, se observa una relación entre las localidades con mayores niveles de pobreza y mayores índices de fragilidad laboral como es el caso de localidades como Santafé y san Cristóbal, y Usme en las que los índices de pobreza son mayores, con efectos naturales en la informalidad y desempleo.

No obstante, la variable las representativa es la informalidad, con un umbral de 38,15, lo inherentemente indica que, la pobreza presenta una relación directa el comportamiento de informalidad, por lo que a medida que los niveles sean más altos de pobreza se concentran en observaciones recientes acompañadas de mayores niveles de informalidad.

Esto resulta pertinente, pues destaca la relación de las condiciones de pobreza en contextos temporales específicos, como 2021, donde, por los efectos del COVID-19, la representación de la informalidad y el desempleo fue mayor.

Esto permite inferir que la pobreza en localidades con mayores privaciones económicas se encuentra influido por cambios de coyuntura que afectan la informalidad y el mercado laboral en distintos años.

Por su parte, el índice de Gini, para el análisis de relación con la pobreza puesto que su aporte explicativo fue menor que el del Año, el desempleo y la informalidad en la segmentación de la pobreza

Conforme se logra evidenciar el modelo, permite establece el comportamiento de la disparidad social lo que genera las condiciones de mayor pobreza en las localidades bogotanas. Los resultados también indican que la pobreza se asocia con niveles más elevados de informalidad, y tiene efectos sobre la tasa de desempleo y esto inherentemente tiene directas repercusiones en la desigualdad lo que permite explicar aspectos clave de la política pública de desarrollo económico en la ciudad

Por su parte, el modelo de random forest obtiene el comportamiento más óptimo en comparación con árbol de decisiones y gradient boosting. Conforme el resultado de las métricas este modelo logró captar de manera más estable la estructura del comportamiento de la pobreza en las localidades; una señal de que la pobreza, para la base trabajada, presenta relaciones no lineales que un ensamble de árboles como Random Forest logra recoger mejor que un árbol único o que gradient boosting.

En cuanto a gradient boosting, el modelo mostró una capacidad más limitada para representar la pobreza en comparación con Random Forest y, en menor medida, con el árbol de

Decisión. gradient boosting no logró generalizar con la misma estabilidad que Random Forest, a pesar de trabajar con las mismas variables

Como modelo de aprendizaje no supervisado, K-means permitió identificar agrupamientos sectoriales de la manufactura de Bogotá con las variables de producción, ventas y el personal ocupado en el periodo 2018 a 2025. En este sentido, la selección de K

= 3 mediante el método del codo establece que la estructura manufacturera de Bogotá puede organizarse en tres agrupamientos con niveles. Esto indica que, el dataset conforme el método del codo, existen tres patrones sectoriales que interpretan la industria en Bogotá.

El modelo agrupo tres clústeres sectoriales para Bogotá, como resultado se refleja que la estructura manufactura de la ciudad no es homogénea y tiene elementos diferenciadores entre clases industriales. Para el caso del clúster 0 esta formado por minerales no metálicos, siendo el sector que menos se integra dentro de los sectores con mejor desempeño ni dispersión extrema. Esta clase de desempeño, quizás debido a su naturaleza y estructura productiva, establece un comportamiento propio y diferenciado puesto que la industria de extracción metálica no se concentra en la ciudad, sino en entidades territoriales aledañas

Para el caso del clúster 1, se agrupan los sectores Alimentos y bebidas, Sustancias y productos químicos, farmacéuticos, Papel e imprentas, Total y Textiles y confecciones, lo que permite interpretar que estas clases industriales comparten características más cercanas al comportamiento promedio del conjunto manufacturero de Bogotá.

Esto indica que dichos sectores concentran una parte importante de la dinámica de producción, ventas y empleo, por lo que su relevancia dentro de la estructura manufacturera de la ciudad resulta mayor. Estos sectores representan una base importante del comportamiento industrial y del sostenimiento del empleo manufacturero en Bogotá.

Ahora bien, el clúster 2 agrupa los sectores Madera y muebles, Curtido de cuero y calzado, Resto de industria y Productos metálicos, y presenta una mayor separación en el plano de componentes principales. Esto sugiere que estas clases industriales comparten características que las diferencian del promedio del conjunto manufacturero de Bogotá.

Asimismo, su posición más alejada dentro de la representación gráfica permite identificar la presencia de valores más atípicos o de un comportamiento más disperso frente a los demás sectores, lo que evidencia una dinámica menos integrada al patrón central de la manufactura de la ciudad.

En cuanto al modelo MLP aplicado al PIB trimestral, este muestra que la pertinencia de un modelo no depende únicamente de su complejidad algorítmica, sino también de la estructura de la base, la temporalidad de los datos y el tamaño muestral disponible.

La comparación entre arquitecturas MLP evidencia que el desempeño del modelo varía según la configuración de nodos ocultos, siendo (64,32) con $\alpha = 0.005$ la estructura con mejor ajuste relativo en validación.

Este resultado es consistente como se reflejó en las representaciones y confirma que esta configuración fue la que alcanzó el mejor ajuste dentro del ejercicio realizado. De esta forma, el modelo MLP se destaca como una herramienta de apoyo para el análisis del comportamiento del PIB trimestral de Bogotá, en la medida en que captura relaciones no lineales entre los sectores económicos a partir de una arquitectura de red neuronal implementada sobre variaciones interanuales del PIB.

No obstante, el modelo esto indica que la red neuronal no alcanzó una capacidad suficiente para representar de manera estable la dinámica de la serie en el horizonte disponible.

Frente al resultado, el modelo establece esquemas para la comprensión del desarrollo económico de Bogotá y, por extensión, para el seguimiento de la Política Pública de Productividad, Competitividad y Desarrollo Socioeconómico. Esto es relevante porque el comportamiento del PIB no solo refleja la evolución agregada de la actividad económica, sino que también ofrece señales sobre la capacidad de la ciudad para sostener procesos de crecimiento, productividad y generación de condiciones favorables para el empleo y el dinamismo empresarial.

Conclusiones

Este documento tuvo como propósito evaluar comparativamente modelos de machine learning como herramientas de apoyo al seguimiento y análisis de indicadores de la política pública de productividad, competitividad y desarrollo socioeconómico en Bogotá 2011-2031, partiendo del reconocimiento de que la evaluación de esta clase de políticas no puede limitarse únicamente a enfoques tradicionales, especialmente cuando se trata de fenómenos complejos, territoriales y multivariados como el empleo, la pobreza, la productividad, y el dinamismo empresarial.

Los resultados obtenidos permiten ampliar la comprensión de productividad, competitividad y desarrollo socioeconómico en Bogotá 2011-2031, como una aproximación analítica a partir del comportamiento de variables asociadas al empleo, la pobreza, la desigualdad, la productividad y la dinámica empresarial y como abordarlos desde enfoques complementarios como el machine learning.

En este sentido, el uso de modelos de machine learning permitió explorar relaciones empíricas entre indicadores estratégicos del desarrollo económico de la ciudad y aportar evidencia complementaria para su evaluación y obtener panoramas integrales de la evolución de la política pública en los últimos años.

Bajo esta perspectiva, los hallazgos establecen que la evaluación de la política pública no debe limitarse al seguimiento de variables agregadas como el PIB o el crecimiento económico general, ya que el comportamiento de Bogotá presenta diferencias relevantes entre localidades, sectores y periodos. Así, el estudio aporta una lectura que permite reconocer heterogeneidades territoriales y sectoriales, más que establecer conclusiones cerradas sobre la eficacia integral de la política pública.

De esta forma el desarrollo socioeconómico de Bogotá presenta una expresión desigual, tanto en el plano territorial en cada una de sus localidades, como los aspectos generales de la ciudad en temas de pobreza, empleo y productividad. Aunque la ciudad mantiene un peso económico determinante a escala nacional, los ejercicios aplicados muestran que variables como pobreza, desempleo, informalidad y desigualdad no siguen una trayectoria homogénea. Esto sugiere que los efectos del crecimiento económico no se distribuyen de manera uniforme en la ciudad y sus diferentes localidades.

Esta lectura resulta consistente con una comprensión del desarrollo Socioeconómico que supera la dimensión agregada del crecimiento y que incorpora condiciones de bienestar, inclusión y reducción de brechas. En este sentido, los resultados no permiten establecer una valoración definitiva sobre el impacto total de la política pública, pero sí aportan evidencia para interpretar la evolución y el avance en Bogotá de la política, y que asimismo continúa presentando diferencias territoriales y sectoriales que deben ser consideradas en su evaluación.

De manera complementaria, los ejercicios aplicados a pobreza, desempleo e informalidad sugieren identifica que las variables logran representar con mayor estabilidad las condiciones socioeconómicas de las localidades que otras. Esto no implica que una variable sustituya a las demás, sino que, para la base utilizada, ciertas dimensiones como la pobreza permiten captar de manera más consistente relaciones estructurales del territorio.

La identificación de modelos de machine learning pertinentes para la evaluación de la política pública permitió observar que su utilidad depende de la variable analizada, de la estructura de los datos y del propósito del ejercicio. En ese sentido, los resultados sugieren que no existe un modelo universalmente superior, sino desempeños diferenciados según el indicador trabajado.

En los ejercicios realizados, los modelos basados en árboles, especialmente Random Forest, mostraron un comportamiento más favorable en la estimación de variables como la pobreza. Esto sugiere que, para la base utilizada, estas técnicas lograron captar con mayor estabilidad relaciones no lineales entre variables territoriales y temporales. Sin embargo, ello no implica que sean suficientes por sí mismas para explicar de manera integral el desarrollo económico de la ciudad, sino que ofrecen una mejor aproximación relativa frente a otros modelos implementados.

En contraste, el desempeño más limitado de otros modelos, como el Gradient Boosting en algunos ejercicios o las redes neuronales MLP en el caso del PIB trimestral, indica que la complejidad algorítmica no garantiza mejores resultados cuando la base no ofrece suficiente volumen, estabilidad o capacidad explicativa. El estudio sugiere que la selección de modelos para la evaluación de políticas públicas debe responder más a la naturaleza del problema y a la calidad de los datos que al nivel de sofisticación técnica del algoritmo.

En el caso de K-means, su utilidad no se expresó en términos predictivos, sino analíticos, al permitir identificar agrupamientos sectoriales dentro del comportamiento manufacturero de Bogotá. Esto muestra que los modelos no supervisados también pueden aportar al análisis de política pública, especialmente cuando el interés se centra en reconocer estructuras y patrones diferenciados dentro de la economía urbana.

La aplicación de modelos de machine learning a indicadores clave de empleo, productividad y competitividad permitió obtener resultados diferenciados según la variable trabajada. En el caso del empleo y la desocupación, los modelos mostraron relaciones parciales entre variables económicas agregadas y el comportamiento del mercado laboral. Esto sugiere que

existe una relación entre crecimiento, inversión y empleo, aunque dicha relación no resulta suficiente para explicar toda la variabilidad observada en la ciudad.

El comportamiento más limitado de algunos modelos frente al desempleo puede interpretarse como una señal de que esta variable depende de factores más amplios que los incorporados en la base, incluyendo elementos coyunturales, sectoriales y territoriales. Por ello, los resultados no deben entenderse como una negación de la relación entre variables económicas y desempleo, sino como evidencia de una capacidad explicativa parcial.

En contraste, cuando la pobreza fue utilizada como variable objetivo, algunos modelos presentaron un mejor ajuste. Esto sugiere que, para la base trabajada, la pobreza logra recoger de manera más estable ciertas condiciones estructurales del territorio, especialmente en relación con desempleo, informalidad, desigualdad y dimensión temporal. En consecuencia, esta variable aparece como particularmente útil para apoyar una lectura complementaria del desarrollo económico y sus efectos sociales.

En relación con la productividad y el comportamiento del PIB, los resultados mostraron un panorama más heterogéneo. Mientras algunas aproximaciones permitieron seguir tendencias generales, la red neuronal MLP presentó un desempeño favorable en el conjunto de prueba ($R^2 = 0.8921$), aunque con alta varianza en la validación cruzada (R^2 entre -0.31 y 0.73 según el fold), lo que refleja sensibilidad al tamaño de la ventana temporal y confirma su alcance como ejercicio exploratorio ajuste robusto. Este comportamiento sugiere que el PIB, aunque central para el análisis del desarrollo económico, no siempre resulta fácilmente modelable bajo las condiciones de la base empleada, especialmente cuando se utilizan arquitecturas más complejas y series relativamente limitadas.

Por su parte, la aplicación de K-means sobre la base manufacturera permitió identificar agrupamientos sectoriales diferenciados dentro de Bogotá. Este hallazgo sugiere que la estructura productiva de la ciudad no responde a un patrón único, sino que se encuentra segmentada entre sectores con dinámicas distintas. Desde la perspectiva de la competitividad y la política pública, esto aporta una lectura relevante, en la medida en que muestra que la base manufacturera de la ciudad no puede analizarse como un bloque homogéneo.

Más que ofrecer una sustitución de los enfoques tradicionales de evaluación, los resultados sugieren que el machine learning puede constituirse en una herramienta metodológica complementaria para el análisis de políticas públicas. Su principal aporte en esta investigación se observó en la posibilidad de comparar desempeños entre modelos, identificar patrones territoriales y sectoriales, y explorar relaciones entre variables que en enfoques tradicionales suelen abordarse de manera más agregada.

Sin embargo, el estudio también pone de presente que la utilidad del machine learning está condicionada por la calidad, disponibilidad y estructura de los datos, así como por la necesidad de mantener interpretabilidad en los resultados. Esto resulta especialmente importante en contextos de política pública, donde no basta con obtener métricas de ajuste aceptables, sino que se requiere que los hallazgos puedan ser comprendidos y utilizados de forma razonable dentro de procesos de análisis, seguimiento y toma de decisiones.

En este sentido, la investigación sugiere que el machine learning puede fortalecer la evaluación de la política pública de desarrollo económico en Bogotá, especialmente al aportar herramientas para examinar heterogeneidades y relaciones complejas. No obstante, sus resultados deben ser interpretados como evidencia complementaria y no como prueba definitiva del impacto integral de la política pública.

El estudio trabajó con información secundaria proveniente de distintas fuentes oficiales, con diferentes niveles de agregación, longitudes de serie y, en algunos casos, presencia de datos faltantes. Estas condiciones influyeron en el desempeño de varios modelos y limitaron la posibilidad de alcanzar ajustes más altos en determinadas variables.

Asimismo, variables como el desempleo y el PIB mostraron mayores dificultades de modelación, lo que sugiere que su comportamiento depende de factores más amplios que los incorporados en la base. De igual forma, el bajo ajuste de algunos modelos complejos, como las redes neuronales MLP, refuerza la idea de que la sofisticación metodológica no garantiza mejores resultados cuando las características de los datos no son las más adecuadas para ese tipo de algoritmo.

Los hallazgos del estudio deben entenderse como una aproximación analítica útil para explorar el potencial del machine learning en la evaluación de políticas públicas, más que como una medición concluyente del impacto total de la política pública de desarrollo económico en Bogotá.

Uno de los componentes centrales de la política pública es el fortalecimiento del tejido productivo, entendido en relación con las unidades productivas, el desarrollo empresarial, la competitividad y la consolidación de alternativas económicas para la ciudad. Desde este enfoque el modelo de regresión logística muestra la relación y el desempeño de localidades El tejido empresarial y consolidar alternativas productivas en la ciudad. Lo que conlleva a que sí sugieren que las dinámicas empresariales continúan operando con intensidades distintas según el territorio.

La aplicación de K-means a la base manufacturera también aporta a la discusión del tejido productivo, al mostrar que la estructura industrial de Bogotá puede organizarse en

agrupamientos sectoriales diferenciados. Esto refleja una estructura segmentada entre sectores más cercanos al promedio productivo, sectores con comportamiento particular y sectores con mayor dispersión relativa.

La inclusión productiva como un eje esencial, en la medida en que reconoce que el desarrollo económico debe beneficiar efectivamente a las personas y ampliar sus oportunidades de ingreso, empleo y participación en la actividad económica en las variables en pobreza, desempleo, ocupación e informalidad, La pobreza mostró una relación consistente con variables como la informalidad, el desempleo y el año, lo que determina que esta dimensión recoge de manera relativamente estable ciertas condiciones socioeconómicas del territorio. De igual forma, la mejor capacidad explicativa observada en la serie de personas desocupadas sugiere que variables como crecimiento del PIB e inversión extranjera guardan una relación parcial con el comportamiento del desempleo.

La política pública actualizada otorga un lugar importante a la gestión espacial de la productividad, al reconocer el papel del territorio, las economías de aglomeración y la distribución espacial de la actividad económica. Además, la política pública reconoce la importancia de las relaciones productivas, la articulación entre actores y las interdependencias que favorecen la productividad y el desarrollo económico. la pobreza mantiene relaciones con desempleo, informalidad.

Por su parte el desempeño empresarial presenta diferencias territoriales, ya que refleja que la estructura manufacturera se organiza en patrones sectoriales distintos. Esta evidencia no permite demostrar causalidades plenas, pero sí aporta una lectura según la cual las dinámicas del desarrollo económico en Bogotá se configuran a partir de relaciones múltiples entre condiciones laborales, estructura productiva, territorio y coyuntura temporal.

Los resultados evidencian que el desarrollo socioeconómico de Bogotá debe analizarse desde una perspectiva que articule tejido productivo, inclusión productiva, territorio, relaciones entre actores y seguimiento institucional. Los modelos implementados permiten observar que las tensiones entre el crecimiento económico agregado y las brechas territoriales persistentes son manifiestas tensiones entre el crecimiento económico con brechas persistentes son evidentes en la ciudad y pueden ser examinadas de manera complementaria mediante herramientas de machine learning.

El modelo de redes neuronales MLP, mediante un tratamiento especial de datos y mediante una arquitectura intermedia con regularización adecuada puede captar de forma razonable la trayectoria reciente del PIB de Bogotá.

Aunque el desempeño del modelo es favorable, logra capturar la trayectoria reciente del PIB, sin embargo, no implica por sí solo una validación predictiva robusta, debido a que el ejercicio sigue condicionado por el tamaño muestral disponible y por la naturaleza temporal de la serie y la disponibilidad de datos. En este sentido, el modelo MLP aporta evidencia útil para complementar el análisis de indicadores económicos asociados a la política pública, en la medida en que permite identificar patrones y apoyar la lectura de tendencias del comportamiento del PIB por sectores para la ciudad de Bogotá.

Los hallazgos también sugieren que el alcance de estos modelos depende de la naturaleza de la variable analizada, de la calidad y extensión de las fuentes disponibles y de la necesidad de mantener interpretabilidad en contextos de política pública

Este documento permitió validar aplicación de modelos de machine learning y establecerse como herramientas metodológicas pertinentes para complementar el seguimiento y

análisis de indicadores de la Política Pública de Productividad, Competitividad y Desarrollo Socioeconómico de Bogotá.

Además, permite identificar, a partir de las variables analizadas y de los modelos implementados, relaciones entre el PIB, la inversión extranjera, las personas ocupadas, las personas desocupadas, la pobreza, la informalidad, el índice de Gini, el desempeño empresarial y la estructura manufacturera, así como diferencias territoriales entre localidades y agrupamientos sectoriales que amplían la comprensión del desarrollo socioeconómico de Bogotá.

Por lo que este escrito aporta evidencia analítica y metodológica útil para fortalecer su seguimiento, enriquecer su evaluación y abrir nuevas posibilidades de investigación en torno al uso de la ciencia de datos en el campo de las políticas públicas y comprender las realidades complejas y establecer alternativas de solución que favorezcan la construcción integral de las ciudades, en pro de la generación de valor público y garantizando condiciones de calidad de vida a todos los habitantes de la ciudad de Bogotá.

Recomendaciones

Fortalecer la disponibilidad, la frecuencia y la calidad de los datos oficiales desagregados por localidad, dado que los resultados del estudio evidenciaron que la utilidad de los modelos de machine learning está condicionada por la cobertura, la homogeneidad y la periodicidad de la información proveniente de fuentes como el DANE, la Secretaría Distrital de Desarrollo Económico, la Secretaría Distrital de Planeación y el Observatorio de Desarrollo Económico de Bogotá.

Incorporar de manera gradual los modelos que presentaron mejor desempeño como herramientas complementarias, y no sustitutas, del seguimiento institucional de la política pública: Random Forest para el análisis de la pobreza por localidad, la regresión logística para el dinamismo empresarial, Support Vector Regression (SVR) para la desocupación y K-means para la identificación de agrupamientos sectoriales.

Ampliar el tamaño muestral y el tratamiento de las series temporales en variables como el PIB y el desempleo, cuya modelación presentó mayores dificultades; en particular, las redes neuronales MLP requieren un mayor número de observaciones para alcanzar un desempeño estable y generalizable.

Avanzar hacia enfoques interpretables y explicables, que integren la capacidad predictiva del machine learning con la solidez teórica de los métodos tradicionales de evaluación, con el fin de preservar la transparencia y la rendición de cuentas en el uso institucional de estas herramientas.

Referencias Bibliográficas

- Abrell, J., Kosch, M., & Rausch, S. (2019). How effective was the UK carbon tax? A machine learning approach to policy evaluation. SSRN Electronic Journal.
<https://doi.org/10.2139/ssrn.3373705>
- Aguilera-Díaz, M. M., Mora-Quiñones, R., & Ramírez-Rodríguez, C. (2020). Bogotá y área de influencia: tendencias económicas en las primeras dos décadas del siglo XXI (Documentos de Trabajo sobre Economía Regional y Urbana No. 291). Banco de la República. <https://doi.org/10.32468/dtseru.291>
- Alcaldía Mayor de Bogotá D. C. (2011). Decreto 064 de 2011, por el cual se adopta la Política Pública de Productividad, Competitividad y Desarrollo Socioeconómico de Bogotá Distrito Capital. Registro Distrital.
- Alcaldía Mayor de Bogotá D. C., & Consejo Distrital de Política Económica y Social del Distrito Capital. (2023). Documento CONPES D. C. 25. Actualización del plan de acción de la Política Pública de Productividad, Competitividad y Desarrollo Socioeconómico 2011–2038. Secretaría Distrital de Planeación.
- Alfonso Roa, Ó. A., & Amézquita, L. L. (2017). La calidad de Bogotá en perspectiva comparada: urbanismo, simbolismos y segregación residencial. En Ó. A. Alfonso Roa (Ed.), Bogotá en la encrucijada del desorden: estructuras socioespaciales y gobernabilidad metropolitana. Universidad Externado de Colombia.
- Aliaga-Linares, L., & Álvarez-Rivadulla, M. J. (2010). Segregación residencial en Bogotá a través del tiempo y diferentes escalas (Documento de trabajo WP10LL1SP). Lincoln Institute of Land Policy. <https://aciur.net/wp-content/uploads/2023/04/segregacion-residencial-bogota-aliaga-lisette-documento-2.pdf>

- Amarasinghe, K., Rodolfa, K. T., Lamba, H., & Ghani, R. (2023). Explainable machine learning for public policy: Use cases, gaps, and research directions. *Data & Policy*, 5, e5. <https://doi.org/10.1017/dap.2023.2>
- Amirullah, I., & Natsir, N. (2024). The influence of public policy on community subsectors: A study on the impact of economic and educational policies. *International Journal of Economics and Management Research*, 2(3), 287–297. <https://doi.org/10.55606/ijemr.v2i3.170>
- Artés, J., & Rodríguez-Sánchez, B. (2022). Métodos de evaluación de políticas públicas. *Papeles de Economía Española*, 172, 18–29.
- Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science*, 355(6324), 483–485. <https://doi.org/10.1126/science.aal4321>
- Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11, 685–725. <https://doi.org/10.1146/annurev-economics-080217-053433>
- Ávila, F. (2024). Marco de las Cinco Seguros para el uso ético de inteligencia artificial en sector público. Banco Interamericano de Desarrollo.
- Barre, R. (1962). *El desarrollo económico: análisis y política*. Fondo de Cultura Económica.
- Battaglini, M., Guiso, L., Lacava, C., Miller, D. L., & Patacchini, E. (2025). Refining public policies with machine learning: The case of tax auditing. *Journal of Econometrics*, 249, 105847. <https://doi.org/10.1016/j.jeconom.2024.105847>
- Bercu, B., Godichon-Baggioni, A., & Portier, B. (2019). An efficient stochastic Newton algorithm for parameter estimation in logistic regressions. arXiv. <https://arxiv.org/abs/1904.07908>

- Botta, A., Porcile, G., & Ribeiro, R. S. M. (2018). Economic development, technical change and income distribution: A conversation between Keynesians, Schumpeterians and Structuralists. Introduction to the special issue. *PSL Quarterly Review*, 71(285), 97–101. https://doi.org/10.13133/2037-3643_71.285_1
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth.
- Buechler, S., Niu, D., & Thompson, A. K. (2021). Predicting urban growth with machine learning. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3784787>
- Bukvic, R. M., & Pavlovic, R. Z. (2014). Neoclassical economic theory and the problems of economic development. *Ekonomika, Journal for Economic Theory and Practice and Social Issues*, 60(3), 1–19. <https://doi.org/10.22004/ag.econ.289165>
- Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review*, 81(5), 825–836. <https://doi.org/10.1111/puar.13293>
- Cámara de Comercio de Bogotá. (2025). Informe anual de dinámica empresarial en Bogotá y Cundinamarca. Observatorio de la Región.
- Céspedes, J. M. S. (2024). Desarrollo de un modelo de inteligencia artificial para ayudar en la toma de decisiones en política pública aplicado a la vocación agrícola desde la perspectiva sociodemográfica en el territorio colombiano [Tesis doctoral, Universidad Distrital Francisco José de Caldas]. Repositorio Institucional UD.
- Chaltseva, O. (2023). Evaluation of public policy as an indicator of its effectiveness: Experience for Ukraine. *Revue Européenne du Droit*, 1, 63–68. <https://doi.org/10.32782/2414-4436/2023-1-9>

- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Chenery, H. B. (1975). The structuralist approach to development policy. *The American Economic Review*, 65(2), 310–316.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1), C1–C68. <https://doi.org/10.1111/ectj.12097>
- Congreso de la República de Colombia. (2012). Ley 1581 de 2012, por la cual se dictan disposiciones generales para la protección de datos personales. *Diario Oficial* No. 48587.
- Consejo Nacional de Política Económica y Social, & Departamento Nacional de Planeación. (2019). Documento CONPES 3975. Política nacional para la transformación digital e inteligencia artificial. Departamento Nacional de Planeación.
- Corona, J. C., Diez, H. G., & Morell, C. (2020). Un estudio empírico del modelo de red neuronal MLP para problemas de predicción con salidas múltiples. *Serie Científica de la Universidad de las Ciencias Informáticas*, 13(6), 1–14.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Cusolito, A. P., & Maloney, W. F. (2018). Productivity revisited: Shifting paradigms in analysis and policy. World Bank Group. <https://doi.org/10.1596/978-1-4648-1334-4>
- Departamento Administrativo Nacional de Estadística. (2021). Pobreza multidimensional en Colombia: Año 2020 [Boletín técnico]. DANE.

https://www.dane.gov.co/files/investigaciones/condiciones_vida/pobreza/2020/boletin_tec_pobreza_multidimensional_20.pdf

Departamento Administrativo Nacional de Estadística. (2025a). Cuentas departamentales y producto interno bruto por departamento. DANE.

Departamento Administrativo Nacional de Estadística. (2025b). Gran Encuesta Integrada de Hogares, microdatos anonimizados. DANE.

Departamento Nacional de Planeación. (2022). Manual de evaluación de políticas públicas, código M-GI-01, versión 9. DNP.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv. <https://arxiv.org/abs/1702.08608>

Engeli, I., & Rothmayr Allison, C. (Eds.). (2014). Comparative policy studies: Conceptual and methodological challenges. Palgrave Macmillan. <https://doi.org/10.1057/9781137314154>

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

Fernandes, R. (2022). Public policies and social progress: Two evaluation models. *Journal of Public Policy*, 43, 199–214. <https://doi.org/10.1017/S0143814X22000289>

Figueroa Burdiles, N. (2012). El desarrollo y las políticas públicas. *Polis. Revista Latinoamericana*, 11(33), 375–391.

Furtado, B., & Andreão, G. (2022). Machine learning simulates agent based model towards policy. arXiv. <https://doi.org/10.48550/arXiv.2203.02576>

Galvis Aponte, L. A., & Meisel Roca, A. (2000). El crecimiento económico de las ciudades colombianas y sus determinantes, 1973–1998. Banco de la República. <https://doi.org/10.32468/dtseru.18>

Gao, J. (2024). Revolutionizing economic insights and development: The transformative power of machine learning. *Applied and Computational Engineering*, 57(1), 166–171.

<https://doi.org/10.54254/2755-2721/57/20241328>

Gao, Z., Wang, Z., & Zhou, M. (2023). Is China's urbanization inclusive? Comparative research based on machine learning algorithms. *Sustainability*, 15(4), 3490.

<https://doi.org/10.3390/su15043490>

Ghimire, B. R. (2024). Public policy and governance: Catalyzing local development and economic growth. *Journey for Sustainable Development and Peace*, 2(2), 39–53.

<https://doi.org/10.3126/jsdpj.v2i2.69562>

Hasan Chy, M. K., & Buadi, O. N. (2024). Role of machine learning in policy making and evaluation. *International Journal of Innovative Science and Research Technology*.

<https://doi.org/10.38124/ijisrt/ijisrt24oct687>

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. [https://doi.org/10.1007/978-0-387-](https://doi.org/10.1007/978-0-387-84858-7)

[84858-7](https://doi.org/10.1007/978-0-387-84858-7)

Hausmann, R. (2013). *The atlas of economic complexity: Mapping paths to prosperity*. MIT Press.

Helpman, E. (2004). *The mystery of economic growth*. Harvard University Press.

Hernández, J. A. (2024). *Modelos de machine learning para la evaluación de la política pública de seguridad alimentaria y nutricional en Bogotá 2019–2031* [Monografía, Universidad Nacional Abierta y a Distancia]. Repositorio Institucional UNAD.

<https://repository.unad.edu.co/handle/10596/67097>

- Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction policy problems. *American Economic Review*, 105(5), 491–495. <https://doi.org/10.1257/aer.p20151023>
- Lamba, H., Rodolfa, K. T., & Ghani, R. (2021). An empirical comparison of bias reduction methods on real-world problems in high-stakes policy settings. *ACM SIGKDD Explorations Newsletter*, 23(1), 69–85. <https://doi.org/10.1145/3468507.3468518>
- Lopez, J. W. M., Lopez, J. M. B., & Aguaded, I. (2025). Modelo predictivo para identificar hogares beneficiarios de programas de transferencias monetarias: Una comparación de técnicas de machine learning. *Revista Ibérica de Sistemas e Tecnologias de Informação*, 57, 3–18.
- Lourenço, L., Weber, L., Garcia, L., Ramos, V., & Souza, J. (2024). Machine learning algorithms to estimate propensity scores in health policy evaluation: A scoping review. *International Journal of Environmental Research and Public Health*, 21(11), 1484. <https://doi.org/10.3390/ijerph21111484>
- Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*, 1(4), 140–147. <https://doi.org/10.38094/jastt1457>
- Mayorga Henao, J. M. (2021). Comparación de los patrones espaciales e incidencia de la segregación residencial en las principales ciudades de Colombia. *Investigaciones Geográficas*, 75, 267–294. <https://doi.org/10.14198/INGEO.17548>
- Medina, F. (2001). Consideraciones sobre el índice de Gini para medir la concentración del ingreso. *Comisión Económica para América Latina y el Caribe*.
- Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (2nd ed.). Autor.

- Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87–106. <https://doi.org/10.1257/jep.31.2.87>
- Muñoz, G. (2022). Evaluación de políticas públicas: Marco conceptual y metodológico. *Boletín de Economía Política*.
- Márquez Ortiz, L. E., Cuétara Sánchez, L. M., Cartay Angulo, R. C., & Labarca Ferrer, N. J. (2020). Desarrollo y crecimiento económico: Análisis teórico desde un enfoque cuantitativo. *Revista de Ciencias Sociales*, 26(1), 233–253. <https://doi.org/10.31876/rcs.v26i1.31322>
- Observatorio de Desarrollo Económico de Bogotá. (2025). Indicadores socioeconómicos por localidad. *Boletín técnico*. Secretaría Distrital de Desarrollo Económico.
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing.
- Ordóñez, H., Cobos, C., & Bucheli, V. (2020). Modelo de machine learning para la predicción de las tendencias de hurto en Colombia. *Revista Ibérica de Sistemas e Tecnologías de Informação*, E29, 494–506.
- Oxford Poverty and Human Development Initiative. (2021). *Global multidimensional poverty index report*. University of Oxford.
- Pasco, L. O., & Bendezu, H. R. (2024). Public policies and their impact on economic development: A systematic review. *International Journal of Professional Business Review*, 9(11), e05068. <https://doi.org/10.26668/businessreview/2024.v9i11.5068>
- Pelsa, I., & Balina, S. (2022). Development of economic theory: From theories of economic growth and economic development to the paradigm of sustainable development. *DIEM*, 7(1), 91–101. <https://doi.org/10.17818/diem/2022/1.10>

- Perboli, G., Tronzano, A., Rosano, M., Tarantino, L., & Velardocchia, F. (2021). Using machine learning to assess public policies: A real case study for supporting SMEs development in Italy. *Proceedings of TEMSCON EUR*, 1–6. <https://doi.org/10.1109/TEMSCON-EUR52034.2021.9488581>
- Polo, D. M., Caballero, L. P., & Gómez, E. M. (2015). Comparación de redes neuronales aplicadas a la predicción de series de tiempo. *Prospectiva*, 13(2), 88–95.
- Programa de las Naciones Unidas para el Desarrollo. (2020). Índice de pobreza multidimensional global: Una nota técnica. PNUD.
- Quintero Montaña, W., Soria Freire, V., & Ruiz Dávila, B. D. (2022). Estructuralismo, neoestructuralismo y la vulnerabilidad macroeconómica en Latinoamérica. *Análisis Económico*, 37(95), 183–195. <https://doi.org/10.24275/uam/azc/dcsh/ae/2022v37n95/quintero>
- Rehill, P., & Biddle, N. (2023). Transparency challenges in policy evaluation with causal machine learning. *Journal of Public Policy*, 43(4), 580–601. <https://doi.org/10.1017/S0143814X23000178>
- Riascos, A. (2025). El potencial impacto del aprendizaje de máquinas en el diseño de las políticas públicas en Colombia: Una década de experiencias (Documento CEDE 2025-09). Universidad de los Andes.
- Rivera, J. (2015). Desarrollo y libertad de Amartya Sen. FLACSO Andes.
- Rodolfa, K. T., Lamba, H., & Ghani, R. (2020). Empirical observation of negligible fairness-accuracy trade-offs in machine learning for public policy. arXiv. <https://arxiv.org/abs/2012.02972>

Rudin, C. (2019). Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215.

<https://doi.org/10.1038/s42256-019-0048-x>

Salama, C., & Picalarga, S. (2024). Promoting policy evaluation across government: The contribution of the OECD recommendation to public policy evaluation. *Public Policy and Administration*. <https://doi.org/10.1177/13563890241234699>

Sandoval, C. G. (2025). Determinantes de la desigualdad en América Latina: Un análisis de datos de panel sobre el índice de Gini y variables socioeconómicas 1990–2022. *Revista IECOS*, 26(2), 96–131.

Secretaría Distrital de Desarrollo Económico. (2024). Informe de seguimiento del plan de acción del CONPES D. C. 25. Corte diciembre 31 de 2023. Alcaldía Mayor de Bogotá.

Secretaría Distrital de Desarrollo Económico. (2025a). Informe de seguimiento del plan de acción del CONPES D. C. 25. Corte diciembre 31 de 2024. Alcaldía Mayor de Bogotá.

Secretaría Distrital de Desarrollo Económico. (2025b). Informe de seguimiento del plan de acción del CONPES D. C. 25. Corte junio 30 de 2025. Alcaldía Mayor de Bogotá.

Secretaría Distrital de Planeación. (2024). Resolución 1634 de 2024, por medio de la cual se adopta la Guía de Evaluación de Políticas Públicas en el Distrito Capital. Alcaldía Mayor de Bogotá.

Secretaría Distrital de Salud de Bogotá, Observatorio de Salud de Bogotá-SaluData. (s. f.).

Pobreza en Bogotá D. C. SaluData. Recuperado el 12 de junio de 2026, de

<https://saludata.saludcapital.gov.co/osb/indicadores/pobreza-y-desigualdad-en-bogota-d-c/>

Sen, A. (1999). *Development as freedom*. Oxford University Press.

Silahtaroglu, G. (2024). Hybrid econometric and machine learning approaches in policy analysis. *Journal of Quantitative Economics*, 22(3), 451–478.

Stiglitz, J. E. (1998). Towards a new paradigm for development: Strategies, policies, and processes. Prebisch Lecture, UNCTAD.

Suguiura, F. O. R. (2022). Árbol de decisión en aprendizaje automático. *Revista Varianza*, 39.

Vargo, A., Zhang, F., Yurochkin, M., & Sun, Y. (2021). Individually fair gradient boosting. arXiv. <https://arxiv.org/abs/2103.16785>

Veale, M., & Brass, I. (2019). Administration by algorithm? Public management meets public sector machine learning. Oxford University Press.

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523), 1228–1242. <https://doi.org/10.1080/01621459.2017.1319839>

Xu, X., Chen, Z., & Chen, S. (2023). Enhancing economic competitiveness analysis through machine learning: Exploring complex urban features. *PLOS ONE*, 18(9), e0293303. <https://doi.org/10.1371/journal.pone.0293303>

Yáñez Cifuentes, M. T. (2022). Principales usos de la evaluación de políticas públicas y programas: Una mirada desde expertos para fomentar su utilización. *Revista Estudios de Políticas Públicas*, 8(1), 27–42. <https://doi.org/10.5354/0719-6296.2022.65039>