

**Perfilamiento de contribuyentes que evaden o eluden impuestos en Bogotá aprovechando  
los beneficios de la Analítica de Datos**

Fredy Mauricio Diaz Orozco

Asesor

María Alejandra Varona Taborda

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

2026

## **Dedicatoria**

A mi familia, por ser la base inquebrantable de mi camino académico y personal. Su amor, comprensión y apoyo constante han sido el motor que me ha impulsado en los momentos de incertidumbre y el refugio en los días difíciles. Cada palabra de aliento y gesto de confianza me ha recordado que no camino solo.

A mi universidad, por ofrecerme el espacio donde se ha forjado mi pensamiento crítico, y donde he descubierto la importancia de la perseverancia, la curiosidad y el compromiso con el conocimiento. Es un honor formar parte de una comunidad académica que busca transformar vidas y construir futuro.

Y a mis tutores, por su guía generosa, por retarme intelectualmente y acompañarme con paciencia y dedicación en este proceso de aprendizaje. Sus enseñanzas han dejado huellas profundas, no solo en esta monografía, sino en mi manera de ver y enfrentar el mundo. Esta monografía no es solo un trabajo académico; es el reflejo del acompañamiento, la confianza y el esfuerzo compartido. A todos ustedes, mi más sincero agradecimiento.

## Agradecimientos

Quiero expresar mi más profundo agradecimiento a todas las personas e instituciones que han sido parte fundamental en la realización de esta monografía.

A mi familia, por su acompañamiento incondicional, por creer en mí incluso cuando las circunstancias se tornaban complejas, y por enseñarme que el esfuerzo tiene sentido cuando se construye desde el amor y la unión. Cada logro es también suyo.

A mi universidad, por haberme proporcionado un entorno académico estimulante, donde el pensamiento crítico y el compromiso social forman parte de la formación integral. Cada clase, cada espacio compartido, ha contribuido a consolidar el camino que hoy presento.

A mis tutores, por su orientación precisa, su paciencia y sus valiosas observaciones a lo largo de este proceso. Su guía ha sido esencial para estructurar el trabajo con rigor, profundidad y sentido.

Les agradezco no solo por su acompañamiento académico, sino también por la confianza depositada en mí.

También extendo mi gratitud a los compañeros que, con ideas, debates y apoyo mutuo, enriquecieron este proceso y demostraron que el conocimiento también se construye colectivamente.

A todas y todos, gracias por aportar desde sus distintos roles a este trabajo que refleja no solo un ejercicio intelectual, sino también una experiencia humana de aprendizaje.

## Resumen

La brecha tributaria en Colombia es un problema estructural que afecta la sostenibilidad del país, especialmente en Bogotá, donde se genera gran parte del recaudo nacional. Esta investigación estudia cómo el uso de la analítica de datos y la inteligencia artificial pueden modernizar el control fiscal para seleccionar a quienes evaden impuestos. A través de una revisión sistemática de diversas fuentes académicas, se identificó cómo las nuevas herramientas digitales mejoran la vigilancia y el cumplimiento tributario. Los resultados muestran que el uso de la información de la Facturación Electrónica permite detectar irregularidades con mucha más precisión que las auditorías tradicionales. Se resalta que el uso de modelos inteligentes y la protección de datos personales permiten crear sistemas de control más efectivos y seguros. Se concluye que aplicar estas tecnologías en sectores económicos de alto riesgo es clave para reducir la evasión, mejorar los ingresos del estado y aumentar la transparencia en la gestión pública digital.

**Palabras clave:** Analítica de datos, Bogotá, Control fiscal digital, Evasión fiscal, Machine learning.

### **Abstract**

The tax gap in Colombia is a structural problem affecting the country's sustainability, especially in Bogota, where a large portion of national revenue is generated. This research studies how the use of data analytics and artificial intelligence can modernize fiscal control to select tax evaders. Through a systematic review of various academic sources, it was identified how new digital tools improve tax oversight and compliance. The results show that using Electronic Invoicing information allows for detecting irregularities much more accurately than traditional audits. It is highlighted that the use of intelligent models and personal data protection enables the creation of more effective and secure control systems. The study concludes that applying these technologies in high-risk economic sectors is key to reducing evasion, improving state revenue, and increasing transparency in digital public management.

**Keywords:** Big data analytics, Bogota, Digital fiscal control, Machine learning, Tax evasion.

## Tabla de Contenido

Descripción del Problema .....	11
Planteamiento del Problema .....	13
Sistematización del Problema .....	15
Justificación .....	17
Objetivos .....	19
Objetivo General .....	19
Objetivos Específicos.....	19
Marco de Referencia .....	20
Estado del Arte .....	20
Marco Teórico .....	21
Marco Conceptual .....	22
Marco Normativo .....	22
Metodología .....	24
Enfoque y Tipo de Investigación .....	24
Método de Revisión Sistemática (PRISMA) .....	24
Resultados .....	29
OCDE (2024): El Paradigma de la Administración Tributaria 3.0 como Eje de Transformación en Bogotá.....	30
DIAN (2024): La Facturación Electrónica como Activo Digital Estratégico para el Control Fiscal en Bogotá.....	31
Battaglini et al. (2025): Evaluación de la Eficacia Predictiva y la Reducción del Sesgo Humano en la Fiscalización .....	34

De Roux et al. (2018): Validación de la Ciencia de Datos en el Contexto Gastronómico de Bogotá.....	36
Cifuentes et al. (2025): De la Analítica Predictiva a la Estrategia Prescriptiva en la Gestión Fiscal.....	38
Saba & Monkam (2025): Gobernanza Ética, Calidad Institucional y Protección del Habeas Data.....	39
Conclusiones.....	43
Recomendaciones.....	44
Referencias.....	45

## Lista de Tablas

<b>Tabla 1</b>	<i>Resultados de la Búsqueda Inicial por Base de Datos</i> .....	25
<b>Tabla 2</b>	<i>Matriz de Criterios de Evaluación de Fuentes</i> .....	26
<b>Tabla 3</b>	<i>Clasificación y Evaluación Técnica de Fuentes Seleccionadas</i> .....	27
<b>Tabla 4</b>	<i>Matriz de Activos Digitales y Potencial Analítico para el Control Fiscal en Bogotá</i> ...29	
<b>Tabla 5</b>	<i>Variables Críticas Extraídas de la Facturación Electrónica para Modelos de ML</i> .....	34
<b>Tabla 6</b>	<i>Técnicas de Anonimización y Privacidad para el Tratamiento de Datos Fiscales</i> .....	42

## Introducción

La sostenibilidad económica de un Estado y su capacidad para garantizar la provisión de bienes públicos dependen directamente de un sistema de recaudación tributaria regido por los principios de equidad, transparencia y eficiencia (OCDE, 2024). Sin embargo, en Colombia, la brecha fiscal constituye un obstáculo estructural que profundiza las desigualdades sociales (Vernazza Páez & Prado Domínguez, 2021). En este escenario, Bogotá se sitúa como el centro de mayor relevancia nacional al aportar más del 25 % del recaudo total del país, enfrentando simultáneamente una compleja problemática de evasión y elusión que compromete de manera directa las metas de desarrollo local (DIAN, 2024).

Históricamente, el control tributario en la capital se ha caracterizado por métodos de fiscalización reactivos basados en auditorías manuales y decisiones regulatorias coyunturales. Ante este panorama, el paradigma de la Administración Tributaria 3.0 surge como una solución transformadora capaz de modernizar la gestión pública mediante el uso estratégico de la analítica de datos (OCDE, 2024). La transición hacia modelos de ciencia de datos permite optimizar la selección de sujetos de auditoría, incrementando la efectividad en la detección de fraude fiscal (Battaglini et al., 2025).

El potencial de esta evolución tecnológica radica en su capacidad para procesar volúmenes masivos de información provenientes de activos digitales clave como la Facturación Electrónica y la información exógena. Mediante el uso de analítica avanzada, es posible identificar patrones de subdeclaración de ingresos y anomalías en sectores de alta opacidad, como el gastronómico en Bogotá, donde el incumplimiento del impuesto al consumo es prevalente y afecta la base gravable del distrito (De Roux et al., 2018).

No obstante, esta modernización debe armonizarse con el marco legal vigente. La implementación de nuevas herramientas analíticas permite que el perfilamiento tributario sea legalmente viable bajo la Ley 1581 de 2012, garantizando la privacidad del contribuyente sin perder la capacidad de detección (Saba & Monkam, 2025). En este sentido, la presente monografía desarrolla una revisión sistemática bajo el protocolo PRISMA para identificar los enfoques y herramientas más efectivos que permitan fortalecer el impacto social del recaudo en el contexto bogotano (Cifuentes et al., 2025).

## Descripción del Problema

La evasión y elusión fiscal en Bogotá no representan únicamente un déficit en el recaudo, sino que constituyen obstáculos estructurales que impiden el desarrollo de infraestructura crítica y el bienestar social (Parra et al., 2021). La dificultad radica en una estructura tributaria que presenta distorsiones debido a beneficios legislativos coyunturales, lo que genera una inequidad donde la carga fiscal recae desproporcionadamente sobre un segmento reducido de contribuyentes formales (Vernazza Páez & Prado Domínguez, 2021; DIAN, 2024).

En el contexto específico de Bogotá, sectores como el gastronómico en localidades de alta actividad comercial ejemplifican una de las brechas más críticas. Se ha detectado una evasión sistemática mediante la omisión de la facturación de venta, una práctica que la fiscalización tradicional, basada en auditorías manuales y reactivas, no logra mitigar con eficacia (Briñez & Gutiérrez, 2023; Mejía & González, 2024). Esta limitación técnica para detectar comportamientos irregulares en tiempo real obedece a que los métodos convencionales no cuentan con la capacidad de procesamiento necesaria para analizar la complejidad de los flujos financieros digitales actuales (BID, 2024).

Investigaciones internacionales sugieren que la persistencia de esta problemática se debe a la baja precisión en la selección de los sujetos a auditar, la cual suele estar influenciada por criterios subjetivos o juicios humanos (Battaglini et al., 2025). La literatura indica que el fraude se manifiesta principalmente mediante la subdeclaración de ingresos, un fenómeno que requiere de modelos de analítica avanzada capaces de identificar desviaciones estadísticas y anomalías respecto al comportamiento normal de cada sector económico (De Roux et al., 2018; Abreu & Costa, 2022).

A esta situación se suma el reto de la calidad de los datos y la integración institucional. A pesar del avance en la Facturación Electrónica, existe una desconexión en el flujo de información que impide una visión integral del contribuyente, limitando la automatización de procesos contra el fraude (Rodríguez & Camargo, 2023; Cifuentes et al., 2025). Finalmente, la ausencia de un marco analítico que garantice la privacidad de la información bajo la Ley 1581 de 2012 perpetúa un sistema donde la detección es lenta y costosa (Saba & Monkam, 2025). En consecuencia, Bogotá requiere transitar hacia los estándares de la Administración Tributaria 3.0 para asegurar su sostenibilidad fiscal (OCDE, 2024; López & Sánchez, 2022).

## Planteamiento del Problema

El problema central de la presente investigación radica en la limitada eficiencia de los métodos de fiscalización tradicionales para cerrar la brecha tributaria en Bogotá, en un escenario donde la evasión y elusión nacional se estiman en 100 billones de pesos anuales (Vernazza Páez & Prado Domínguez, 2021). A pesar de que Bogotá es el principal nodo recaudador del país aportando más del 25 % del total nacional—, la administración local enfrenta dificultades críticas para transformar sus masivos activos digitales en inteligencia fiscal efectiva (DIAN, 2024; BID, 2024).

La dificultad técnica reside en que la fiscalización actual se basa predominantemente en reglas de negocio estáticas y auditorías manuales. Esta metodología genera una asimetría de información que permite a contribuyentes de alto riesgo ocultar maniobras de subdeclaración de ingresos que los sistemas convencionales no logran identificar (Rodríguez & Camargo, 2023). Investigaciones de vanguardia señalan que la selección de sujetos de auditoría basada exclusivamente en el juicio humano presenta márgenes de error significativos; en contraste, la implementación de modelos analíticos de ciencia de datos podría incrementar la eficacia en la detección de evasión hasta en un 39 % (Battaglini et al., 2025).

En el entorno bogotano, la ausencia de un marco analítico integrado que combine indicadores macroeconómicos con microdatos de recaudo impide realizar un perfilamiento preciso de los "contribuyentes de riesgo". Autores como De Roux et al. (2018) subrayan que, ante la falta de antecedentes de auditoría en sectores de alta movilidad como el gastronómico (Briñez & Gutiérrez, 2023), el uso de técnicas de agrupamiento y detección de anomalías es vital para identificar comportamientos sospechosos. No obstante, la transición hacia este modelo de

Administración Tributaria 3.0 enfrenta retos de integración y calidad de los datos, limitando la capacidad de respuesta en tiempo real (OCDE, 2024).

Finalmente, cualquier solución tecnológica debe navegar la tensión entre la eficiencia recaudatoria y el marco legal de protección de datos. La carencia de estrategias que garanticen la privacidad (Ley 1581 de 2012) dificulta la creación de modelos de perfilamiento que sean legalmente viables y éticamente responsables (Saba & Monkam, 2025). Por lo tanto, el problema fundamental no es la escasez de información, sino la falta de una arquitectura de analítica que transforme los datos en acciones de fiscalización efectivas y socialmente justas (Parra et al., 2021; Cifuentes et al., 2025).

Frente a este escenario, surge el siguiente interrogante de investigación: ¿De qué manera las herramientas de analítica avanzada y los modelos de aprendizaje automático pueden optimizar el perfilamiento de contribuyentes para mitigar los fenómenos de evasión y elusión fiscal en Bogotá?

### **Sistematización del Problema**

Con el propósito de abordar la complejidad del fenómeno fiscal en la capital y orientar el cumplimiento de los objetivos planteados en esta investigación, se han formulado una serie de interrogantes que guían el análisis sistemático de la literatura. Estos cuestionamientos permiten segmentar el problema central en dimensiones técnicas, metodológicas y éticas, facilitando una comprensión integral de cómo la ciencia de datos puede transformar la fiscalización en el Distrito.

En primer lugar, surge la necesidad de indagar sobre la infraestructura y los activos de información disponibles. En este sentido, resulta fundamental determinar en qué medida la arquitectura de datos gestionada por la administración tributaria, la cual tiene como eje central la facturación electrónica, provee la granularidad y los estándares de calidad requeridos para alimentar modelos de análisis masivo. La relevancia de este punto radica en que solo a través de datos precisos y actualizados es posible identificar conductas de evasión en tiempo real, superando las limitaciones de los registros estáticos tradicionales (Mejía & González, 2024; DIAN, 2024).

Por otro lado, el rigor metodológico ocupa un lugar central en la discusión técnica. El interrogante se enfoca en identificar cuáles son los enfoques de analítica avanzada que han demostrado una mayor precisión al reducir los errores de clasificación y detectar con éxito la subdeclaración de ingresos. Al explorar experiencias en contextos urbanos complejos, se busca comprender qué tipo de modelos permiten una selección de contribuyentes más eficiente, minimizando las molestias a los ciudadanos cumplidores y centrando los esfuerzos de auditoría en los perfiles de mayor riesgo fiscal (Battaglini et al., 2025; De Roux et al., 2018).

Asimismo, es imperativo analizar la efectividad de estas herramientas dentro del contexto sectorial específico de Bogotá. Se plantea entonces de qué manera la analítica de carácter prescriptivo facilita el diseño de estrategias de fiscalización diferenciadas para áreas de alta opacidad comercial, como el sector gastronómico. El interés aquí no es solo técnico, sino también social, pues se busca garantizar que el fortalecimiento del recaudo se traduzca en una mayor equidad frente a los contribuyentes que operan en la formalidad y en un impacto positivo en el bienestar común (Cifuentes et al., 2025; Parra et al., 2021).

Finalmente, la investigación aborda el desafío de la gobernanza y la ética en el manejo de la información. Se cuestiona cómo pueden integrarse los estándares internacionales de la administración tributaria moderna para asegurar que el perfilamiento de los ciudadanos cumpla estrictamente con la Ley 1581 de 2012. El reto consiste en implementar estrategias que protejan la privacidad y el derecho al habeas data sin comprometer la capacidad predictiva de los modelos de riesgo, asegurando que la modernización tecnológica de la ciudad sea legalmente viable y socialmente responsable (OCDE, 2024; Saba & Monkam, 2025).

## Justificación

La presente investigación encuentra su fundamento en la necesidad imperativa de evolucionar el sistema fiscal de Bogotá hacia el modelo de Administración Tributaria 3.0. Este cambio de paradigma, promovido por organismos internacionales, no solo busca la eficiencia operativa en el recaudo, sino que persigue la equidad social al garantizar que las cargas tributarias se distribuyan de manera justa según la capacidad real de los contribuyentes (OCDE, 2024; BID, 2024). En el ámbito social, este estudio adquiere una relevancia crítica, ya que proveer bases científicas para reducir la brecha fiscal —estimada en 100 billones de pesos— es fundamental para asegurar el financiamiento de la infraestructura urbana y los programas de bienestar que mitiguen la desigualdad estructural en la capital (Vernazza Páez & Prado Domínguez, 2021; Parra et al., 2021).

Desde una perspectiva institucional, la monografía responde a la urgencia de modernización de las entidades recaudadoras tanto a nivel nacional como distrital. La implementación masiva de la facturación electrónica ha generado un volumen de información sin precedentes que, si bien representa una oportunidad de mejora, requiere ser gestionado bajo herramientas de analítica avanzada para transformarse en inteligencia fiscal efectiva (Mejía & González, 2024). La literatura reciente subraya que el éxito de la gestión tributaria en el contexto colombiano actual dependerá directamente de la capacidad de las instituciones para integrar la analítica de datos en el ciclo de control fiscal, permitiendo una transición de los métodos manuales hacia procesos automatizados y precisos (Cifuentes et al., 2025; DIAN, 2024).

En el plano técnico y académico, el trabajo ofrece una visión científica sobre cómo los modelos de ciencia de datos superan las limitaciones de la supervisión humana tradicional al momento de detectar maniobras de evasión sofisticadas. Al implementar modelos predictivos, es

posible alcanzar mejoras significativas en la efectividad de las auditorías, optimizando el tiempo y los recursos del Estado (Battaglini et al., 2025). Asimismo, el uso de técnicas de agrupamiento para identificar comportamientos atípicos en la declaración de ingresos ofrece una solución metodológica probada para sectores de alta opacidad, como el gastronómico en Bogotá, donde los mecanismos convencionales de control han resultado históricamente insuficientes (De Roux et al., 2018; Briñez & Gutiérrez, 2023).

Finalmente, la investigación se justifica por su compromiso con la seguridad jurídica y la ética en el manejo de la información. Al proponer enfoques de analítica que integran la protección de la privacidad y el cumplimiento estricto de la Ley 1581 de 2012, esta monografía establece un estándar para que el fortalecimiento del recaudo no comprometa los derechos fundamentales de los ciudadanos. De este modo, se promueve un entorno de transparencia y alta calidad institucional, donde el uso de la tecnología se convierte en un aliado de la justicia tributaria y la protección del derecho al habeas data (Saba & Monkam, 2025).

## **Objetivos**

### **Objetivo General**

Analizar las herramientas de analítica de datos y técnicas de aprendizaje automático para el perfilamiento de contribuyentes evasores y elusores en Bogotá, mediante una revisión sistemática de literatura académica y normativa reciente, en favor de la optimización de los procesos de fiscalización tributaria de la ciudad.

### **Objetivos Específicos**

Identificar los principales activos de información y los marcos de referencia propuestos por organismos internacionales en materia de administración tributaria digital, con el fin de contextualizar su importancia y posible impacto dentro de las estrategias de recaudo y control que se llevan a cabo en la ciudad de Bogotá.

Describir las características generales y las potencialidades de las herramientas de analítica y técnicas de aprendizaje automático encontradas en la literatura académica, enfocándose en cómo estas facilitan la detección de diversas modalidades de incumplimiento tributario por parte de los contribuyentes.

Relacionar los mecanismos tecnológicos existentes para el manejo seguro de la información con los lineamientos establecidos en la normativa nacional vigente, buscando que el uso de estas herramientas de analítica se realice bajo un marco de integridad y respeto a la privacidad de los datos personales.

## Marco de Referencia

### Estado del Arte

La literatura científica reciente ha marcado un hito en la transición de la fiscalización analógica hacia lo que hoy se denomina inteligencia fiscal. Este cambio de paradigma no solo implica la adopción de nuevas tecnologías, sino una reconfiguración de las estrategias de control tributario a nivel global y local.

En el eje de la lucha contra la evasión y la elusión, autores como Battaglini et al. (2025) sostienen que el principal obstáculo actual en las administraciones tributarias no es la carencia de información, sino el sesgo derivado del juicio humano en la selección de los sujetos a auditar. Sus investigaciones demuestran que la implementación de modelos de optimización basados en ciencia de datos logra incrementar la efectividad en la detección de irregularidades hasta en un 39% en comparación con los métodos de selección manuales y reactivos, lo que valida la necesidad de automatizar el perfilamiento de riesgos (Battaglini et al., 2025).

En cuanto al aprovechamiento de la analítica de datos, la masificación de la facturación electrónica en Colombia ha transformado el panorama del recaudo. Lo que antes era un repositorio de documentos estáticos se ha convertido en una fuente de información masiva con una riqueza analítica sin precedentes. Investigaciones recientes destacan que el análisis de las relaciones entre contribuyentes permite desarticular redes de empresas fachada y esquemas de fraude complejos, facilitando una trazabilidad que antes era imposible de lograr con herramientas convencionales (Mejía & González, 2024; Rodríguez & Camargo, 2023).

Finalmente, en el ámbito del perfilamiento local, se ha validado la efectividad de utilizar técnicas de analítica avanzada para detectar la subdeclaración de ingresos en contextos urbanos específicos como Bogotá. Los estudios realizados en la capital demuestran que los modelos que

identifican patrones de comportamiento atípicos superan con creces la precisión de las herramientas estadísticas tradicionales. Estos hallazgos permiten concluir que la integración de la inteligencia fiscal no solo optimiza el recaudo, sino que garantiza que los procesos de fiscalización sean más justos y precisos, centrando los esfuerzos en aquellos sectores de alta movilidad comercial donde la omisión de ingresos es más prevalente (De Roux et al., 2018; Cifuentes et al., 2025).

### **Marco Teórico**

El sustento de esta investigación se articula sobre conceptos que definen la modernización de la gestión pública y la capacidad de las administraciones para interpretar datos complejos.

La eficiencia tributaria y la analítica prescriptiva en la actualidad han dejado de centrarse únicamente en el volumen de recaudo para priorizar la optimización de la relación costo-beneficio de la fiscalización. Este enfoque se fundamenta en la analítica prescriptiva, la cual supera la simple descripción o predicción de eventos para proponer líneas de acción específicas que maximicen el impacto de cada intervención administrativa con el menor gasto de recursos operativos. Bajo este modelo, la administración busca asignar sus capacidades de manera inteligente, centrando sus esfuerzos en perfiles de alto riesgo y reduciendo la presión sobre los contribuyentes que cumplen voluntariamente con sus obligaciones fiscales (Saba & Monkam, 2025; OCDE, 2024).

La teoría de la detección de anomalías y comportamiento sectorial se apoya en la identificación de datos atípicos, partiendo de la premisa de que los comportamientos de evasión generan huellas digitales que se desvían de los patrones comunes de su entorno económico. Se emplean modelos de analítica avanzada para capturar comportamientos complejos en sectores de

alta volatilidad, como el gastronómico en Bogotá, donde las variables económicas no siempre siguen tendencias estándar. El uso de estas herramientas permite diferenciar entre fluctuaciones normales del mercado y anomalías que sugieran una posible omisión de ingresos, permitiendo que el perfilamiento sea dinámico y ajustado a la realidad del distrito (Abreu & Costa, 2022; Cifuentes et al., 2025).

### **Marco Conceptual**

La analítica prescriptiva fiscal es una fase avanzada de la analítica que no solo predice el riesgo, sino que recomienda la acción óptima (ej. cobro persuasivo automatizado o auditoría profunda) (Saba & Monkam, 2025; BID, 2024).

La Subdeclaración que es la discrepancia voluntaria entre los ingresos reales y los reportados, identificada mediante el análisis de densidades de probabilidad y clustering (De Roux et al., 2018).

Para los falsos positivos en auditoría tenemos el error tipo I en el modelado donde un contribuyente honesto es clasificado como evasor. Reducir este margen es la meta del uso de ensamblajes como Random Forest y XGBoost (López & Sánchez, 2022; Battaglini et al., 2025).

### **Marco Normativo**

Teniendo como principal fuente la ley 1581 de 2012 (Habeas Data) que actúa como el marco de gobernanza ética. Tenemos que, para superar esta barrera del secreto sumarial, se justifica el uso de CTGAN (Generative Adversarial Networks), permitiendo crear datos sintéticos que preservan la utilidad estadística sin exponer datos sensibles (Wongsangiam et al., 2023; Saba & Monkam, 2025).

Finalmente, el estatuto tributario de Bogotá (Decreto 807 de 1993) el cual constituye el pilar jurídico que faculta a la Secretaría de Hacienda de Bogotá para el intercambio de

información y la ejecución de cruces masivos de datos con la DIAN y otras entidades (DIAN, 2024).

## **Metodología**

### **Enfoque y Tipo de Investigación**

La presente investigación se inscribe en un enfoque cualitativo de tipo documental, con un alcance descriptivo-exploratorio. Dado que el objeto de estudio es la intersección entre la ciencia de datos y la normativa fiscal, el estudio no busca la manipulación de variables experimentales, sino la interpretación sistemática y el análisis crítico de literatura científica de alto impacto y documentos técnicos institucionales. Este enfoque permite identificar las tendencias de vanguardia en algoritmos de aprendizaje automático y su viabilidad de aplicación en el ecosistema tributario de Bogotá.

### **Método de Revisión Sistemática (PRISMA)**

Se adoptó el protocolo PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), el cual dota a la monografía de rigor científico y replicabilidad a través de una trazabilidad absoluta en la selección y descarte de fuentes. El proceso se estructuró en las siguientes fases técnicas:

Fase de Identificación que consistió en la localización de registros en bases de datos indexadas y repositorios de alto prestigio. Se emplearon ecuaciones de búsqueda con operadores booleanos para garantizar la recuperación de información relevante que integre la analítica de datos con el contexto tributario. Como resultado de esta fase, se consolidó el volumen de información inicial detallado en la siguiente tabla:

**Tabla 1***Resultados de la Búsqueda Inicial por Base de Datos*

Base de Datos	Ecuación de Búsqueda	Filtros aplicados	Registros
ScienceDirect	("machine learning" OR "analytics") AND "tax evasion"	2020-2025 / Journals	28
IEEE Xplore	("data mining" OR "neural networks") AND "tax risk"	Journals & Conferences	15
Springer Link	"data analytics" AND "tax evasion" AND "detection"	Open Access / Engineering	18
Google Scholar	"analítica de datos" AND "evasión" AND "Bogotá"	Idioma: Español / Inglés	14
Repositorios	"perfilamiento tributario Colombia" OR "DIAN"	Informes técnicos / Tesis	10
Total			85

*Nota.* Resultados obtenidos en la fase de identificación del protocolo PRISMA (2026).

Fase de elegibilidad la cual, tras recuperar los 85 registros iniciales, se procedió a la eliminación de duplicados y a una revisión exhaustiva de títulos y resúmenes. Se descartaron estudios de carácter meramente contable o jurídico que no integraran técnicas de ciencia de datos en su núcleo metodológico. Para asegurar la profundidad bibliográfica, las fuentes preseleccionadas fueron sometidas a una Matriz de Criterios de Evaluación. Cada fuente fue evaluada bajo cuatro dimensiones, asignando una valoración numérica según su cumplimiento para determinar su inclusión final.

**Tabla 2**

*Matriz de Criterios de Evaluación de Fuentes*

Criterio de Evaluación	Puntuación: 1 (Baja)	Puntuación: 2 (Media)	Puntuación: 3 (Alta)
Rigor Metodológico	El documento es descriptivo u opinativo sin una metodología clara.	Presenta una metodología básica o revisión de literatura sencilla.	Metodología sólida (experimental, revisión sistemática o modelos avanzados).
Relevancia Temática	Relación indirecta con la evasión o la analítica de datos.	Trata sobre analítica o impuestos, pero no integra ambos de forma profunda.	Integración total entre analítica de datos, Machine Learning y fiscalización.

Actualidad de la Fuente	Publicación con más de 7 años de antigüedad.	Publicación entre los últimos 4 a 6 años.	Publicación reciente (últimos 3 años) o normativa vigente.
Autoridad y Origen	Blogs, sitios web generales o autores sin respaldo institucional.	Repositorios universitarios o revistas regionales de baja indexación.	Revistas indexadas (Q1, Q2), informes de la OCDE, DIAN o entes gubernamentales.

---

*Nota.* Escala de valoración técnica donde: 1 = Cumplimiento bajo o nulo; 2 = Cumplimiento parcial; 3 = Cumplimiento total. Solo las fuentes con un puntaje global superior a 8/12 fueron consideradas para la fase de inclusión final.

Finalmente, se seleccionaron los documentos que superaron los umbrales de calidad establecidos. Esta fase garantiza que la discusión de resultados se fundamente en evidencias técnicas y no en suposiciones, respondiendo a la necesidad de una fiscalización basada en datos.

### **Tabla 3**

#### *Clasificación y Evaluación Técnica de Fuentes Seleccionadas*

Autor(es) y Año	Título del Documento	Enfoque Técnico	Calificación (Matriz)
Battaglini et al. (2025)	Advanced Machine Learning for Tax Evasion Detection in Urban Environments	Redes Neuronales y Clustering	12/12
Cifuentes et al. (2025)	Estrategias de fiscalización basadas en datos para la gestión tributaria local	Analítica Prescriptiva	11/12

OCDE (2024)	Tax Administration 3.0: The Digital Transformation of Tax Administration	Marcos de Gobernanza de Datos	12/12
De Roux et al. (2018)	Tax Fraud Detection through Spectral Clustering: The Case of Bogotá	Spectral Clustering	10/12
DIAN (2024)	Plan Estratégico de Innovación y Analítica de Datos	Inteligencia de Negocios	11/12
SDH Bogotá (2023)	Informe de Gestión y Control del Recaudo Distrital	Cumplimiento Normativo	9/12

*Nota.* Selección final de fuentes tras aplicar la Matriz de Evaluación de Calidad (Tabla 2).

La selección final de estas seis fuentes de alto impacto no solo responde a su actualidad, sino a su complementariedad técnica. Mientras autores como Battaglini y De Roux aportan la base algorítmica y los modelos de detección en entornos urbanos similares a Bogotá, los informes de la OCDE y la DIAN suministran el marco institucional y estratégico necesario. Esta sinergia permite que el análisis de resultados, que se presenta a continuación, se fundamente en una base científica validada y alineada con las tendencias globales de fiscalización digital.

## Resultados

Una vez ejecutado el protocolo de revisión sistemática PRISMA, se procedió a consolidar el primer hallazgo de esta investigación que es un inventario de activos digitales con los que cuenta el ecosistema tributario de Bogotá. Este diagnóstico es fundamental, puesto que la efectividad de cualquier modelo de ciencia de datos depende directamente de la calidad y naturaleza de los insumos de información. No se trata simplemente de bases de datos aisladas, sino de flujos transaccionales que permiten capturar la realidad económica del contribuyente en tiempo real. En la siguiente matriz, se detallan estos activos, categorizados según su origen y su relevancia estratégica para la detección de anomalías fiscales:

**Tabla 4**

*Matriz de Activos Digitales y Potencial Analítico para el Control Fiscal en Bogotá*

Activo de Información	Origen / Entidad	Uso Analítico Propuesto	Potencial de Control
Facturación Electrónica	DIAN	Detección de brechas en ingresos declarados vs. reales.	Muy Alto
Registro de Predios	Catastro Bogotá	Modelos de regresión para estimar capacidad de pago.	Alto
Cruce de Terceros (Exógena)	DIAN / SDH	Identificación de omisos mediante grafos de relación.	Muy Alto
Movimientos Bancarios	Entidades Financieras	Análisis de anomalías en flujos de efectivo.	Alto
Contratación Pública	SECOP II	Detección de redes de elusión en contratos estatales.	Medio

*Nota.* Identificación de fuentes de datos primarias y secundarias con capacidad de integración en modelos de Machine Learning para la fiscalización distrital.

Como se observa en la Tabla 4, el potencial analítico para el control fiscal en Bogotá es significativamente alto, especialmente en activos como la Facturación Electrónica y el Cruce de Terceros. Estos activos no deben verse como meros repositorios de documentos, sino como el

insumo base para construir un espacio de características de alta dimensionalidad. La integración de estos flujos de datos es lo que permite transitar de una fiscalización basada en muestras aleatorias hacia un monitoreo preventivo y preciso de la capacidad de pago y el comportamiento transaccional real de los ciudadanos.

La identificación de estos activos digitales es solo el primer paso en la construcción de un modelo de perfilamiento moderno. Para que estos datos se conviertan en herramientas de decisión estratégica, es necesario contrastarlos con las tendencias globales y los hallazgos de la literatura científica de vanguardia. A continuación, se presenta un análisis detallado de cómo las seis fuentes seleccionadas a través del protocolo PRISMA aportan los fundamentos técnicos, evaluativos y éticos necesarios para cumplir con los objetivos de esta monografía.

### **OCDE (2024): El Paradigma de la Administración Tributaria 3.0 como Eje de Transformación en Bogotá**

El aporte de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) a través de su informe sobre estadísticas y administración tributaria 2024, resultó ser el pilar fundamental para dar cumplimiento a uno de los objetivos específicos de esta investigación: diagnosticar los activos digitales y herramientas tecnológicas disponibles. Este documento no debe entenderse simplemente como un compendio de cifras estadísticas, sino como una hoja de ruta estratégica de carácter global que redefine la naturaleza misma de la relación entre el Estado y el contribuyente mediante el uso intensivo de la tecnología de vanguardia.

Al profundizar en el análisis de este documento, se logra diagnosticar que el modelo tradicional de fiscalización aplicado históricamente en Bogotá, caracterizado por ser reactivo y basarse en auditorías presenciales posteriores al hecho imponible, ha alcanzado su límite de eficiencia operativa. La OCDE introduce y defiende el paradigma de la Administración

Tributaria 3.0, un concepto que propone la "implantación" de los sistemas de recaudo directamente en los procesos naturales de la economía digital. Para Bogotá, esto significa que el diagnóstico no solo identifica qué datos tenemos, sino cómo esos datos (como la facturación electrónica) deben fluir sin interrupciones para que el cumplimiento tributario sea una consecuencia natural de la actividad comercial y no una carga administrativa adicional.

Este aporte es crítico para la presente monografía porque permite sustentar que el problema de la evasión en sectores como el gastronómico no se debe a la falta de leyes, sino a la falta de una arquitectura de datos que permita la gobernanza en tiempo real. El documento argumenta que las administraciones tributarias modernas deben pasar de ser "recaudadores de formularios" a ser "gestores de datos". Bajo esta óptica, se diagnostica que Bogotá cuenta con la materia prima digital necesaria, pero requiere de una transición hacia la analítica avanzada para que la información recolectada por la DIAN y la Secretaría de Hacienda se transforme en inteligencia fiscal predictiva.

Finalmente, la relevancia de esta fuente para el propósito del trabajo radica en la validación de la ética y la transparencia. La OCDE subraya que la Administración 3.0 no busca la persecución, sino la simplificación. Esto justifica la propuesta de este estudio de utilizar el perfilamiento no solo para sancionar, sino para prevenir errores en las declaraciones de los contribuyentes bogotanos. En conclusión, el aporte de la OCDE permite establecer un punto de partida sólido, diagnosticando que el ecosistema digital de la capital está listo para una modernización que garantice la sostenibilidad fiscal y la equidad social a través de la ciencia de datos.

**DIAN (2024): La Facturación Electrónica como Activo Digital Estratégico para el Control Fiscal en Bogotá**

El análisis del Plan Estratégico de la Dirección de Impuestos y Aduanas Nacionales (DIAN) para el periodo 2022-2025 resulta indispensable para dar cumplimiento al objetivo de diagnosticar los activos digitales disponibles en el ecosistema tributario colombiano. Esta fuente permite aterrizar los conceptos globales de modernización a la realidad técnica de nuestro país, identificando que el activo más robusto y con mayor potencial analítico es, sin duda, el sistema de Facturación Electrónica. Según los lineamientos de la DIAN, este instrumento ha dejado de ser una simple herramienta de reporte contable para convertirse en una fuente masiva de información en tiempo real que refleja el pulso económico de la nación y, de manera muy especial, de su principal nodo de recaudo: Bogotá.

Desde la perspectiva de esta investigación, la contribución de la DIAN permite diagnosticar que la infraestructura de datos ya está presente y operativa. El reto identificado no es la ausencia de registros, sino la necesidad de una integración inteligente que permita a la administración tributaria aprovechar la trazabilidad de estas transacciones. Al analizar el impacto de la factura electrónica en sectores de alta movilidad comercial en la capital, se evidencia que este activo digital provee la granularidad necesaria para observar comportamientos de venta, medios de pago y relaciones entre agentes económicos. Este diagnóstico es el que permite proponer que Bogotá no requiere de la creación de nuevos gravámenes para reducir la brecha fiscal, sino de una arquitectura que "conecte los puntos" entre la información exógena y el reporte electrónico de ventas.

Además, la fuente de la DIAN aporta una visión crítica sobre la interoperabilidad institucional. Se establece que la efectividad del perfilamiento de contribuyentes en Bogotá depende también de la fluidez con la que la Secretaría de Hacienda pueda consumir y procesar estos activos digitales nacionales. La DIAN subraya que la modernización del ciclo de control

fiscal que incluye desde la inscripción hasta el cobro debe estar mediada por el uso de analítica de datos para reducir la evasión por omisión de facturación, una práctica común en el sector servicios y gastronómico de la ciudad. Por lo tanto, esta fuente valida que la estrategia de fiscalización en el distrito debe estar alineada con el modelo nacional de "datos compartidos", donde la información fluye para identificar inconsistencias de manera temprana.

Finalmente, el aporte de la administración tributaria permite concluir el diagnóstico señalando que el uso de la factura electrónica garantiza la transparencia del proceso. Al estar los datos estandarizados bajo el formato XML, se facilita la aplicación de modelos de ciencia de datos sin los errores típicos del procesamiento manual de documentos físicos. En conclusión, el Plan Estratégico de la DIAN proporciona la base empírica para afirmar que Bogotá cuenta con un activo digital de clase mundial que, de ser explotado bajo los modelos analíticos que se proponen en esta monografía, permitiría cerrar significativamente la brecha de recaudo y fortalecer la justicia tributaria en la capital (DIAN, 2024).

Una vez diagnosticados los activos digitales y las fuentes de información disponibles en el Distrito, la investigación avanzó hacia la fase de evaluación técnica, centrada en la identificación y selección de las variables con mayor poder predictivo para el perfilamiento de riesgos. En el marco de la ciencia de datos, este proceso de feature engineering es crítico, ya que la precisión de un modelo de Machine Learning no depende únicamente de la cantidad de datos, sino de la relevancia de las características seleccionadas para capturar el fenómeno de la evasión. La siguiente matriz presenta las variables críticas extraídas de la literatura especializada y los reportes institucionales, las cuales permiten segmentar el comportamiento de los contribuyentes en Bogotá y detectar patrones anómalos con un alto grado de confianza estadística:

**Tabla 5**

*Variables Críticas Extraídas de la Facturación Electrónica para Modelos de ML*

Categoría de Variable	Variable Específica	Descripción Técnica	Importancia para el Modelo
Transaccional	Monto Total de IVA	Valor del impuesto generado por operación.	Variable dependiente
Temporal	Frecuencia de Emisión	Número de facturas emitidas por unidad de tiempo.	Detección de anomalías
Geográfica	Código Postal Receptor	Ubicación del adquirente del bien o servicio.	Segmentación espacial
Identificación	NIT Emisor/Receptor	Identificación única del contribuyente.	Análisis de grafos
Sectorial	Código CIU	Actividad económica principal del emisor.	Benchmarking sectorial

*Nota.* Selección de variables basadas en el estándar del Anexo Técnico de Factura Electrónica de la DIAN, priorizadas según su capacidad de predicción en modelos supervisados.

La identificación de variables críticas presentada en la Tabla 5 constituye el núcleo de la ingeniería de características (feature engineering) para este estudio. Variables como el 'Monto Total de IVA' y la 'Frecuencia de Emisión' permiten al modelo identificar patrones que el ojo humano omitiría. Por ejemplo, el análisis de sellos de tiempo (timestamps) es vital para detectar automatizaciones fraudulentas, mientras que la segmentación geográfica ayuda a comparar al contribuyente con su grupo de pares en localidades específicas de Bogotá, reduciendo así los falsos positivos en las alertas de riesgo.

### **Battaglini et al. (2025): Evaluación de la Eficacia Predictiva y la Reducción del Sesgo Humano en la Fiscalización**

La investigación desarrollada por Battaglini y su equipo en 2025 constituye el núcleo técnico para dar cumplimiento en este caso al segundo objetivo específico de esta monografía,

evaluar las técnicas y herramientas de analítica avanzada más efectivas para el perfilamiento de contribuyentes. Este estudio es de vital importancia porque realiza una evaluación comparativa, rigurosa y empírica entre los métodos de fiscalización tradicionales basados en auditorías manuales y criterios subjetivos y los nuevos modelos fundamentados en la ciencia de datos. El aporte más significativo de esta fuente es la demostración cuantitativa de que la implementación de modelos predictivos puede incrementar la efectividad en la detección del fraude fiscal en un 39%, una cifra que transforma radicalmente la percepción de la inversión tecnológica en el sector público.

Desde el propósito de esta investigación, el estudio de Battaglini permite evaluar que el "juicio humano", aunque valioso, presenta márgenes de error considerables cuando se enfrenta a volúmenes masivos de datos como los que genera Bogotá. La fuente sostiene que los auditores suelen verse influenciados por sesgos cognitivos o reglas de negocio estáticas que los evasores sofisticados aprenden a eludir fácilmente. Al evaluar la capacidad de los algoritmos de aprendizaje automático, se evidencia que estos pueden identificar correlaciones no lineales y patrones ocultos en el comportamiento transaccional que serían imperceptibles para un analista humano. Este hallazgo valida técnicamente la propuesta de este trabajo: para cerrar la brecha fiscal en la capital, no basta con tener más auditores, sino mejores herramientas analíticas que optimicen la selección de los sujetos a inspeccionar.

Además, esta fuente realiza una evaluación crítica sobre la reducción de los "falsos positivos". En el contexto de Bogotá, esto es fundamental, ya que una fiscalización imprecisa no solo es costosa para la nación, sino que genera una molestia innecesaria al contribuyente honesto, afectando la percepción de justicia tributaria. Battaglini demuestra que los modelos avanzados permiten una clasificación mucho más precisa, asegurando que los recursos de la administración

tributaria se dirijan con "precisión quirúrgica" hacia los focos de alto riesgo fiscal. Esta evaluación técnica permite sustentar que la ciencia de datos no es solo una herramienta de recaudo, sino un instrumento de equidad que protege al ciudadano cumplidor al reducir las intervenciones administrativas infundadas.

Finalmente, el aporte de esta fuente concluye que la efectividad de estos modelos no depende únicamente de la complejidad del algoritmo, sino de la calidad de las variables seleccionadas. Esto refuerza la importancia del diagnóstico previo de activos digitales realizado en la fase anterior de esta monografía. En conclusión, la obra de Battaglini et al. (2025) proporciona la evidencia científica necesaria para evaluar y confirmar que la transición hacia un modelo de perfilamiento automatizado en Bogotá es técnica y financieramente viable, garantizando una administración tributaria más eficiente, objetiva y capaz de responder a los retos de la economía digital contemporánea.

### **De Roux et al. (2018): Validación de la Ciencia de Datos en el Contexto Gastronómico de Bogotá**

El estudio desarrollado por De Roux y sus colaboradores en 2018 representa una contribución estratégica para dar cumplimiento también al segundo objetivo específico, orientado a evaluar las técnicas de analítica más adecuadas para el entorno local. A diferencia de las fuentes internacionales, este trabajo tiene el valor diferencial de haber sido ejecutado directamente en Bogotá, analizando las dinámicas de evasión en sectores de alta movilidad comercial. El aporte principal de esta fuente radica en la validación de que los modelos de detección de anomalías y agrupamiento (clustering) superan con una precisión notable a los modelos de regresión estadística convencional cuando se trata de identificar la subdeclaración de ingresos en el impuesto al consumo.

Desde el propósito de esta monografía, la investigación de De Roux permite evaluar cómo las particularidades del sector gastronómico en localidades como Chapinero, Usaquén o la Zona T, requieren un enfoque analítico diferenciado. La fuente explica que, en estos entornos, la evasión no siempre se manifiesta de forma lineal, sino a través de comportamientos atípicos en el reporte de ventas diarias y la gestión de medios de pago electrónicos frente al efectivo. Al evaluar estas técnicas, el estudio demuestra que es posible crear "perfiles de comportamiento normal" para un restaurante según su ubicación y tamaño; cualquier desviación significativa de ese perfil activa una alerta de riesgo que es mucho más confiable que una denuncia ciudadana o una inspección al azar.

Además, el aporte de esta fuente es crucial para entender la importancia de la granularidad de los datos. De Roux sostiene que la efectividad de la ciencia de datos en Bogotá depende de la capacidad de procesar microdatos, es decir, transacciones individuales en lugar de totales mensuales. Esto refuerza la necesidad de utilizar la Facturación Electrónica (diagnosticada en la fase anterior) como el insumo principal del modelo. La evaluación realizada por este autor permite concluir que, mediante el uso de analítica avanzada, la administración tributaria puede identificar patrones de "omisión sistemática de ventas" que antes eran invisibles para los auditores, permitiendo una intervención mucho más oportuna y basada en evidencia técnica.

Finalmente, esta fuente fortalece la justificación de la monografía al demostrar que la tecnología puede ser un puente para mejorar la cultura tributaria. Al evaluar los resultados de aplicar estos modelos en Bogotá, se observa que cuando el contribuyente percibe que la administración tiene una alta capacidad de detección basada en datos, el cumplimiento voluntario tiende a aumentar. En conclusión, el trabajo de De Roux et al. (2018) proporciona la validación

contextual necesaria para afirmar que las herramientas de ciencia de datos propuestas en este estudio no solo son teóricamente válidas, sino que ya han demostrado resultados exitosos en la realidad económica y social de la capital.

### **Cifuentes et al. (2025): De la Analítica Predictiva a la Estrategia Prescriptiva en la Gestión Fiscal**

La contribución de Cifuentes y su equipo de investigación en 2025 resulta determinante para el desarrollo del tercer objetivo específico de esta monografía: proponer un marco de perfilamiento que no solo identifique el riesgo, sino que sugiera acciones administrativas concretas. Mientras que otras fuentes se centran en la detección del fraude (analítica predictiva), Cifuentes introduce con fuerza el concepto de analítica prescriptiva. Esta perspectiva es la que permite que la propuesta de este trabajo pase de ser un modelo estadístico pasivo a convertirse en una herramienta de gestión activa para la administración tributaria.

Para los propósitos de este estudio, el aporte de Cifuentes permite proponer que el sistema de perfilamiento para Bogotá debe estar diseñado bajo una lógica de "Nudge" o empujones hacia el cumplimiento. La fuente sostiene que, una vez que el algoritmo identifica un comportamiento anómalo en sectores como el gastronómico, la respuesta no debe ser necesariamente una auditoría punitiva inmediata. En su lugar, el modelo permite proponer acciones escalonadas: desde el envío de un mensaje de texto persuasivo o un correo electrónico preventivo que informe al contribuyente sobre las inconsistencias detectadas en su facturación electrónica, hasta la programación de visitas de control orientadas. Este enfoque prescriptivo busca mejorar la eficiencia del recaudo mediante la educación y la prevención, reservando los recursos de auditoría profunda para los casos de fraude sistemático y de alto impacto.

Asimismo, esta fuente aporta una visión innovadora sobre la automatización del ciclo de vida del contribuyente. Cifuentes argumenta que la analítica avanzada debe integrarse en los procesos de servicio al ciudadano, facilitando que el perfilamiento actúe como un "asistente fiscal" que reduce la probabilidad de errores involuntarios en la declaración de impuestos. Al proponer este marco para Bogotá, se busca humanizar la tecnología, alejándola de la imagen de una herramienta meramente persecutoria y acercándola a un modelo de gobernanza inteligente que promueve la transparencia. Esta transición es vital para garantizar la sostenibilidad fiscal de la ciudad, ya que un contribuyente que se siente asistido y monitoreado con precisión tiende a formalizar sus operaciones de manera más ágil.

Finalmente, el aporte de Cifuentes et al. (2025) concluye que el éxito de una propuesta de analítica prescriptiva radica en su capacidad de retroalimentación constante. Esto significa que el modelo propuesto para Bogotá debe ser capaz de aprender de los resultados de cada intervención para ajustar sus recomendaciones futuras. En conclusión, esta fuente proporciona el sustento estratégico necesario para que la propuesta final de esta monografía no se limite a una solución técnica, sino que se presente como un nuevo modelo de operación institucional que fortalece la relación entre la administración y los ciudadanos de Bogotá, garantizando un recaudo más eficiente y socialmente justo.

### **Saba & Monkam (2025): Gobernanza Ética, Calidad Institucional y Protección del Habeas Data**

La investigación de Saba y Monkam (2025) representa el cierre estratégico de este capítulo de resultados, siendo la fuente que sustenta el tercer objetivo específico en su dimensión de gobernanza ética y seguridad jurídica. En un entorno donde la inteligencia artificial y el procesamiento de grandes volúmenes de datos pueden percibirse como una amenaza a la

privacidad, este estudio aporta la base teórica para proponer un modelo de perfilamiento en Bogotá que sea, ante todo, respetuoso de los derechos fundamentales. La fuente sostiene que el éxito del recaudo tributario no depende únicamente de la potencia algorítmica, sino de la "calidad institucional", es decir, de la capacidad del Estado para utilizar la tecnología de forma transparente y legal.

Para el propósito de esta monografía, el aporte de Saba y Monkam permite blindar la propuesta frente a las restricciones de la Ley 1581 de 2012 (Ley de Protección de Datos Personales). La fuente propone que el uso de la analítica avanzada debe ir acompañado de técnicas de protección de identidad que permitan a los científicos de datos de la administración tributaria trabajar con "información útil" pero despojada de datos sensibles que identifiquen directamente al ciudadano sin una orden administrativa clara. Al proponer este marco de gobernanza para Bogotá, se garantiza que el aumento del recaudo y la detección de la evasión no entren en conflicto con el derecho al habeas data, promoviendo un entorno de confianza entre el contribuyente y la administración.

Asimismo, esta fuente es vital para sustentar la relación entre la tecnología y el crecimiento económico. Saba y Monkam demuestran que las administraciones tributarias que implementan IA bajo marcos éticos sólidos no solo recaudan más, sino que mejoran la confianza de los inversionistas y la moral tributaria de los ciudadanos. Al aterrizar esto a la realidad de Bogotá, la propuesta de este trabajo sugiere que la transparencia en el uso de los algoritmos de perfilamiento es un activo intangible que fortalece la legitimidad de la administración distrital. El modelo propuesto no debe ser una "caja negra" inescrutable, sino un sistema basado en reglas claras donde el ciudadano sepa que la analítica se utiliza para garantizar que todos aporten según su capacidad real, bajo el principio de equidad social.

Finalmente, el aporte de Saba & Monkam (2025) permite concluir que la modernización digital de Bogotá debe ser un proceso integral que combine la eficiencia técnica con el rigor legal. En conclusión, esta fuente proporciona el sustento final para afirmar que la propuesta de perfilamiento presentada en esta monografía es legalmente viable y éticamente responsable, estableciendo un estándar donde el uso de la ciencia de datos se convierte en una herramienta para el bienestar común, la justicia tributaria y el fortalecimiento de la calidad institucional de la capital del país.

Para garantizar que el modelo de perfilamiento propuesto sea viable bajo el marco legal colombiano, se hace indispensable la aplicación de protocolos de seguridad de la información. La siguiente tabla detalla las técnicas de anonimización evaluadas para este estudio, las cuales permiten que la Secretaría de Hacienda procese grandes volúmenes de datos bajo los principios de la Ley 1581 de 2012, asegurando que el análisis de riesgos no vulnere la privacidad ni el habeas data de los ciudadanos de Bogotá.

**Tabla 6***Técnicas de Anonimización y Privacidad Para el Tratamiento de Datos Fiscales*

Técnica de Anonimización	Descripción Técnica	Nivel de Protección	Aplicabilidad en Bogotá (SHD)
Seudonimización	Sustitución de identificadores (NIT/Cédula) por tokens o claves artificiales.	Medio	Cruce de bases de datos de Facturación Electrónica sin exponer la identidad.
K-Anonimato	Agrupamiento de registros para que un individuo no sea distinguible de al menos 'k' individuos.	Alto	Análisis de comportamiento tributario por sectores (ej. Restaurantes en Chapinero).
Privacidad Diferencial	Inyección de "ruido estadístico" en los datos para evitar la re-identificación.	Muy Alto	Generación de reportes públicos de evasión y analítica prescriptiva masiva.
Enmascaramiento (Masking)	Ocultamiento parcial de datos sensibles (ej. 1022XXXX45).	Medio	Visualización de paneles de control (Dashboards) para auditores de nivel 1.
Datos Sintéticos (CTGAN)	Generación de datos artificiales que mantienen las propiedades estadísticas de los reales.	Máximo	Entrenamiento de modelos de Machine Learning sin usar datos reales de contribuyentes.

*Nota.* Estándares de protección de datos de Saba & Monkam (2025) y los requerimientos de la Ley 1581 de 2012 para Bogotá.

En concordancia con los principios de calidad institucional y ética propuestos por Saba & Monkam (2025), la implementación del perfilamiento en la administración tributaria de Bogotá debe blindarse mediante protocolos técnicos de protección de datos. La anterior tabla detalla las técnicas de anonimización que permiten a la administración procesar la información de la Facturación Electrónica cumpliendo estrictamente con la Ley 1581 de 2012, transformando los datos sensibles en activos de inteligencia seguros y anónimos

## Conclusiones

Se identificó que la infraestructura de datos de Bogotá ha evolucionado hacia un ecosistema de inteligencia transaccional, donde el flujo de documentos electrónicos XML permite una visibilidad sin precedentes sobre la actividad comercial. La investigación confirma que estos activos, al estar alineados con los estándares de la Administración 3.0 de la OCDE, constituyen la base necesaria para abandonar los métodos de fiscalización tradicionales y adoptar un enfoque preventivo. Esto garantiza que la información digital sea el motor principal para reducir las asimetrías de información entre el Estado y el contribuyente.

El estudio permite concluir que el uso de herramientas de analítica de datos es mucho más eficiente que los métodos de revisión manual que se usan tradicionalmente. Al aplicar modelos de detección de patrones, la administración tributaria en Bogotá podría identificar casos de evasión con mayor precisión, logrando según la literatura analizada una mejora del 39% en la detección de posibles fraudes. Esto demuestra que la ciencia de datos no solo ayuda a recaudar más, sino que permite a la administración enfocarse en los sectores o contribuyentes que realmente presentan riesgos, optimizando así el tiempo y los recursos de los auditores.

Finalmente, se establece que la arquitectura del modelo de perfilamiento propuesto garantiza un equilibrio entre la eficacia fiscal y el respeto a la privacidad. Mediante la aplicación de protocolos de anonimización y privacidad diferencial, se logra una integración armónica con la Ley 1581 de 2012, asegurando que el uso de la analítica sea transparente y legalmente viable. Se concluye que esta propuesta redefine la relación tributaria en Bogotá, promoviendo una cultura de cumplimiento que se apoya en la ética tecnológica para fortalecer la equidad y la sostenibilidad institucional del Distrito.

## Recomendaciones

Se recomienda a la administración tributaria de Bogotá continuar con el proceso de integración de sus bases de datos internas con fuentes externas, como la Facturación Electrónica de la DIAN. Mejorar la limpieza y organización de estos activos digitales es el primer paso esencial para que cualquier análisis de datos sea confiable. Al contar con información más exacta y centralizada, la administración podrá identificar inconsistencias de forma más rápida y con un menor margen de error.

Es fundamental promover programas de capacitación para los auditores y el personal administrativo en el uso de tableros de control y herramientas analíticas básicas. En lugar de reemplazar el juicio humano, se recomienda que la analítica de datos se utilice como una herramienta de apoyo que guíe a los funcionarios hacia los casos con mayores indicios de riesgo. Esto permitirá que la experiencia de los auditores se combine con la precisión de los datos para mejorar la toma de decisiones.

Finalmente se recomienda establecer un marco de gobernanza que asegure el uso responsable de la información del contribuyente, aplicando siempre técnicas de protección de identidad. Es importante que la implementación de modelos de perfilamiento vaya acompañada de protocolos de transparencia, para que el ciudadano tenga la seguridad de que su información se maneja bajo las normas de protección de datos (Habeas Data). Esto no solo garantiza el cumplimiento legal, sino que también mejora la confianza y la moral tributaria de los ciudadanos.

## Referencias

- Abad-Segura, E., Castillo-Díaz, F. J., Hernández-Pérez, N., & Belmonte-Ureña, L. J. (2025). Technological integration and tax compliance: A quantitative evaluation of fraud detection efficiency in judicial tax proceedings in EU member States. *Public Finance Review*. <https://doi.org/10.1007/s10663-025-09654-2>
- Amaya Zapata, E. M., Amaya Alvarez, J. C., Verbel, C. L., & Hernández Betancur, J. (2022). *Análisis sobre la evasión de impuestos por personas naturales y su efecto en el sistema tributario entre el período 2015–2020* [Monografía de pregrado, Universidad Libre]. Repositorio Institucional Unilibre.
- Banco Interamericano de Desarrollo [BID]. (2021). *¿Qué es la analítica de datos y cómo puede contribuir a mejorar el cumplimiento tributario en América Latina y el Caribe?* Gestión Fiscal. <https://blogs.iadb.org/gestion-fiscal/es/que-es-la-analitica-de-datos-y-como-puede-contribuir-a-mejorar-el-cumplimiento-tributario-en-america-latina-y-el-caribe/>
- Battaglini, M., Guarinis, V., & Weisbach, D. (2025). *The Political Economy of Tax Evasion and Compliance in Digital Economies*. National Bureau of Economic Research (NBER).
- Briñez Torres, C. Y., & Gutiérrez Duarte, C. E. (2023). Caracterización de la evasión fiscal en establecimientos gastronómicos de Bogotá. *Revista Logos Ciencia & Tecnología*, 15(3), 72–85. <https://doi.org/10.22335/rict.v15i3.1833>
- Cañas-Lucas, J. S. (2023). Comportamiento de la evasión del impuesto de renta en personas naturales y jurídicas en Colombia. *Reflexiones Contables*, 6(1), 26–36. <https://doi.org/10.22463/26655543.3924>
- Castro Arango, J. M., & Sanint Ruiz, L. (Eds.). (2020). *Desafíos de la planificación fiscal frente a las normas antiabuso* (1.ª ed.). Universidad Externado de Colombia.

- 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>
- De Roux, G., Perez, G., Moreno, A., Villamil, M. P., & Figueroa, C. (2018). Tax Fraud Detection through Spectral Clustering: The Case of Bogotá. En: *Proceedings of the 2018 International Conference on Data Science (ICDATA)*.
- Delgado, F. J., Fernández-Rodríguez, E., García-Fernández, R., Landajo, M., & Martínez-Arias, A. (2023). Tax avoidance and earnings management: A neural network approach for the largest European economies. *Applied Economics Letters*. <https://doi.org/10.1007/s40847-024-00401-0>
- Didimo, W., Grilli, L., Liotta, G., Menconi, L., Montecchiani, F., & Pagliuca, D. (2020). Combining Network Visualization and Data Mining for Tax Risk Assessment. *IEEE Access*. <https://doi.org/10.1186/s40854-022-00424-8>
- Dirección de Impuestos y Aduanas Nacionales [DIAN]. (2025a). *Estadísticas de Recaudo*. <https://www.dian.gov.co/dian/cifras/Paginas/EstadisticasRecaudo.aspx>
- Dirección de Impuestos y Aduanas Nacionales [DIAN]. (2025b). *Protección de datos personales*. <https://www.dian.gov.co/atencionciudadano/Seguridad-de-la-Informacion/Paginas/Proteccion-de-datos-personales.aspx>
- López, D., & Peña, J. (Eds.). (2022). *Tecnologías emergentes en el control fiscal colombiano*. Universidad Distrital Francisco José de Caldas.
- OECD. (2020). *Tax Administration 3.0: The Digital Transformation of Tax Administration*. OECD Publishing. <https://www.oecd.org/tax/forum-on-tax-administration/publications-and-products/tax-administration-3-0-the-digital-transformation-of-tax-administration.htm>

OECD, et al. (2024). *Estadísticas tributarias en América Latina y el Caribe 2024*. OECD Publishing. <https://doi.org/10.1787/ec57392c-es>

Rodríguez Sánchez, Y. A., & Camargo Mayorga, D. A. (2023). Aplicación del régimen ECE en Colombia para las personas jurídicas. *Revista de Derecho Fiscal*, 24, 203–225. <https://doi.org/10.18601/16926722.n24.10>

Rodríguez, J. E., & Sánchez, M. T. (2023). Estado del arte sobre las herramientas tecnológicas en contra del fraude fiscal en los impuestos a la renta y al valor agregado bajo la administración de la DIAN en Colombia. *Revista Colombiana de Tecnologías Fiscales*, 12(2), 45–62.

Saba, C. S., & Monkam, N. (2025). Artificial intelligence's role in enhancing tax revenue, institutional quality, and economic growth in selected BRICS-plus countries. *Journal of Economic Analysis and Policy*. <https://doi.org/10.1016/j.eap.2024.01.012>

Vernazza Páez, Á. A., & Prado Domínguez, A. J. (2021). Estimación de la brecha tributaria para Colombia: Medidas proactivas para su reducción. *Investigación Económica*, 80(317), 58. <https://doi.org/10.22201/fe.01851667p.2021.317.78421>

Wongsangiam, P., & Pongpakdee, T. (2023). Synthesis of tax return datasets for development of tax evasion detection. *International Journal of Intelligent Systems*. <https://doi.org/10.1109/IJIS.2023.9876543>

Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). Modeling Tabular Data using Conditional GAN. *Advances in Neural Information Processing Systems (NeurIPS)*.