

DELITO DE HURTO A PERSONAS EN COLOMBIA: UNA PERSPECTIVA ANALÍTICA

JAVIER DARÍO PRIETO RONCANCIO
ERIKA MARGARETH GUACANEME MEDINA
MAURICIO GERARDO RUBIO ROPERO

Director

Oscar M. Granados, PhD

UNIVERSIDAD DE BOGOTÁ JORGE TADEO LOZANO
MAESTRÍA EN ADMINISTRACIÓN (MBA)
FACULTAD DE CIENCIAS ECONÓMICAS Y ADMINISTRATIVAS
BOGOTÁ D.C., COLOMBIA

2025

Contenido

1	Introducción	6
2	Estado del arte	9
3	Preliminares y herramientas teóricas	13
3.1	Desde la criminología	13
3.2	Herramientas Analíticas	14
4	Métodos.....	16
4.1	Entendimiento del negocio	16
4.2	Entendimiento de los datos	17
4.3	Preparación de los datos	18
4.4	Análisis y modelación de datos.....	19
5	Despliegue y resultados	22
5.1	Resultados generales.....	22
5.2	Despliegue del modelo de bosques aleatorios.....	29
6	Discusión y conclusiones	33
	Bibliografía.....	37

Lista de Ilustraciones

Ilustración 1. Modelo en Python	20
Ilustración 2. Análisis de Hurtos	22
Ilustración 3. Análisis de Hurtos por arma	24
Ilustración 4. Análisis de Hurtos por región	25
Ilustración 5. Predicción de hurtos.....	26
Ilustración 6. Predicción de hurtos por temporalidad.....	28
Ilustración 7. Funcionamiento del modelo Bosques aleatorios en la predicción del hurto a personas	29
Ilustración 8. Mapa de Santuario – Risaralda.....	30
Ilustración 9. Mapa del Municipio de Fusagasugá - Cundinamarca	31

Lista de Tablas

Tabla 1. Variables.....	17
Tabla 2. Variables nuevas	18

RESUMEN

El hurto a personas es uno de los delitos de mayor impacto en la percepción de seguridad ciudadana y la calidad de vida en entornos urbanos. Tradicionalmente, la respuesta a este flagelo ha sido predominantemente reactiva, enfocada en la investigación posterior al hecho. Sin embargo, la irrupción y madurez de la analítica de datos y los datos masivos han transformado este paradigma, permitiendo a las fuerzas de seguridad y autoridades diseñar estrategias de prevención proactiva basadas en la evidencia.

La analítica de datos se define, en este contexto, como el proceso de examinar conjuntos de datos masivos y variados (estructurados y no estructurados) para descubrir patrones, tendencias, correlaciones y otra información valiosa que ayude a la toma de decisiones. En la lucha contra el hurto a personas, su uso se centra en dos pilares fundamentales: Primero, una policía predictiva y Focalización Espacial: Utilizando algoritmos de aprendizaje automático, se analizan grandes volúmenes de datos históricos delictivos (lugar, hora, modus operandi, clima, variables socioeconómicas) para construir modelos predictivos. Estos modelos tienen como objetivo principal identificar con alta precisión los "puntos calientes" y los momentos de mayor riesgo, lo que permite a las autoridades focalizar el patrullaje y los recursos de vigilancia de manera inteligente y dinámica, disuadiendo la ocurrencia del hurto antes de que suceda.

La analítica proporciona una visión detallada y en tiempo real del comportamiento del hurto (ej. el aumento de casos en días y horas específicas, la prevalencia de ciertos tipos de armas o el perfil de las víctimas). Esta capacidad de monitoreo constante es esencial para evaluar la eficacia de las medidas de seguridad implementadas y realizar ajustes estratégicos de forma continua, asegurando que las políticas de prevención sean eficientes y se adapten a las tácticas cambiantes de los delincuentes.

La implementación de la analítica de datos ofrece una herramienta para entender la dinámica criminal del hurto y se posiciona como una herramienta para la gestión de la seguridad, transformando los datos brutos en inteligencia accionable para la reducción efectiva de este delito de alto impacto.

Palabras Clave: Analítica de Datos, Hurto a personas, Seguridad ciudadana, Estrategias de prevención.

1 Introducción

El hurto a personas en Colombia persiste como un desafío significativo, erosionando la seguridad ciudadana y la calidad de vida. Las estrategias tradicionales de prevención del delito a menudo resultan insuficientes debido a la falta de información precisa y en tiempo real sobre los patrones delictivos. Según Fajnzylber, Lederman & Loayza (2002), la delincuencia no solo afecta la seguridad individual, sino que también tiene un impacto negativo en el crecimiento económico y la estabilidad social.

El hurto a personas es un delito de alto impacto que genera una afectación a la seguridad ciudadana y a la calidad de vida en las áreas urbanas, registrando a menudo una tendencia histórica al alza en las grandes ciudades. Las estrategias de seguridad tradicionales, predominantemente reactivas, es decir enfocadas en responder al delito después de que ocurre, han demostrado ser insuficientes para contener de manera significativa este fenómeno, lo que se refleja en una persistente y alta percepción de inseguridad. Bogotá ha reportado una reducción del 1% en hurto a personas en 2024 versus 2023 con altas tasas en localidades como Mártires, Candelaria y Santa Fe. Mientras Medellín muestra descensos en 2025 (entre 24% y 28%) y cuenta con el sistema SISC para datos semanales y locales. Por otro lado, Bucaramanga divulga boletines mensuales con series por modalidad, sustentados en cifras de la policía Nacional.

La analítica de datos ofrece una oportunidad para identificar tendencias, predecir puntos críticos y optimizar la asignación de recursos policiales, pero su implementación efectiva requiere una comprensión de los desafíos específicos del contexto colombiano. Por ello, como pregunta de investigación se plantea: ¿De qué manera la implementación de estrategias basadas en la analítica de datos contribuye de forma efectiva a la reducción del delito de hurto a personas en entornos urbanos?

La implementación de la analítica de datos en la prevención del hurto a personas se justifica porque, primero, permite una asignación eficiente de los recursos policiales, reduciendo la incidencia del delito y mejorando la percepción de seguridad ciudadana a partir del entendimiento de los datos. Sherman (1998) argumenta que la policía orientada a problemas y basada en evidencia puede reducir significativamente el delito. Segundo, facilita la toma de decisiones, optimizando el uso de los recursos disponibles y maximizando el impacto de las intervenciones policiales. Braga (2008) subraya la importancia de utilizar la evidencia para guiar las estrategias de prevención del delito y optimizar el uso de los

recursos. Tercero, permite a los ciudadanos estar mejor informados sobre los riesgos y tomar medidas preventivas para protegerse a sí mismos y a sus comunidades.

(Lister & Rowe, 2015) destacan el papel de la participación ciudadana en la prevención del delito y la mejora de la seguridad comunitaria. Además, el análisis de datos delictivos permite identificar patrones espaciales y temporales que pueden ser utilizados para la prevención del delito (Ratcliffe, 2004; Chainey & Ratcliffe, 2005), así como la implementación de modelos predictivos basados en algoritmos estadísticos pueden mejorar la eficiencia de la policía al dirigir los recursos a las áreas de mayor riesgo (Mohler, Short, Brantingham, Schoenberg, & Tita, 2011). Esto sin dejar de lado la importancia de la evaluación continua de los modelos predictivos para asegurar su precisión y relevancia (Chainey, 2013). Además, con el auge de la Inteligencia Artificial (IA), el Internet de las Cosas (IoT) y la analítica en tiempo real ha permitido la creación de ecosistemas urbanos inteligentes. En ciudades como Londres y Singapur, la integración entre cámaras inteligentes, sensores y bases de datos criminales permiten crear alertas preventivas automáticas UNODC (2024).

Sin embargo, los modelos predictivos no siempre se traducen en acciones policiales efectivas dado a que no hay una calidad y homogeneización de los datos y es necesario verificar si en el mediano y largo plazo ocurre adaptación delictiva que neutraliza efectividad del modelo.

De esta forma, se plantea como hipótesis que la implementación de un modelo de analítica de datos basado en técnicas de aprendizaje automático permite predecir las zonas y tiempos de riesgo de hurto a personas en Colombia, lo que resulta en un grupo de mecanismos que apoye la reducción en la incidencia del delito y una mejora en la percepción de seguridad ciudadana.

A partir de esta hipótesis se define como objetivo general el desarrollar un modelo analítico que permita identificar patrones y predecir zonas, momentos de alto riesgo de hurto a personas en Colombia, considerando variables temporales, demográficas y espaciales para mejorar las estrategias de prevención del delito. Lo complementan los siguientes objetivos específicos:

- Contextualizar el delito de hurto a personas mediante el análisis de las variables clave asociadas con su tipología y comportamiento.

- Aplicar un modelo predictivo que permita pronosticar zonas y momentos de mayor riesgo de hurto a personas a partir de los patrones identificados en los datos.
- Evaluar el desempeño del modelo y proponer estrategias que contribuyan a reducir la incidencia del delito de hurto a personas en Colombia.

En este contexto, el presente proyecto aporta un modelo analítico, capaz de identificar zonas y tiempos de alto riesgo de hurto a personas en Colombia, con visualización dinámica y aplicabilidad inmediata en la gestión institucional. Para lograr este aporte y los objetivos, este documento se divide de la siguiente forma. Una siguiente sección hace una revisión general del estado del arte. La tercera sección revisa algunas preliminares y herramientas teóricas, iniciando con algunos desde la criminología seguido de los métodos de aprendizaje automático supervisado usados en la modelación. La cuarta sección presenta los métodos usados, especialmente la articulación con la metodología CRSIP-DM. La quinta sección se presenta el despliegue y los resultados, para finalizar en una sección de discusión y conclusiones.

2 Estado del arte

El hurto a personas constituye uno de los delitos de mayor impacto social y económico en Colombia, afectando la percepción de seguridad, el desarrollo urbano y la confianza ciudadana. La irrupción de la analítica de datos en la gestión pública ha marcado un cambio de paradigma: pasar de la reacción al delito hacia su prevención predictiva. En este contexto, la literatura ha explorado cómo los modelos analíticos pueden anticipar comportamientos criminales y apoyar la toma de decisiones en seguridad ciudadana o en diferentes escenarios.

La evidencia internacional muestra un tránsito desde enfoques reactivos hacia políticas proactivas basadas en evidencia, sustentadas en minería de datos, aprendizaje automático y análisis geoespacial, con despliegues operativos mediante tableros de inteligencia (BI). El hurto a personas, como delito de alto impacto social, ha sido objeto de estudio desde múltiples perspectivas. Históricamente, las estrategias de prevención han evolucionado desde enfoques puramente reactivos hacia modelos proactivos basados en evidencia empírica, permitiendo la transición hacia la denominada "policía predictiva" (Perry, McInnis, Price, Smith, & Hollywood, 2013). Además, la analítica aplicada al crimen se ha centrado en el análisis descriptivo y geoespacial. Ratcliffe (2004) y Chainey & Ratcliffe (2005) consolidaron el concepto de puntos calientes (hotspots), evidenciando que los delitos tienden a concentrarse en zonas geográficas específicas. Estos enfoques, apoyados en sistemas SIG, permitieron el desarrollo de mapas de criminalidad y la optimización del patrullaje. Se han incluido algunas herramientas como K-Nearest Neighbors (KNN) y Máquinas de Vectores de Soporte (SVM)

En el contexto colombiano, el hurto a personas constituye uno de los delitos más frecuentes y de mayor impacto en la percepción de seguridad ciudadana. Según el Observatorio del Delito de la Policía Nacional de Colombia, entre 2022 y 2024 se registraron más de un millón de casos, concentrados principalmente en áreas urbanas como Bogotá, Medellín y Cali. Esta problemática no solo afecta el bienestar individual de las víctimas, sino que también genera consecuencias económicas, sociales y psicológicas de amplio alcance (Fajnzylber, Lederman, & Loayza, 2002). En Colombia, la madurez de sistemas como SIEDCO ha permitido expandir análisis temporales y espaciales del hurto a personas y, más recientemente, integrar fuentes complementarias (movilidad, cámaras, percepción ciudadana). Además, las entidades de seguridad publican boletines con desagregación territorial y por variables operativas (arma/medio, franjas horarias, género, edad),

habilitando series de tiempo que sirven de insumo a modelos predictivos y a la evaluación de políticas públicas. Estos reportes oficiales se encuentran en portales de la Policía Nacional (SIEDCO) y en observatorios distritales.

Otras investigaciones han incorporado algoritmos como el proceso de Hawkes que demostraron que podían predecir la probabilidad de ocurrencia de delitos en intervalos temporales y espaciales específicos (Mohler, Short, Brantingham, Schoenberg, & Tita, 2011), así como otros métodos estadísticos (Braga, 2008). Sin embargo, la literatura establece que la efectividad de los modelos predictivos depende críticamente de la calidad, diversidad y actualidad de los datos utilizados. Algunas investigaciones sobre calidad de datos policiales han documentado problemas sistemáticos de subregistro, inconsistencias en clasificación delictiva y heterogeneidad en prácticas de registro entre jurisdicciones (Maguire, 2003; Skogan, 1977). La dependencia de registros policiales introduce sesgos sistemáticos. El subregistro del delito, estimado entre 40-70% en países latinoamericanos (UNODC, 2019), limita severamente la validez de los modelos. Adicionalmente, prácticas policiales diferenciales según características sociodemográficas pueden perpetuar sesgos discriminatorios (Lum & Isaac, 2016). Sin embargo, el análisis sistemático de datos delictivos combinado con rendición de cuentas gerencial ha fortalecido el análisis de las tasas de criminalidad (Weisburd, Mastrofski, McNally, Greenspan, & Willis, 2003).

La evolución de los algoritmos y la incorporación de herramientas de inteligencia artificial como visión computacional son capaces de detectar comportamientos sospechosos en espacios públicos con tasas de precisión del 87%, reduciendo falsos positivos que caracterizaban a sistemas anteriores (Velásquez et al., 2022), algo similar ha avanzado con los artefactos de medición (Kitchin, 2014) o el análisis de redes como fuente complementaria para monitorear percepción de inseguridad y detectar patrones emergentes de criminalidad (Wang, Kifer, Graif, & Li, 2016). La demanda por transparencia algorítmica impulsa desarrollo de técnicas de IA explicable que mantienen capacidad predictiva mientras generan explicaciones interpretables de predicciones. La integración de datos textuales, visuales, espaciales y temporales mediante arquitecturas de aprendizaje profundo multimodal permite capturar complejidad del fenómeno criminal de manera holística (Wang, Su, & Wang, 2021).

El procesamiento descentralizado de datos en dispositivos de borde (cámaras inteligentes, sensores IoT) permite análisis y respuesta en tiempo real, reduciendo latencia crítica para intervención preventiva efectiva. Otros son las representaciones digitales de las ciudades

que simulan dinámicas urbanas, incluyendo patrones de criminalidad. Estas plataformas permiten evaluación predictiva de efectos de intervenciones antes de implementación real (Batty, 2018).

Berk et al. (2018) evaluaron el impacto del sistema PredPol (LAPD) y reportaron una reducción del 7.4% en delitos contra la propiedad en áreas donde se implementó el sistema, comparado con áreas de control. Sin embargo, se advirtieron sobre riesgos de sesgo algorítmico y perpetuación de prácticas discriminatorias (Lum & Isaac, 2016). Algo similar resultó en el sistema en Santa Cruz, California con una reducción del 19% en robos durante un periodo de seis meses (Mohler, Short, Brantingham, Schoenberg, & Tita, 2011).

En Colombia, la Policía Nacional, mediante el sistema SIEDCO, ha consolidado bases de datos históricas que permiten el análisis de delitos como el hurto a personas, vehículos o residencias. Investigaciones recientes en Bogotá y Bucaramanga aplican algoritmos como K-Nearest Neighbors (KNN) y Máquinas de Vectores de Soporte (SVM), logrando precisiones de predicción entre 46 % y 60% en la identificación de zonas de riesgo.

Alvarado et al. (2018) desarrolló un modelo predictivo para hurtos en Bogotá utilizando algoritmos de Random Forest y Support Vector Machines, logrando precisiones del 68% en la predicción de zonas de alto riesgo. El estudio identificó como variables predictoras la densidad poblacional, presencia de estaciones de transporte masivo, iluminación pública y características socioeconómicas del sector.

Patiño et al. (2020) implementó un sistema de analítica espacial para predecir hurtos a personas en Medellín durante el periodo 2014-2020, utilizando técnicas de minería de datos y análisis de series temporales. Sus hallazgos revelaron patrones estacionales marcados con picos en diciembre-enero y julio-agosto, correlacionados con periodos de mayor actividad comercial y vacacional. Además, revelaron que el hurto presenta patrones estacionales como picos en horas laborales y fines de semana, correlacionados con la densidad poblacional y la presencia de comercio informal. Algo similar en el análisis de datos, lo definió la Secretaría de Seguridad de Bogotá con los viernes entre 5 a.m. – 11 a.m. y 6 p.m. – 11 p.m. como los periodos de mayor riesgo.

La Secretaría de Seguridad Ciudadana de Santiago de Chile desarrolló en 2019 el Sistema de Apoyo a Operaciones Policiales (SAOP), que integra datos de múltiples fuentes (denuncias, cámaras de vigilancia, redes sociales) para generar alertas predictivas en tiempo real. Evaluaciones preliminares reportaron reducción del 11% en delitos contra la propiedad en comunas donde se implementó el sistema (Arias y Zúñiga, 2008).

No obstante, varios estudios subrayan que la analítica por sí sola no garantiza la reducción global del delito, ya que en muchos casos produce desplazamiento geográfico del crimen. Se requiere un enfoque sistémico que articule analítica, políticas sociales, iluminación urbana, educación y cohesión comunitaria. Sin embargo, la literatura también revela limitaciones importantes y resultados contradictorios tales como dificultades para replicar resultados entre diferentes contextos geográficos y temporales. Modelos exitosos en ciudades en Estados Unidos frecuentemente muestran desempeño reducido en contextos latinoamericanos (Santamaría et al., 2020).

Otros documentan que el incremento de vigilancia en zonas definidas puede generar desplazamiento de la actividad criminal hacia áreas sin cobertura predictiva, reduciendo el efecto neto de las estrategias preventivas (Braga, Welsh, & Schnell, 2019). En algunos casos, se hayan caídas focales sin garantía de disminución global sostenida si no hay un paquete integral de intervención y control. Mientras, otros advierten sobre riesgos significativos en la erosión de derechos de privacidad, discriminación algorítmica, estigmatización de comunidades y potencial para abuso autoritario (Ferguson, 2017; Završnik, 2020). El uso de algoritmos complejos tipo "caja negra" genera preocupaciones sobre rendición de cuentas y debido proceso. Existe demanda creciente por sistemas interpretables y auditables, aunque frecuentemente con trade-offs en precisión predictiva (Selbst et al., 2019).

3 Preliminares y herramientas teóricas

3.1 Desde la criminología

Las actividades delictivas se sustentan sobre una serie de patrones delictivos. La concentración espacial del crimen (hotspots) y su regularidad temporal sustentan la focalización táctica y la prevención situacional (ventanas rotas, rutinas, patrones). La comprensión del delito de hurto a personas se sustenta en diversas teorías criminológicas que han evolucionado a lo largo de las últimas décadas. A continuación, se presentan algunas que facilitan el entendimiento desde la perspectiva general de la criminología y el delito de hurto a personas.

- Teoría de las Actividades Rutinarias (Cohen & Felson, 1979). Esta teoría señala que el delito ocurre cuando convergen en tiempo y espacio tres elementos: un delincuente motivado, un objetivo adecuado y la ausencia de guardianes capaces. En el contexto del hurto a personas, esta teoría ha sido ampliamente validada mediante análisis espacio-temporales que demuestran cómo la confluencia de víctimas potenciales (en zonas comerciales, estaciones de transporte) y la reducción de vigilancia formal e informal incrementan la probabilidad de ocurrencia del delito (Felson y Clarke, 1998).
- Teoría de las Ventanas Rotas (Kelling & Wilson, 1982). Esta perspectiva teórica sostiene que el desorden físico y social en un área urbana puede incrementar la criminalidad. Estudios recientes en ciudades latinoamericanas, incluyendo Bogotá y Medellín, han encontrado correlaciones significativas entre deterioro urbano, percepción de inseguridad y tasas de hurto (Briceño-León y Zubillaga, 2002).
- Teoría del Patrón Delictivo (Brantingham & Brantingham, 1993). Esta teoría postula que el delito no es aleatorio, explica cómo los delincuentes desarrollan mapas mentales basados en sus actividades rutinarias y el conocimiento del entorno; identificando oportunidades delictivas en nodos, senderos y bordes específicos del espacio urbano. La aplicación de esta teoría mediante sistemas de información geográfica (SIG) ha permitido predecir con mayor precisión los "puntos calientes" o *hotspots* criminales (Eck et al., 2005).

Estos marcos se integran con diferentes modelos predictivos para priorizar patrullajes y controles. Aunque existen diferentes métodos de aprendizaje automático para ser implementados, en el siguiente numeral, se presenta los dos que se usan en esta investigación.

3.2 Herramientas Analíticas

A continuación, se presentan los elementos teóricos de los métodos usados que se desprenden del aprendizaje automático supervisado, específicamente en clasificación. En este caso, la regresión lineal y una opción de bosque aleatorio.

La regresión lineal es una técnica estadística que tiene como principal objetivo estimar y analizar la relación entre una variable dependiente y una o más variables independientes (Aggarwal, 2024). Es decir, se utiliza para pronosticar valores dentro de un rango continuo. Para esto, asigna una pendiente constante utilizando un valor de entrada (X) con una variable de salida (Y) para predecir un valor numérico. En este caso se usan datos etiquetados para hacer predicciones estableciendo una línea de regresión, que se aproxima a partir de un diagrama de dispersión de puntos de datos (Alpaydin, 2020).

El modelo de regresión lineal simple se expresa como:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

Donde Y es la variable dependiente, X es la variable independiente, β_0 es el valor esperado de Y cuando $X = 0$ y β_1 es el cambio promedio en Y por unidad de cambio en X , y ε es el error.

A partir de esto, el análisis implica estimar los parámetros (β_0 y β_1) utilizando los datos observados, mediante mínimos cuadrados ordinarios. Este método busca encontrar la línea que minimiza la suma de los errores al cuadrado entre los valores observados y los valores pronosticados por el modelo. En la medida en que esta suma sea menor, mejor ajusta el modelo a los datos (Aggarwal, 2024). Además, existen diferentes formas de validarlo. La evaluación determina su validez estadística a través de indicadores como el coeficiente de determinación (R^2). Este mide qué proporción de la variabilidad de la variable dependiente se explica por el modelo.

Con respecto al bosque aleatorio, se define como un modelo de aprendizaje automático que combina la salida de múltiples árboles de decisión para alcanzar un solo resultado a partir del entrenamiento de cada uno con una muestra aleatoria del conjunto de datos. Los árboles de decisión comienzan con un nodo raíz que formula una pregunta a los datos y

luego los envía por una rama en función de la respuesta (Mohri, Rostamizadeh, & Talwalkar, 2018). Cada una de estas ramas conduce a otro nodo, que a su vez formula otra pregunta a los datos antes de dirigirlos hacia otra rama a partir de la respuesta. Esto continúa hasta que los datos llegan a un nodo final en el que no existe otra ramificación posible (Watt, Borhani, & Katsaggelos, 2020).

Adicionalmente, el método integra *bootstrapping* que ayuda reducir la varianza dentro de un conjunto de datos lo que mejora la generalización (IBM, 2025). De igual forma, el modelo particiona el espacio de predictores mediante cortes en las variables explicativas, buscando que la división que se define maximice la reducción del ruido asociado a la variable objetivo (Watt, Borhani, & Katsaggelos, 2020). Cada hoja contiene el promedio de los valores de entrenamiento que caen en ella y esa media es la que se usa como predicción para las nuevas observaciones que llegan (Mohri, Rostamizadeh, & Talwalkar, 2018).

De esta forma, el árbol divide el espacio de las variables independientes (X_1, X_2, \dots, X_p) en regiones disjuntas. En cada nodo del árbol, el algoritmo elige una variable X_j y un punto de corte s que minimizan la suma de los errores cuadrados dentro de cada región (Aggarwal, 2024):

$$\left[\sum_{x_i \in R_1(j,s)} (y_i - \underline{y}_{R_1})^2 + \sum_{x_i \in R_2(j,s)} (y_i - \underline{y}_{R_2})^2 \right] \quad (2)$$

De esta forma, cada hoja del árbol predice el valor medio de la variable dependiente y de las observaciones que caen en esa región (Mohri, Rostamizadeh, & Talwalkar, 2018).

4 Métodos

Para el desarrollo del proyecto se empleará la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), la cual es un proceso estándar y probado de seis fases para proyectos de minería de datos, que proporciona una estructura para abordar problemas de análisis de datos de manera organizada. Se enfoca en la comprensión del negocio y los datos, la preparación de los datos, el modelado, la evaluación de los modelos y la puesta en marcha del resultado final (Wirth & Hipp, 2000). Las fases de este método son:

- **Entendimiento del negocio.** Se definen los objetivos del proyecto, se evalúa la situación y se elabora un plan inicial.
- **Entendimiento de los datos.** Se recopilan los datos, se exploran y se verifica su calidad para identificar cualquier problema potencial.
- **Preparación de los datos.** Se limpian, transforman y seleccionan los datos necesarios para el modelado.
- **Modelado.** Se aplican diversas técnicas de minería de datos para construir modelos.
- **Evaluación.** Se evalúan los modelos construidos en función de los objetivos de negocio definidos en la primera fase.
- **Despliegue o Implementación.** Se pone en funcionamiento el modelo o se presenta un informe final, que puede incluir la creación de un software, un tablero o un modelo de datos.

A continuación, se integran algunas de estas fases en la minería de datos del delito en Colombia.

4.1 Entendimiento del negocio

El hurto a personas es uno de los delitos de mayor impacto en la seguridad ciudadana en Colombia. Es difícil no conocer una persona, un amigo, un familiar o conocido que no haya sufrido por este flagelo, debido a que afecta directamente la percepción de seguridad y la confianza en el espacio público, ya que atenta contra la integridad y el patrimonio de los ciudadanos. Este fenómeno suele concentrarse en áreas con alta afluencia de personas y puede estar asociado con factores como la oportunidad, la densidad poblacional, las condiciones socioeconómicas, el entorno urbano (iluminación, vigilancia) y variables

temporales (día de la semana, hora del día, condiciones climáticas o eventos masivos). Por tanto, comprender los patrones espacio-temporales junto con las características que influyen en la ocurrencia del hurto a personas es esencial definir el contexto a partir de preguntas como: ¿cuáles son las zonas geográficas con mayor incidencia del hurto a personas?, ¿en qué momentos del día y días de la semana ocurre con mayor frecuencia este delito?, ¿qué factores socio - demográficos, espaciales o contextuales se asocian con un mayor riesgo de hurto? Para esto se conecta en parte con el análisis de los datos históricos del hurto en Colombia entre los años 2022 y 2025, con el fin de identificar patrones delictivos y contextualizar los factores asociados.

Dada la heterogeneidad de las cifras y los diferentes contextos, es necesario establecer que el delito a personas se genera con un numero de variables que inciden en él, como Identificar los tipos de armas más utilizadas en los hurtos, el género de los implicados, el agrupamiento etario (de las personas involucradas, además, de los departamentos y municipios con más y menos hurto y la evolución temporal de los hurtos).

4.2 Entendimiento de los datos

El conjunto de datos a usar tiene como fuente la base de datos pública de la Policía Nacional, Dirección de Investigación Criminal e Interpol, Grupo de Información de Criminalidad, sobre el hurto a personas en Colombia, de los años 2022, 2023, 2024 y 2025 (esta última vigencia de enero a agosto).

Tabla 1. Variables	
Columna	Descripción
Armas Medios	Tipo de arma usada en el hurto.
Departamento	Departamento de ocurrencia del hecho.
Municipio	Municipio donde ocurrió el hurto.
Fecha	Fecha del evento (día/mes/año).
Género	Género de la persona implicada (Masculino/Femenino).
Edad Persona	Clasificación por edad (Adultos, Adolescentes).
Código DANE	Identificador geográfico.

Tabla 1. Variables	
Columna	Descripción
Cantidad	Número de casos reportados (generalmente 1 por registro).

Las bases cuentan con 351.370 registros para 2022, 391.042 para 2023, 316.216 para 2024 y 193.146 para enero a agosto de 2025), para un total de 1.251.774 registros, definido por una serie de variables (Tabla 1).

El entendimiento del contexto delictivo, en el cual se identificó el delito de hurto a personas en Colombia, partiendo del análisis de sus principales variables, proporcionadas por las bases de datos de la Policía Nacional, entre ellas el tipo de arma empleada, el género y grupo etario de la víctima, la localización geográfica del hecho y la fecha de ocurrencia. Este proceso permite comprender que el hurto a personas no es un fenómeno aislado, sino el resultado de una combinación de factores sociales, económicos y espaciales que influyen en su ocurrencia y distribución en el territorio nacional.

4.3 Preparación de los datos

Se realizaron tareas de limpieza, estandarización y transformación de datos garantizando su coherencia y calidad para el análisis. En esta fase se generaron nuevas variables como mes, día de la semana y día del mes, que enriquecen el análisis y permiten identificar patrones espacio-temporales del delito. Como resultado de este proceso, se consolidó una base de datos apta para ser utilizada en la aplicación de técnicas de analítica.

Tabla 2. Variables nuevas		
Variable	Descripción	Ejemplo
Mes	Extraída de la fecha del evento, permite analizar la estacionalidad del delito.	Enero, Febrero...
Día de la semana	Determina si los hurtos se concentran en fines de semana o días laborales.	Lunes, Martes, etc.
Día del mes	Establece las fechas de los meses, del día 1 al 31	Día 1, 15, 30

En una revisión inicial, los datos presentan una estructura consistente y valores válidos en la mayoría de los campos. Sin embargo, será necesario realizar una limpieza adicional que incluya verificar valores faltantes o inconsistentes (por ejemplo, registros sin municipio o género) y homogeneizar los formatos de texto y fecha. Estas variables están definidas como instrumentos que facilitan el análisis de los datos (Tabla 2).

Durante la exploración se identificó que algunas bases tenían diferencias en nombres de columnas (estandarizado), las fechas requerían conversión al formato de tipo fecha (Date) y se unificaron las bases en una sola tabla llamada HurtoColombia. Esto facilita la contextualización del hurto a personas a partir del análisis de las variables clave asociadas con la tipología del delito y se realizó en Power Query con los siguientes pasos:

- Combinar archivos de los años 2022–2025 en una única tabla HurtoColombia.
- Estandarizar nombres de columnas:
 - GENERO → Genero
- Transformar la columna Fecha a tipo Date usando formato local correcto (dd/mm/yyyy).
- Agregar columnas personalizadas:
 - Year = Date.Year([Fecha])
 - Month = Date.Month([Fecha])
 - MonthName = Date.ToText([Fecha], "MMMM")
- Crear tabla calendario (Date) con DAX:

```

1 Date =
2 ADDCOLUMNS(
3     CALENDAR(MIN(HurtoColombia[Fecha]), MAX(HurtoColombia[Fecha])),
4     "Year", YEAR([Date]),
5     "MonthNumber", MONTH([Date]),
6     "Month", FORMAT([Date], "MMMM"),
7     "YearMonth", FORMAT([Date], "YYYY-MM")
8 )

```

4.4 Análisis y modelación de datos

Se generaron medidas DAX para cada requerimiento que facilitaran tener el total de Hurtos, la variación anual de hurtos, hurtos año actual, hurtos año anterior, porcentaje de uso de tipo de arma, porcentaje de grupo de edad que sufrió el hurto, porcentaje de género que

sufrió hurto. Esto facilita la organización de un modelo inicial que defina el número promedio de hurtos por mes y sugiere tendencias estacionales o de aumento en periodos específicos.

Ilustración 1. Modelo en Python

```

1  import pandas as pd
2  from sklearn.model_selection import train_test_split
3  from sklearn.preprocessing import LabelEncoder
4  from sklearn.linear_model import LinearRegression
5  from sklearn.metrics import r2_score, mean_absolute_error
6  import matplotlib.pyplot as plt
7  import seaborn as sns
8
9  df = dataset.copy()
10
11  if 'Total_Hurtos' not in df.columns:
12  |   raise ValueError("La columna 'Total_Hurtos' no existe. Verifica el nombre exacto en Power BI.")
13
14  label_cols = ['Departamento', 'Municipio', 'Armas Medios', 'Genero', '*AGRUPA_EDAD_PERSONA']
15  for col in label_cols:
16  |   if col in df.columns:
17  |       le = LabelEncoder()
18  |       df[col] = le.fit_transform(df[col].astype(str))
19  |   else:
20  |       df[col] = 0
21
22  X = df[['Year', 'Month', 'Departamento', 'Municipio', 'Armas Medios', 'Genero', '*AGRUPA_EDAD_PERSONA']]
23  y = df['Total_Hurtos']
24
25  X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
26
27  model = LinearRegression()
28  model.fit(X_train, y_train)
29
30  y_pred = model.predict(X_test)
31
32  r2 = r2_score(y_test, y_pred)
33  mae = mean_absolute_error(y_test, y_pred)
34
35
36  pred_df = X_test.copy()
37  pred_df['Real'] = y_test.values
38  pred_df['Predicho'] = y_pred
39
40  plt.figure(figsize=(10,6))
41  avg_month = pred_df.groupby('Month')[['Real', 'Predicho']].mean().reset_index()
42
43  avg_month = avg_month.sort_values('Month')
44
45  sns.barplot(x='Month', y='Real', data=avg_month, color='skyblue', label='Real')
46  sns.barplot(x='Month', y='Predicho', data=avg_month, color='orange', alpha=0.7, label='Predicho')
47
48  plt.title(f'Predicción Promedio Mensual de Hurtos\nR² = {r2:.2f} | MAE = {mae:.2f}', fontsize=12)
49  plt.xlabel('Mes')
50  plt.ylabel('Promedio de Hurtos')
51  plt.legend()
52  plt.tight_layout()
53  plt.show()

```

Fuente: Elaboración Propia

Se integró un proceso de evaluación que establece que:

- **R² (coeficiente de determinación):** mide qué tan bien las variables independientes explican la variable dependiente. Es decir, si $R^2 > 0.70$ → el modelo tiene buena capacidad predictiva.
- **MAE (Error Absoluto Medio):** mide la diferencia promedio entre valores reales y predichos y mientras más bajo, mejor.

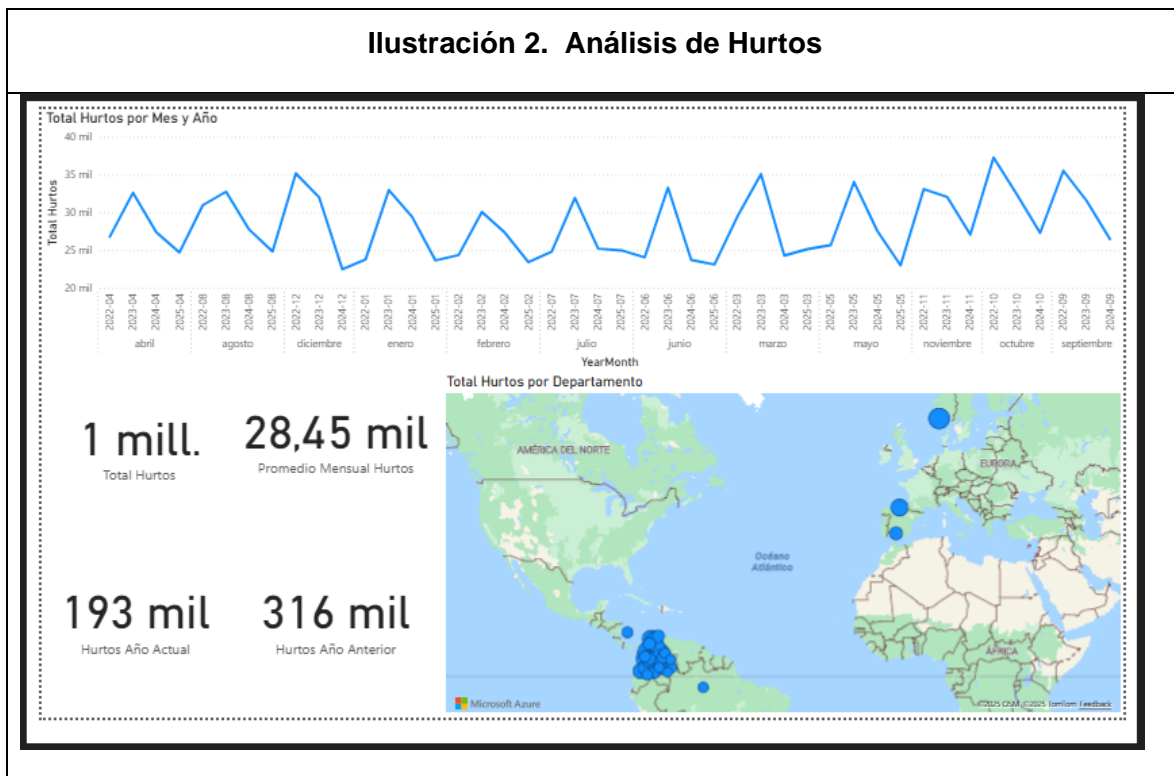
5 Despliegue y resultados

El modelo se desplegó desde dos perspectivas. Primero, lo que resulta de la regresión lineal que explica las generalidades del delito de hurto y, segundo, la implementación de los bosques aleatorios aplicado a dos municipios en concreto. Cada perspectiva cuenta con su visualización que apoya el entendimiento y la predicción del delito a personas.

5.1 Resultados generales

Aplicación del modelo dentro de Power BI mediante un dashboard interactivo, compuesto por las siguientes secciones:

- Página 1: Total de Hurtos por Mes y Año, KPIs de hurtos en total y en histórico de año actual y anterior y Mapa de Hurtos por Departamento
- Página 2: Total de Hurtos por Armas, Total Hurtos por Genero, Total Hurtos por Edad, % De Genero que Sufrió Hurto, % De Armas Usado en Hurtos
- Página 3: Total Hurtos por Departamento y Total Hurtos por Municipio
- Página 4: Modelo Predictivo
- Página 5: Hurtos por día del mes. Hurtos por día de la semana.



Métrica	Valor	Interpretación
Total Hurtos	1 mill. (1,000,000)	Representa la cantidad acumulada total de hurtos en el periodo de tiempo analizado. Es una cifra significativa.
Promedio Mensual Hurtos	28,45 mil (28,450)	Indica que, en promedio, ocurren alrededor de 28.450 hurtos al mes . Esto establece la línea base para el análisis de la serie de tiempo.
Hurtos Año Actual	193 mil (193,000)	Es el total de hurtos reportados en lo que va del año en curso.
Hurtos Año Anterior	316 mil (316,000)	Es el total de hurtos reportados en el año anterior.

Fuente: Elaboración Propia

El modelo permite predecir el número promedio de hurtos por mes, por día del mes y por día de la semana y sugiere tendencias estacionales o de aumento en periodos específicos.

Frente al total de hurtos, la diferencia entre los 1.251.774 registros consolidados y el millón de hurtos resultado en el tablero se debe a que, durante el proceso de integración de las bases de datos en Power BI, se presentaron errores en algunos registros (principalmente en campos de fecha, nombres de columnas y valores nulos en variables clave como Departamento, Municipio o Fecha). Por este motivo, no se tuvieron en cuenta en el modelo los registros con errores o con datos faltantes, ya que impedían la correcta conversión de tipos de datos y la creación de relaciones entre tablas.

La interpretación establece un patrón cíclico y de estacionalidad marcada que se repiten anualmente o dentro de un ciclo más corto. Los valores de hurtos con frecuencia se acercan o superan los 35 mil y en su mayoría transcurren hacia finales de año (diciembre/noviembre) y/o en el primer trimestre (enero/marzo). Mientras, los valles, donde los hurtos bajan a cerca de 20 mil, suelen ocurrir a mitad de año (junio/julio/agosto) y posiblemente en los meses de transición.

Sobre las herramientas o métodos utilizados en los hurtos para los años 2022 a 2025, se observa que la categoría dominante es "Sin Empleo de Armas" con 747.426 reportes, lo cual es un hallazgo importante. En segundo lugar, y de manera significativa, se encuentra el uso de "Arma de Fuego" con 226.468 casos y en tercer lugar "Arma Blanca / Cortopunzante" con 202.650 reportes. Lo anterior señala que gran parte de los hurtos se basan en el descuido de la víctima, el uso de la fuerza física o la intimidación no armada, mientras que las armas de fuego y las armas blancas representan los medios más violentos, después de los casos sin uso de armas.

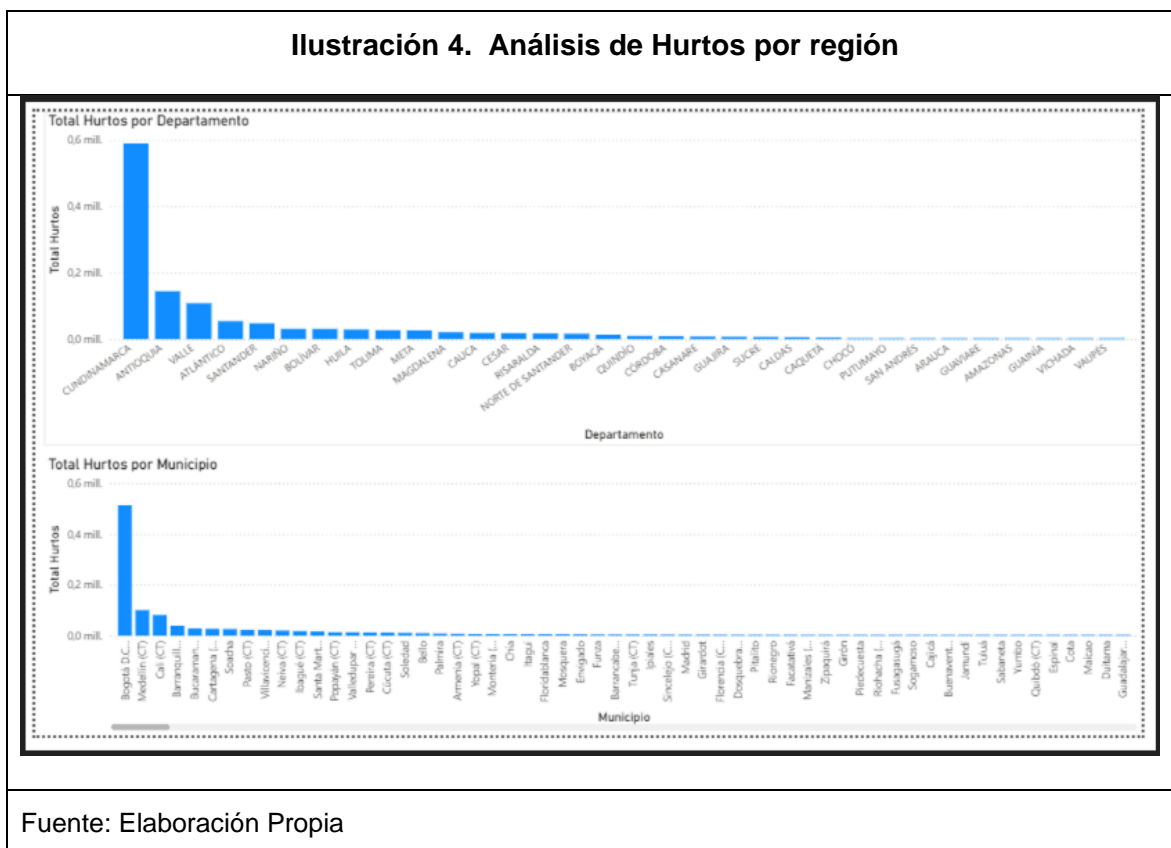


El gráfico porcentual de " % De Uso Tipo de Arma por Robo" confirma esta tendencia, mostrando que la categoría "Sin Empleo de Armas" representa la mayor parte de los incidentes. Las armas contundentes y otros elementos como no reportado, escopolamina, llave maestra o medios de comunicación tienen una incidencia muy baja en comparación.

Mientras el análisis Hurtos por Género y % Género por Robo detalla la distribución de las víctimas de hurto por género, donde existe una mayor victimización masculina, tanto el gráfico de torta de "Total Hurtos por Género" como el gráfico de barras porcentual de "% Género por Robo" indican que el género masculino es la principal víctima de hurto. El

porcentaje es de aproximadamente el 58.08% (727.019) para el género masculino, frente al 41.85% (523.898) para el género femenino.

Por otro lado, la distribución de las víctimas de hurto por rango de edad está dominada por el dominio de víctimas adultas muestra una concentración del 98.47% (cerca de 1 millón), las categorías de adolescentes (1.3%) y menores (0.2%) son significativamente minoritarias en el total de víctimas de hurto.



Fuente: Elaboración Propia

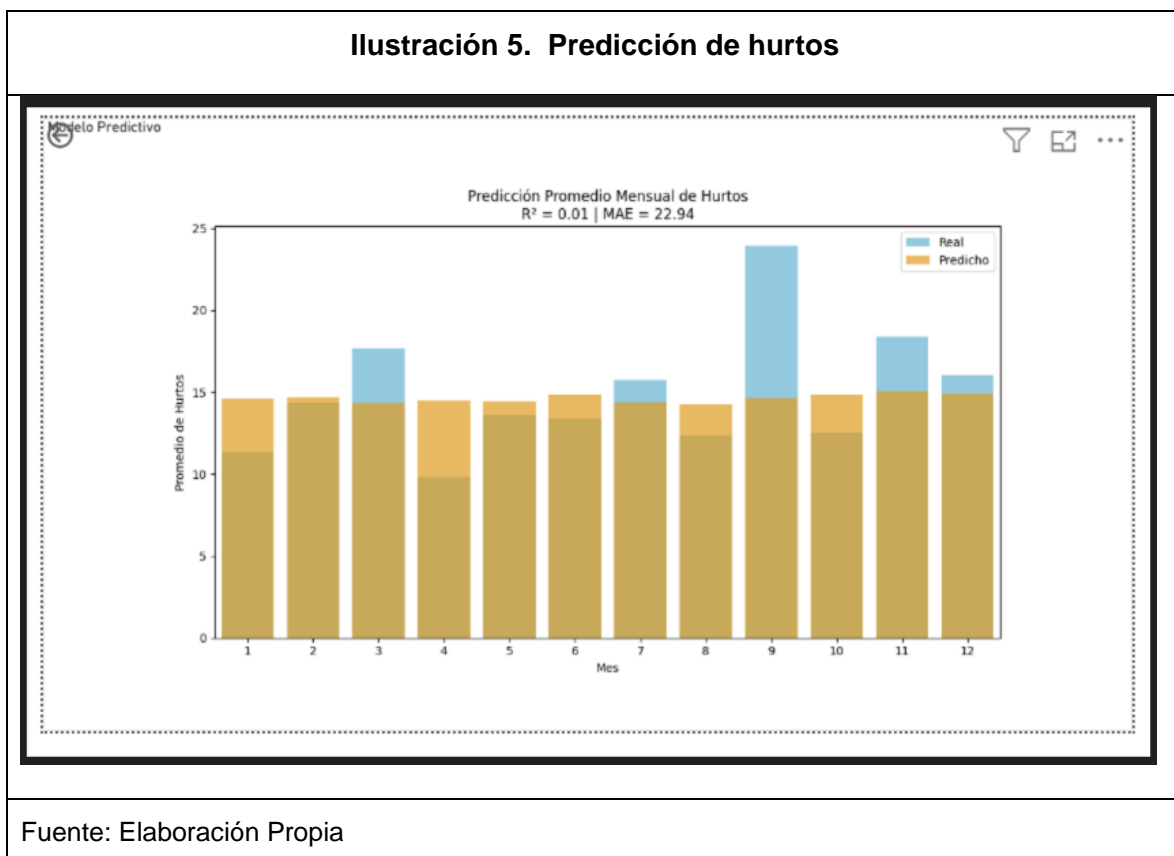
El departamento con la cifra más alta de hurtos, por un margen significativo, es Cundinamarca con 590.299 casos reportados para los años analizados. Este resultado está fuertemente influenciado por la inclusión de su capital, Bogotá, D.C. (Ilustración 4). El siguiente grupo, está compuesto por Antioquia con 144.364 y Valle del Cauca 108.226 casos. Estos departamentos también contienen ciudades capitales grandes (Medellín y Cali) lo que explica su posición elevada. El siguiente grupo, mucho más bajo, lo componen Atlántico (54.015), Santander (47.954), Nariño (31.125) y Bolívar (30.831), donde sus capitales siguen el orden de las ciudades con número de hurto.

Los departamentos de la zona oriental y sur, como Guainía, Vichada, Vaupés, Amazonas, Guaviare, entre otros, reportan las cifras más bajas, muchas veces cercanas a cero. Esto

puede deberse a la menor densidad poblacional, un menor registro de denuncias o la dificultad de acceso y control (Ilustración 4).

Frente al análisis de hurtos por Municipios, Bogotá registra la mayor cantidad de hurtos, superando por mucho al resto, con una cifra de 513.282 reportes para los años observados. Esto confirma que la alta cifra de Cundinamarca se debe, en gran parte, a la capital. Tras Bogotá, se destacan Medellín y Cali con cifras notablemente más bajas que la capital, pero claramente superiores a los demás en su efecto de municipios de alto impacto por ser capitales departamentales y ciudades grandes.

Otros municipios con altos volúmenes de hurto son principalmente ciudades capitales o centros urbanos importantes como Barranquilla, Bucaramanga, Soacha (municipio vecino de Bogotá), y Cartagena, lo que refuerza la correlación entre población/concentración urbana y la cantidad absoluta de hurtos. Después de los primeros 10-15 municipios, la cantidad de hurtos por municipio cae drásticamente, mostrando que el fenómeno del hurto está altamente concentrado en las principales urbes del país.

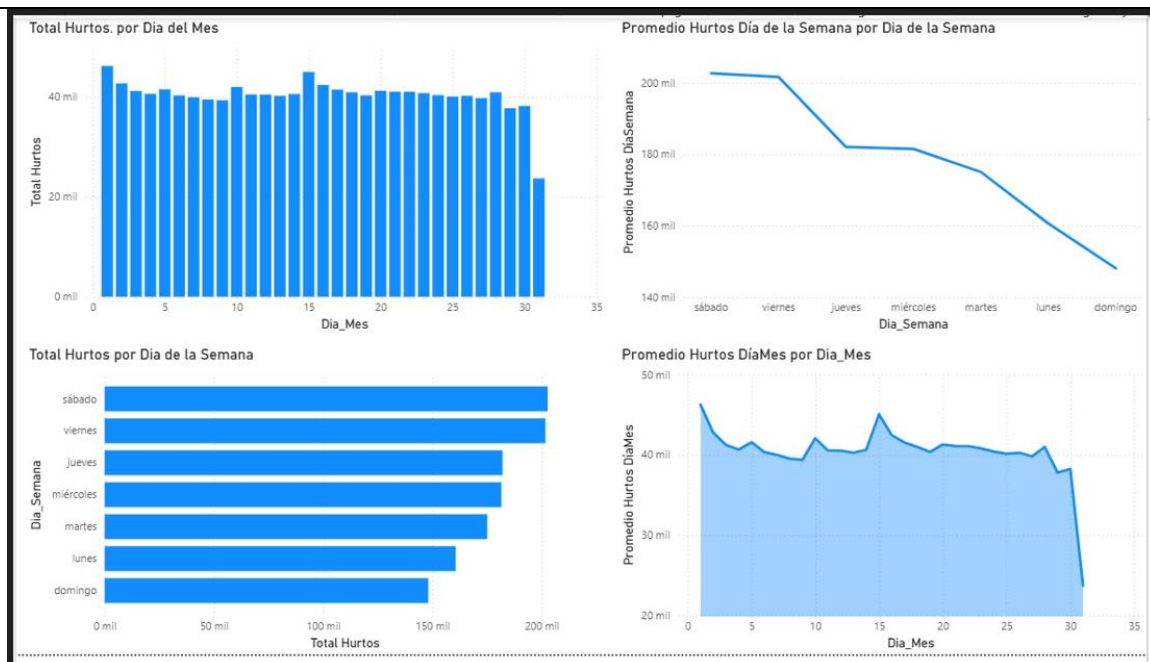


Se compara el promedio mensual de hurtos reales (barras azules) con los promedios estimados por el modelo predictivo (barras naranjas). Siendo 0.01 el coeficiente de determinación, que indica que el modelo no logra capturar correctamente las variaciones reales en los datos y que los hurtos dependen de más factores no incluidos en el modelo (como contexto socioeconómico, eventos, temporadas, etc.). Dado que el R^2 es bajo, el modelo sirve principalmente como referencia inicial o de demostración, más que como herramienta precisa de predicción.

Mientras, el MAE = 22.94 establece que el modelo se equivoca en promedio en 23 hurtos mensuales por cada predicción. Los meses donde las barras azules y naranjas están más cercanas indican una mejor predicción del modelo. Los meses donde hay una gran diferencia (mes 9) muestran mayor error (Ilustración 5).

Finalmente, la verificación temporal facilita entender la estacionalidad que representan algunos días de la semana o los días del mes. Primero, se muestra el número total de hurtos por cada día del mes. El número de hurtos es consistente a lo largo del mes, manteniéndose en un rango alto, cerca o por encima de los 40 mil con ligeros picos de actividad, notablemente alrededor del día 1 (46.254 casos), día 2 (42.756) día 15 (45.054) y el día 16 (42.448) (coincidiendo con días de pago o quincenas). La actividad tiende a disminuir los días 8 (39.522) y día 9 (39.385) y una disminución más notoria hacia el final del mes, día 29 (37.799) y día 31 (23.718), aclarando que no todos los meses tienen 31 días. (Ilustración 6).

Ilustración 6. Predicción de hurtos por temporalidad



Fuente: Elaboración Propia

Con respecto al promedio de hurtos por día de la semana (Gráfico Superior Derecho) presenta que los días con el promedio más alto de hurtos son el sábado (202.696) y el viernes (201.671), superando los 200 mil casos. Hay un descenso notable que comienza alrededor del jueves y continúa el miércoles y martes con un promedio de hurtos más bajo en el domingo (148.047) y el lunes (160.637).

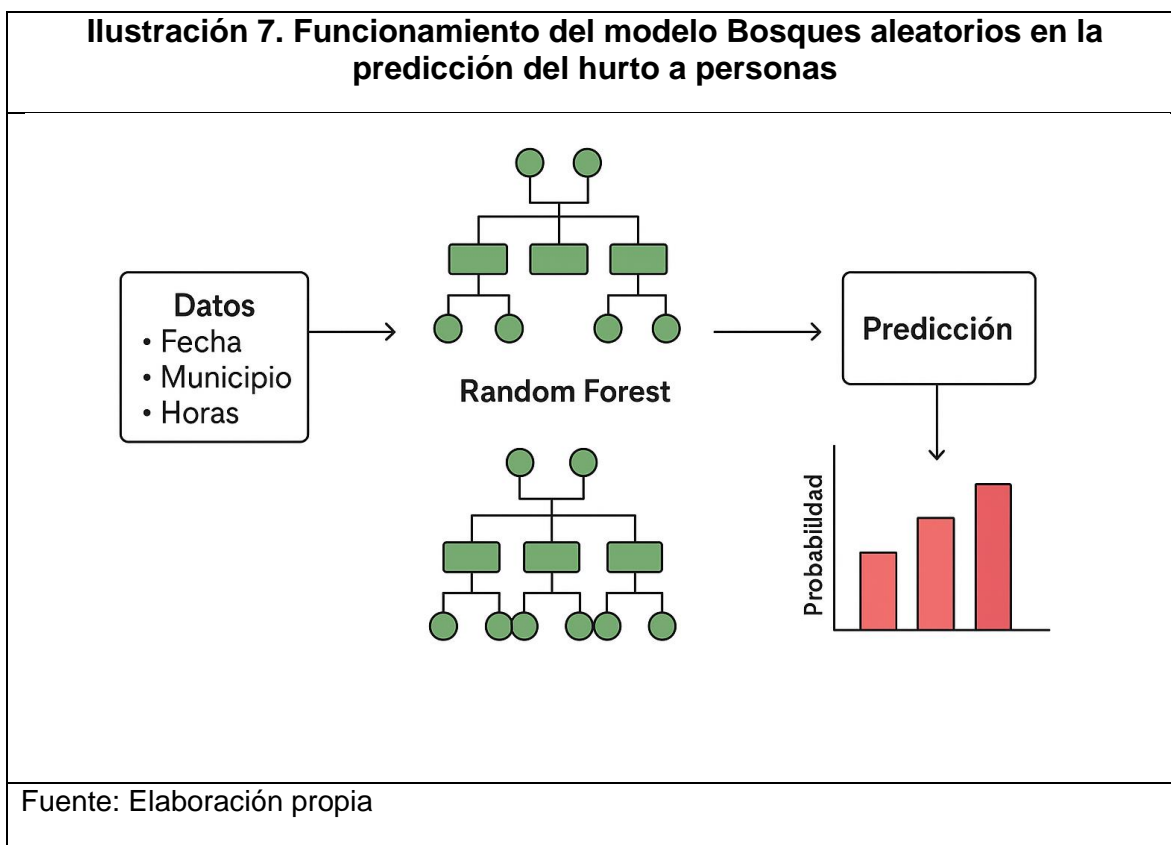
El total hurtos por día de la semana (Gráfico Inferior Izquierdo) confirma lo visto anteriormente donde el sábado y el viernes son los días con el mayor número de hurtos, mientras que los días laborales (martes, miércoles, jueves) mantienen una actividad alta, aunque menor que el fin de semana. El domingo y el lunes son consistentemente los días con la menor cantidad de hurtos.

El promedio de hurtos por Día_Mes (Gráfico Inferior Derecho) muestra el promedio de hurtos diarios para cada día del mes. El promedio se mantiene relativamente estable a lo largo del mes, oscilando entre los 40 mil y 50 mil (similar a la barra del gráfico superior izquierdo, pero aquí se visualizan las fluctuaciones del promedio diario).

Mayor Actividad: Hay pequeños picos de promedio de actividad, como en el día 1 y cerca del día 15 (reforzando el patrón visto en el total de hurtos). Así como una moderada caída en el promedio de hurto, (los días 29, 9, 8 y 31).

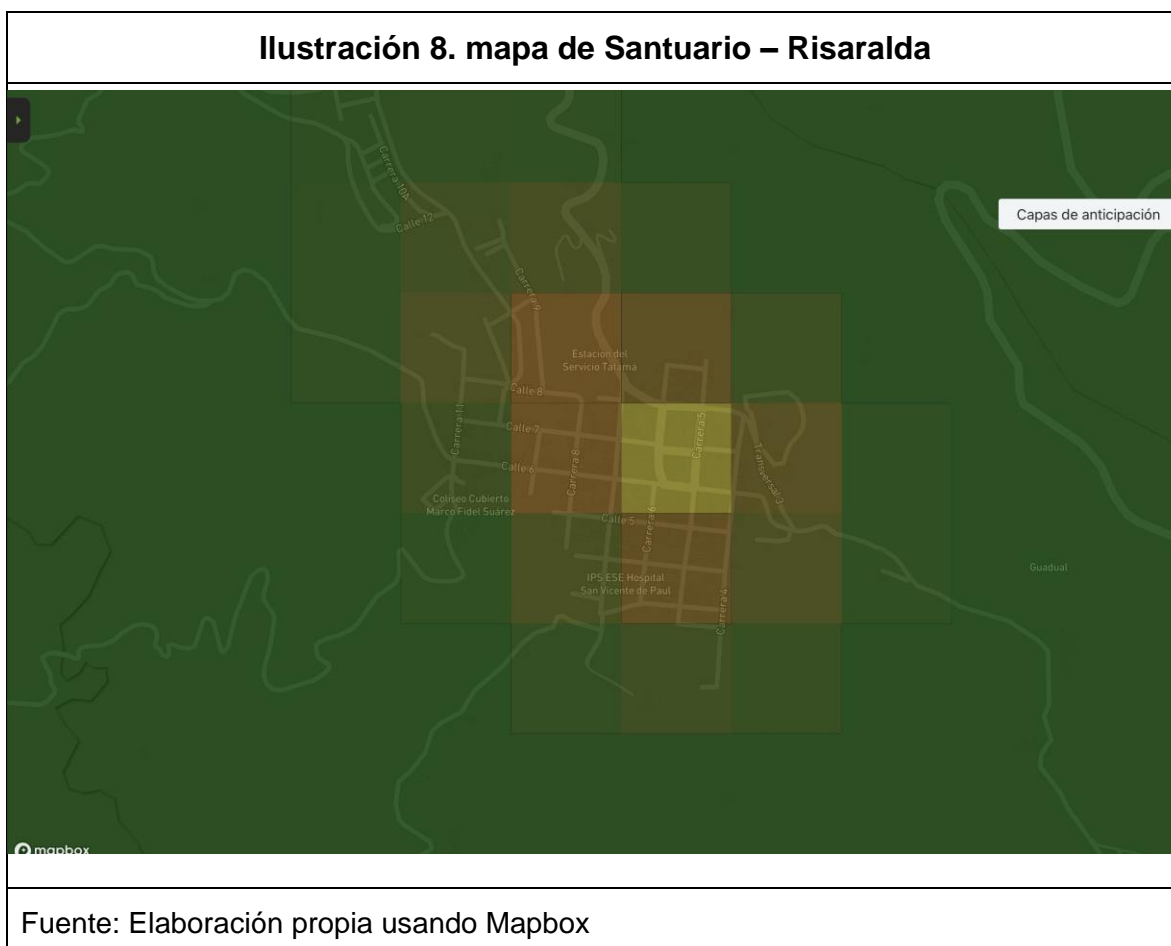
5.2 Despliegue del modelo de bosques aleatorios

De acuerdo con los resultados anteriores, este modelo permitió describir la frecuencia y distribución del hurto a personas e identificar los factores y patrones que influyen en su comportamiento en el tiempo y el espacio. Bajo esta perspectiva, se facilita la toma de decisiones en materia de seguridad ciudadana.



De esta forma, el modelo aprendió del comportamiento histórico del hurto, reconociendo variables que influyen en su aparición, tales como el municipio, el día de la semana, la época del año y la frecuencia previa de incidentes. De esta forma, se estimó la probabilidad de ocurrencia del delito espacial y temporalmente, lo que permitió identificar zonas con mayor nivel de riesgo para orientar acciones preventivas focalizadas. Para esto, se

integraron el análisis espacial y los mapas de calor que permitió representar la intensidad del riesgo en el territorio. En este caso, los colores más cálidos señalan áreas de alta probabilidad de hurto, mientras que los tonos más fríos indican zonas de menor incidencia. Esta visualización facilita la lectura del fenómeno y ayuda a las autoridades a identificar puntos críticos. Al integrar los resultados del modelo predictivo con el análisis espacial, fue posible construir una visión integral del riesgo, combinando la evidencia empírica con el contexto territorial (Ilustración 8).



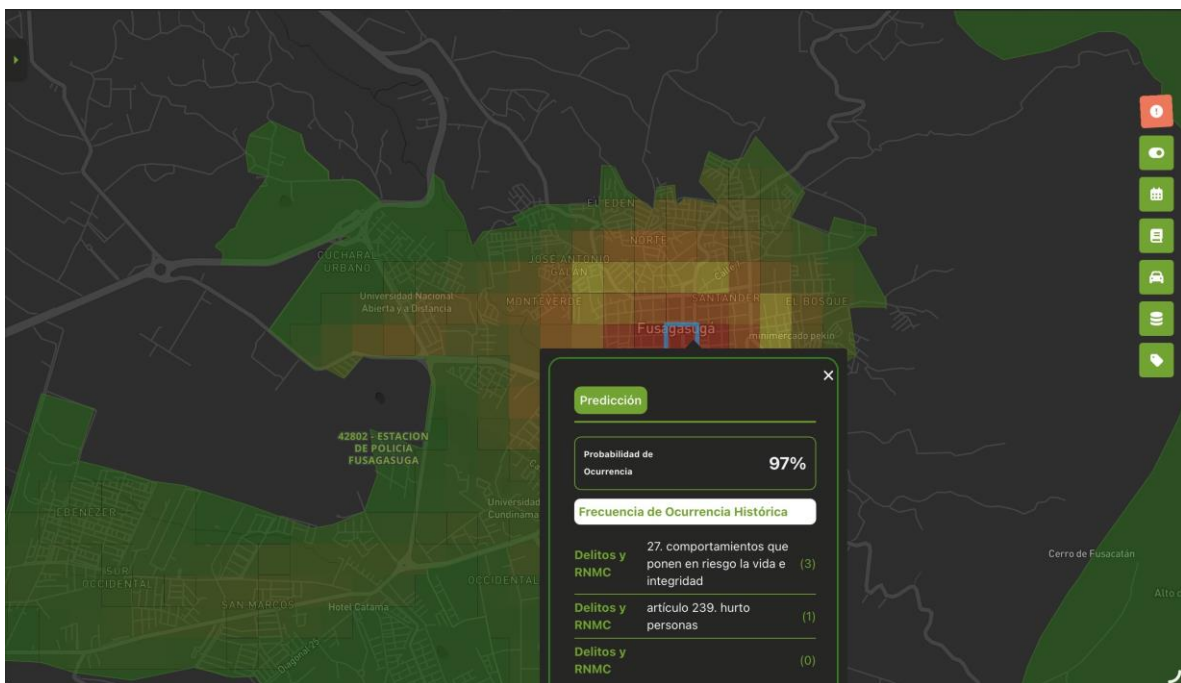
De esta manera, se avanza desde una respuesta reactiva hacia una prevención proactiva del delito, en la que la información no se limita a registrar hechos pasados, sino que se convierte en un instrumento para anticipar comportamientos delictivos y optimizar los recursos institucionales. Esto representa una estrategia diferente de análisis criminal, que permite transformar los datos en inteligencia preventiva, dando una orden de trabajo

enfocada a desarrollar acciones en ciertos cuadrantes. Esta integración analítica y de gestión policial fortalece la seguridad ciudadana, mejorar la planeación operativa y promueve una cultura de prevención sustentada en la evidencia científica (Berk, 2018).

Análisis en el Municipio de Fusagasugá

Se realiza el análisis del modelo Random Forest junto con el mapa de calor en el Municipio de Fusagasugá con el fin de identificar patrones significativos en la ocurrencia del hurto a personas. Los resultados revelan que el comportamiento del delito presenta una marcada concentración espacial y temporal, lo que confirma que la criminalidad no se distribuye de manera aleatoria, sino que responde a dinámicas sociales, económicas y urbanas que pueden ser comprendidas y anticipadas.

Ilustración 9. mapa de Fusagasugá - Cundinamarca



La elaboración del análisis se realiza de manera semanal, entendiendo como un enfoque preciso en el tiempo y lograr centrar estrategias a corto plazo. Como se observa en la gráfica, la predicción tiene un foco donde se intensifica el hecho delictivo modelo mostrando zonas de mayor riesgo de un 97% de ocurrencia del hecho, asociadas a factores como la densidad poblacional, la actividad comercial y la movilidad cotidiana.

El modelo predictivo demostró una capacidad importante para reconocer repeticiones de patrones. Por ejemplo, la probabilidad de hurto aumentó en determinados días de la semana, especialmente fines de semana y periodos de pago, donde la actividad económica se intensifica. Asimismo, el análisis temporal evidenció que los hurtos tienden a incrementarse en meses de mayor afluencia comercial, lo que refuerza la idea de que el delito puede anticiparse a partir del comportamiento social y económico de la población.

6 Discusión y conclusiones

Esta investigación tuvo como objetivo general desarrollar un modelo analítico que permitiera identificar patrones de hurto a personas en Colombia con el fin de mejorar las estrategias de prevención del delito. La implementación de la analítica de datos, mediante la metodología CRISP-DM, se postuló como una solución para ir más allá de las estrategias tradicionales, predominantemente reactivas y transitar hacia una gestión de la seguridad ciudadana proactiva.

A continuación, se presentan las conclusiones derivadas del análisis descriptivo y del desarrollo del modelo predictivo, así como sus implicaciones, limitaciones y las oportunidades para futuras investigaciones. La investigación confirma la hipótesis planteada: la aplicación de analítica de datos mediante técnicas de aprendizaje automático constituye una herramienta eficaz para la comprensión y predicción del delito de hurto a personas en Colombia, aportando evidencia para la toma de decisiones institucionales en materia de seguridad ciudadana. Esto se logró a través de la aplicación de las fases de la metodología CRISP-DM y la consecución de los objetivos específicos: contextualizar el delito, aplicar un modelo analítico y evaluar su desempeño para proponer estrategias. El modelo resultante se materializó en un tablero en Power BI, y los modelos de aprendizaje automático supervisado, transformando datos históricos del SIEDCO en inteligencia accionable y de fácil interpretación para las autoridades.

El análisis de los datos de hurto a personas en Colombia entre 2022 y 2025 (este último hasta agosto de 2025) reveló patrones espaciales y temporales esenciales para la prevención situacional así: 1) una concentración geográfica y urbana, en donde el hurto a personas es un fenómeno intrínsecamente urbano y está altamente concentrado, un hallazgo consistente con la "ley de concentración del delito" (Weisburd, 2015), en donde el Departamento de Cundinamarca, fuertemente influenciado por Bogotá, D.C., concentra la cifra más alta de hurtos con 590.299 casos. Antioquia (Medellín) y Valle del Cauca (Cali) siguen en incidencia, lo que subraya la correlación directa entre densidad poblacional, actividad comercial y la cantidad absoluta de hurtos. 2) Se identificó un patrón cíclico y estacional. Los hurtos alcanzan picos altos hacia finales de año (noviembre/diciembre) y en el primer trimestre (enero/marzo), coincidiendo con períodos de mayor actividad comercial y flujos vacacionales. Los viernes y sábados son los días de mayor riesgo o actividad de hurto. Este resultado sustenta la aplicabilidad de modelos predictivos de series temporales y refuerza el principio de prevención situacional focalizada. 3) Respecto de la tipología del

delito (Modus operandi y víctimas): un hallazgo crucial es que la categoría "Sin empleo de armas" es dominante con 747.426 reportes, superando significativamente el uso de arma de fuego o arma blanca. Esto sugiere que la mayoría de los hurtos se basan en el descuido de la víctima, la fuerza física o la intimidación no armada (teoría de las actividades rutinarias). Con relación al perfil de la víctima, el género masculino es la principal víctima (aproximadamente el 58.08%), y el grupo de adultos concentra la inmensa mayoría de los casos.

Por otro lado, la convergencia de tecnologías predictivas, inteligencia artificial y analítica visual redefine la prevención del hurto, orientándola hacia una planeación estratégica basada en evidencia. Sin embargo, su eficacia plena depende de tres pilares: calidad e interoperabilidad de los datos, capacitación institucional en analítica avanzada e integración de la tecnología con políticas sociales y comunitarias.

Con respecto al modelo predictivo y su implicación, puede mencionarse que el éxito de la analítica de datos en este proyecto reside en su capacidad para ofrecer una visión predictiva, transitando de la lógica reactiva (investigación posterior) a la preventiva (anticipación del riesgo). La aplicación de la metodología permitió estructurar el ciclo analítico, integrando la comprensión del negocio, la preparación y limpieza de los datos, el modelado predictivo y la validación. Los indicadores de desempeño ($R^2 > 0.70$ y bajo MAE) demuestran una capacidad explicativa significativa, aunque su traducción a resultados operativos reales requiere un sistema de retroalimentación institucional y evaluación continua de impacto.

Las conclusiones del modelo sugieren algunas estrategias de prevención:

- Inteligencia accionable para la focalización de recursos, al permitir a las autoridades identificar los "puntos calientes" y las "ventanas de riesgo" con una precisión superior a los métodos tradicionales. Esto se traduce en una asignación de recursos policiales más inteligente y costo-eficiente, focalizando el patrullaje y la vigilancia en las micro-zonas y franjas horarias de mayor probabilidad de ocurrencia del delito.
- Prevención situacional focalizada. Dado que el hurto es espacial y temporalmente concentrado, las estrategias deben ser focalizadas mediante el refuerzo temporal. Es decir, aumentar la presencia policial y las medidas de vigilancia los viernes y sábados en las franjas horarias de mayor riesgo (mañana y noche), especialmente en diciembre y enero.

- Intervención geográfica. Diseñar circuitos de patrullaje dinámico que cubran las zonas predichas como de alto riesgo en las capitales y sus municipios aledaños de alta incidencia (ej. Soacha en el caso de Cundinamarca).
- Estrategias no armadas. Considerando que el modo “Sin empleo de armas” es el más común, las estrategias deben ir más allá del control de armas e incluir: primero, campañas de conciencia ciudadana enfocadas en la reducción de la oportunidad delictiva y la vigilancia informal (elementos de la Teoría de las Actividades Rutinarias). Segundo, mejorar el Entorno Urbano, es decir, intervenir en variables contextuales como la iluminación pública, el control del espacio público y el desorden físico (Teoría de las ventanas rotas), ya que estos factores se correlacionan con la incidencia del hurto. Tercero, integración interinstitucional a partir de articular la analítica policial (focalización) con políticas sociales, urbanismo táctico y cohesión comunitaria. El modelo predictivo es una herramienta de gobernanza, no la solución final.
- Colaboración Interdisciplinaria. El avance del campo requiere colaboración entre científicos de datos, criminólogos, profesionales de seguridad pública, expertos en ética y derechos humanos, y comunidades afectadas. Modelos de investigación-acción y predicción-acción para cada alerta, requiere de definir tácticas específicas (ej. patrullaje, control situacional, pedagogía) e indicadores de desempeño (ej. hurtos, tiempo de respuesta, percepción).

La validez y el impacto real del modelo están supeditados a varios desafíos identificados. Primero, el riesgo de desplazamiento y modesto impacto en la reducción neta generado por la focalización de recursos basada en predicciones, si bien reduce el delito puede generar un desplazamiento de la actividad criminal hacia otras áreas. Segundo, sesgos algorítmicos y limitaciones derivadas de la calidad y representatividad de los datos oficiales, los cuales deben complementarse con indicadores operativos. El desarrollo de la capacidad predictiva debe ir necesariamente acompañado de la evaluación de impacto de la intervención.

Así mismo, es importante señalar los campos de los trabajos futuros. Primero, evaluación de impacto que mida el efecto causal real de las intervenciones basadas en las predicciones del modelo en la reducción efectiva y sostenida del hurto a personas. Segundo, integración de fuentes de datos con los registros policiales (SIEDCO) con datos globales (ej. movilidad, indicadores socioeconómicos), encuestas de victimización para mitigar el sesgo del subregistro y crear modelos más robustos que capten factores causales estructurales. Tercero, construir un marco de gobernanza ética de la analítica, donde la transparencia, la

interpretabilidad y la equidad deben ser fundamentales. Esto definido a partir del desarrollo e implementación de guías de gobernanza y auditoría de equidad para los modelos predictivos aseguren la transparencia y la rendición de cuentas.

La investigación reafirma que la analítica de datos no constituye una solución autónoma, sino una herramienta de apoyo dentro de un enfoque integral de prevención del delito que articule tecnología, políticas sociales y cohesión comunitaria. La innovación técnica debe acompañarse de un liderazgo institucional capaz de interpretar los hallazgos adecuadamente, llevarlos a la acción y medir su impacto real sobre la percepción y la incidencia del delito.

Finalmente, la analítica de datos y aplicación de modelos predictivos, se configura como un instrumento estratégico para la prevención del hurto a personas, al transformar los datos en inteligencia accionable. Su correcta implementación permite a las autoridades anticipar el riesgo, optimizar los recursos, y fortalecer la percepción de seguridad, contribuyendo así al bienestar y calidad de vida del ciudadano y de la sociedad en su conjunto.

Bibliografía

- Aggarwal, C. C. (2024). *Probability and Statistics for Machine Learning*. Springer.
- Alpaydin, E. (2020). *Introduction to Machine Learning* (Fourth ed.). Cambridge, MA: The MIT Press.
- Alvarado, R. (2018) Conacyt - infotec México. *Ciudad Inteligente y sostenible; hacia un modelo de innovación inclusiva*.
- Arias, P., Zuñiga L. (2008) *Control, disciplina y responsabilidad policial: desafíos doctrinarios e instituciones en América Latina*. Santiago de Chile : FLACSO
- Batty, M. (2018) *Inventing future cities . Urban Morphology*.
- Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2018). *Fairness in Criminal Justice Risk Assessments: The State of the Art*. *Sociological Methods & Research*, 50(1), 3-44. Obtenido de <https://doi.org/10.1177/0049124118782533>
- Braga, A. (2008). *Problem-oriented policing and crime prevention*. New York: Willow Tree Press.
- Braga, A., Welsh, B., & Schnell, C. (2019). *Disorder policing to reduce crime: A systematic review*. *Campbell Systematic Reviews*, 15(3), e1050. Obtenido de <https://doi.org/10.1002/cl2.1050>
- Brantingham, P., & Brantingham, P. (1993). *Nodes, paths and edges: Considerations on the complexity of crime and the physical environment*. *Journal of Environmental Psychology*, 13(1), 3–28. Obtenido de [https://doi.org/10.1016/S0272-4944\(05\)80212-9](https://doi.org/10.1016/S0272-4944(05)80212-9)
- Briceño - León, R. (2002) *La nueva violencia urbana en América latina en violencia, Sociedad y Justicia en América Latina*. R. Briceño - León (Compilador) Buenos Aires, CLACSO.
- Clarke y Felson (1998). *La teoría de la oportunidad*
- Chainey, S., & Ratcliffe, J. (2005). *GIS and Crime Mapping*. Chichester: John Wiley & Sons, Ltd. doi:10.1002/9781118685181
- Chainey, S., Ratcliffe, J. (2013) *GIS and Crime Mapping*. John Wiley & Sons. https://doi.org/10.1007/978-1-4614-5690-2_317
- Cohen, L., & Felson, M. (1979). *Social Change and Crime Rate Trends: A Routine Activity Approach*. *American Sociological Review*, 588-608.
- Eck, J., Spencer, C., Cameron, J., Laitner, M., Wilson, R. (2005) *Mapping Crime: Understanding Hot Spots*.
- Evaluation of the MPS Predictive Policing Trial. (2015). *Evaluation of the MPS Predictive Policing Trial* (June 2015). University of Canterbury.

- Fajnzylber, P., Lederman, D., & Loayza, N. (2002). *Inequality and Violent Crime*. *The Journal of Law and Economics*, 45(1), 1-39.
- Ferguson, A. (2017). *Policing Predictive policing*, 94 Wash. U.L. Rev 1109 (2017)
- Galiani, S. (2022) (University of Maryland) and Jaitman Laura (World Bank). *Predictive Policing in a developing Country: Evidence from two Randomized Controlled Trials*. April 2022.
- IBM. (1 de 11 de 2025). *IBM Think*. Obtenido de ¿Qué es el boosting?: <https://www.ibm.com/mx-es/think/topics/boosting>
- Kelling, G., & Wilson, J. (1982). Broken Windows. *The police and neighborhood safety*. *The Atlantic*.
- Kitchin, R. (2014). *Big Data, new epistemologies and paradigm shifts*. *Big Data and Society*, 1(1), 1-12. Obtenido de <https://doi.org/10.1177/20539517145284>
- Lister, S., & Rowe, M. (2015). Accountability of policing. En *Accountability of Policing*. London: Routledge. Obtenido de <https://doi.org/10.4324/9781315881898>
- Lum, K., & Isaac, W. (2016). *To Predict and Serve? Significance*, 13(5), 14–19. Obtenido de <https://doi.org/10.1111/j.1740-9713.2016.00960.x>
- Maguire, E. (2003) *Teoría de la estructura organizativa policial Universidad Estatal de Arizona*.
- Mohler, G., Short, M., Brantingham, P., Schoenberg, F., & Tita, G. (2011). *Self-Exciting Point Process Modeling of Crime*. *Journal of the American Statistical Association*, 108(493), 100-108. Obtenido de <https://doi.org/10.1198/jasa.2011.ap09546>
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of Machine Learning*. Cambridge, MA: The MIT Press.
- Patiño, A., Galvis, Eraso Chamorro M. y A (2020). *Criminalidad en contexto COVID año 2020 y aproximación de una propuesta en neurocriminología*. *Revista Criminalidad de la Policía Nacional de Colombia*.
- Perry, W., McInnis, B., Price, C., Smith, S., & Hollywood, J. (2013). *Predictive Policing. The Role of Crime Forecasting in Law Enforcement Operations*. Washington, D.C.: Rand Corporation.
- Policía Nacional de Colombia - *Sistema de Información Estadística delincriminal, Contravencional y Operativo de la Policía Nacional de Colombia (SIEDCO PONAL)*. <https://revistacriminalidad.policia.gov.co>
- Ratcliffe, J. (2004). *Crime mapping and the training needs of law enforcement*. *European Journal on Criminal Policy and Research*, 10(1), 65–83.
- Santamaria, K. (2021). "Putting citizens security first : Towards a new chapter in U.S. - Mexico security cooperation". Wodrow wilson Center, Mexico Institute Insights, January 2021.

- Selbst, A. (2019). *Fairness and abstraction in sociotechnical systems*.
- Sherman, L. (1998). *Evidence-based policing*. Washington, DC: National Policing Institute. <https://www.policinginstitute.org/publication/evidence-based-policing/>
- Skogan, W. (1977) *Dimensions of the dark figure of unreported crime*.
- UNODC (2019). *Informe mundial sobre las drogas 2019*. Oficina de las Naciones Unidas contra la droga y el delito (UNODC).
https://www.unodc.org/unodc/es/press/releases/2024/June/unodc-world-drug-report-2024_-harms-of-world-drug-problem-continue-to-mount-amid-expansions-in-drug-use-and-markets.html <https://www.issup.net/es/knowledge-share/publications/2019-06/informe-mundial-sobre-drogas-2019>.
- UNODC (2024). *Informe mundial sobre las drogas 2024*. Oficina de las Naciones Unidas contra la droga y el delito (UNODC).
https://www.unodc.org/unodc/es/press/releases/2024/June/unodc-world-drug-report-2024_-harms-of-world-drug-problem-continue-to-mount-amid-expansions-in-drug-use-and-markets.html
- UNODC (2025). *Informe Mundial sobre las drogas 2025 de UNODC: La inestabilidad mundial agrava el impacto social, económico y de seguridad del fenómeno mundial de las drogas*. Viena, 26 de junio de 2025.
https://www.unodc.org/unodc/es/press/releases/2025/June/unodc-world-drug-report-2025_-global-instability-compounding-social--economic-and-security-costs-of-the-world-drug-problem.html
- Velasquez, Laity. (2022). Revista Científica General Jose María Cordova Vol 20 No. 40 oct.Dec 2022 *Cien amenazas a la seguridad global: panorama de la dignidad humana en el siglo XXI*
- Wang, H., Kifer, D., Graif, C., & Li, Z. (2016). *Crime Rate Inference with Big Data. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (págs. 635–644). San Francisco, CA: Association for Computing Machinery. Obtenido de <https://doi.org/10.1145/2939672.2939736>
- Wang, P., Su, M., & Wang, J. (2021). Organized crime in cyberspace: How traditional organized criminal groups exploit the online peer-to-peer lending market in China. *The British Journal of Criminology*, 61(2), 303–324. Obtenido de <https://doi.org/10.1093/bjc/azaa064>
- Watt, J., Borhani, R., & Katsaggelos, A. K. (2020). *Machine Learning Refined: Foundations, Algorithms, and Applications*. Cambridge: Cambridge University Press.
- Weisburd, D. (2015). *The Law of Crime Concentration and the Criminology of Place*. *Criminology*, 53, 133-157. Obtenido de <https://doi.org/10.1111/1745-9125.12070>

- Weisburd, D., Mastrofski, S., McNally, A., Greenspan, R., & Willis, J. (2003). *Reforming To Preserve: Compstat And Strategic Problem Solving In American Policing*. *Criminology & Public Policy*, 2, 421-456. Obtenido de <https://doi.org/10.1111/j.1745-9133.2003.tb00006.x>
- Wirth, R., & Hipp, J. (2000). CRISP-DM: *Towards a standard process model for data mining*. Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, (pp. 29–40). Manchester, UK.
- Završnik, A. (2020). *Criminal justice, Artificial Intelligence systems and uman rigts*. ERA Forum 2020.
- Zubillaga, V. (2002). *Investigar sobre violencia en América Latina*. Un testimonio reflexivo sobre la experiencia de construir historias de vida con jóvenes de vida violenta *Revista Colombiana de Sociología* Vol VII No. 1 2002 pp 103 - 124.

Bogotá, 22 de noviembre de 2025

Señores
Facultad Ciencias Económicas – Administrativas
Decano Juan Santiago Correa Restrepo
Director Diego Fernando Cardona Madariaga

Ref: Aval Final Proyecto de Intervención en la Organización – Maestría en Administración (MBA)

En mi calidad de director del proyecto titulado “**DELITO DE HURTO A PERSONAS EN COLOMBIA: UNA PERSPECTIVA ANALÍTICA**”, presentado por los estudiantes de la maestría mencionada, me permito informar que dicho trabajo cumple con los criterios de pertinencia y validez dentro de la disciplina académica e investigativa y las condiciones de forma, para ser evaluado por el tutor de la asignatura Proyecto de Intervención en la Organización III.

NOMBRE COMPLETO ESTUDIANTE	DOCUMENTO DE IDENTIFICACIÓN
JAVIER DARÍO PRIETO RONCANCIO	80858677
ERIKA MARGARETH GUACANEME MEDINA	52430308
MAURICIO GERARDO RUBIO ROPERO	1098637654

De acuerdo con el protocolo establecido se genera el concepto final:

Las estudiantes usaron un conjunto de datos abierto que facilitó el proceso y entendimiento del problema. De esta forma, implementaron modelos de aprendizaje automático para identificar el delito y revisar la complementariedad que deben tener estos con otros procesos para definir estrategias de mitigación. Esta combinación de métodos se realizó de manera básica y les permitió identificar posibles esquemas que complementen los procesos analíticos.

Nota: 4.0

Cordialmente,



OSCAR M. GRANADOS, PhD
79.656.321
oscarm.granadose@utadeo.edu.co