

**FACULTAD DE INGENIERÍA Y ARQUITECTURA**

**CARRERA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

**Implementación de un Aplicativo Predictivo de Criminalidad basado en Machine Learning para mejorar la asignación de recursos policiales en Lima Metropolitana**

**PROBLEMA**

La ineficiente asignación de recursos policiales en Lima Metropolitana

**OBJETIVO**

Mejorar la asignación de recursos policiales en Lima Metropolitana

**AUTORES:**

Albert André Palacios Carillo (ORCID: 0009-0005-3173-9814)

Jhusbeht Casallo Veliz (ORCID: 0009-0002-2447-5281)

Luis Antony Huamani Gonzales (ORCID: 0009-0009-1874-1994)

**Profesor:**

Dr. Hugo Froilán Vega Huerta

**Lima - Perú**

**2025**

# DEDICATORIA

A nuestra familia, de las que resaltan nuestros padres, quienes a lo largo del presente proyectos mantuvieron firme su apoyo, constituyéndose en la motivación para continuar avanzando a pesar de los retos que se presentaron.

# AGRADECIMIENTO

A la Universidad Autónoma del Perú que, con su plana docente, nutrió nuestros conocimientos consolidándolos para permitir el desarrollo del presente proyecto, que contribuye a la mejora de nuestra capacidad de investigar para llegar a la verdad y presentar soluciones a los problemas de la sociedad.

# EPÍGRAFE

No se trata de hacer bien las cosas, sino de hacer las cosas bien.

(Peter Drucker, 1966)

# RESUMEN

Título : Implementación de un Aplicativo Predictivo de Criminalidad basado en Machine Learning para mejorar la asignación de recursos policiales en Lima Metropolitana

Autores : Jhusbeht Casallo Veliz

Luis Antony Huamani Gonzales

Albert André Palacios Carillo

Asesor de tesis : Dr. Hugo Froilán Vega Huerta

Fecha : Abril 2025

El presente proyecto analiza la implementación de un Aplicativo Predictivo de Criminalidad basado en Machine Learning en una comisaría de la Policía Nacional del Perú de Lima Metropolitana para mejorar la asignación de recursos materiales en el patrullaje, mediante la aplicación del método Regresión de Vectores de Soporte (SVR).

Se analizaron plataformas para predicción de delitos en sus dimensiones predicción de tiempo y predicción de espacio, para la asignación óptima de los recursos policiales de la Policía Nacional del Perú.

Se planteó realizar la implementación de un Aplicativo Predictivo de Criminalidad basado en Machine Learning en base a denuncias policiales correspondientes a los años 2018 a 2022 de delitos contra el patrimonio. Al proporcionar datos precisos sobre dónde y cuándo es más probable que ocurran delitos, se permite al personal policial asignar recursos de manera eficiente.

**PALABRAS CLAVE:**

Asignación de Recursos Policiales, Predicción del Delito, Machine Learning

# ABSTRACT

Title : Implementation of a Machine Learning-Based Predictive Crime Application to Improve the Allocation of Police Resources in Metropolitan Lima

Authors : Jhusbeht Casallo Veliz

Luis Antony Huamani Gonzales

Albert André Palacios Carillo

Thesis advisor : Dr. Hugo Froilán Vega Huerta

Date : April 2025

This project analyzes the implementation of a Machine Learning-based Crime Predictive Application in a Peruvian National Police station in Metropolitan Lima to improve the allocation of material resources during patrols, using the Support Vector Regression (SVR) method.

Crime prediction platforms were analyzed in their time and space prediction dimensions for the optimal allocation of police resources for the Peruvian National Police.

The project proposed implementing a Machine Learning-based Crime Predictive Application based on police reports for property crimes from 2018 to 2022. By providing accurate data on where and when crimes are most likely to occur, it enables police personnel to allocate resources efficiently.

**KEY WORDS:**

Police Resource Allocation, Crime Prediction, Machine Learning

# ÍNDICE

[**DEDICATORIA 2**](#_heading=h.xv2if6mng920)

[**AGRADECIMIENTO 3**](#_heading=h.jbffjtufjc0a)

[**EPÍGRAFE 4**](#_heading=h.981gidt6ovdf)

[**RESUMEN 5**](#_heading=h.jmq5p837etie)

[**ABSTRACT 6**](#_heading=h.b8vpysqmlz3e)

[**ÍNDICE 7**](#_heading=h.skzoy47o1s)

[**INTRODUCCIÓN 1**](#_heading=h.bke24mcdjq2w)

[**CAPÍTULO I: VISIÓN DEL PROYECTO 3**](#_heading=h.wbrpi71rp3fo)

[**1.1 Antecedentes del Problema 3**](#_heading=h.gw8dswb6r6e0)

[1.1.1 El Negocio 3](#_heading=h.wmfsc819ukd)

[1.1.2 Procesos del Negocio 5](#_heading=h.20pjw65iehi4)

[1.1.3 Organigrama 6](#_heading=h.jhccd6r2ykrr)

[**1.2 Formulación del Problema 6**](#_heading=h.n3gf4ecpt5ey)

[1.2.1 Realidad Problemática 6](#_heading=h.p6h5uqjv1ls)

[1.2.2 Descripción del Problema 8](#_heading=h.42yqwojemc60)

[**1.3 Objetivos del Proyecto 9**](#_heading=h.rspi0w2tiouq)

[1.3.1 Marco Lógico 9](#_heading=h.42na72hpkgbi)

[1.3.2 Objetivo General 10](#_heading=h.ieu7bomfqg1d)

[1.3.3 Objetivos Específicos 11](#_heading=h.h9r49lqmxhc6)

[**1.4 Justificación del Proyecto 11**](#_heading=h.tms5nb3zz90r)

[1.4.1 Justificación Académica 11](#_heading=h.4fu4oqvnvd0i)

[1.4.2 Beneficios Tangibles 12](#_heading=h.cqmlsztkzysm)

[1.4.3 Beneficios Intangibles 12](#_heading=h.pn68kf1mi2uq)

[**1.5 Alcance del Proyecto 12**](#_heading=h.2f0zqk9lxyrg)

**CAPÍTULO I**I**:** **MARCO** **TEÓRICO………………………………………………………………..13**

**CAPÍTULO III: ESTADO DEL ARTE 16**

**3.1. Artículos………………………………………………………………………………………. 16**

[**REFERENCIA BIBLIOGRÁFICA 13**](#_heading=h.xr1sdni8ds5n)

# INTRODUCCIÓN

En la actualidad gracias a las Tecnologías de la Información y Comunicación (TICs) entidades del estado que luchan contra la delincuencia aprovechan sus beneficios; puesto que, les ha permitido conocer indicadores y estadísticas de los delitos (Ordóñez et al., 2020). Así, de las tecnologías emergentes de la industria 4.0 resalta la inteligencia artificial, dentro de la cual tenemos a la Machine Learning, un modelo que permite, entre otras cosas la predicción del delito. De esa forma, se constituyen en una herramienta útil para construir estrategias de prevención en ciudades principales (Gelvez et al., 2022). Lo interesante es que este modelo, se alimenta de la información que ha sido previamente almacenada en bases de datos mediante sistemas de información como pueden ser las denuncias policiales por delitos contra el patrimonio.

En la Policía Nacional del Perú el 99,2% de comisarías básicas cuenta con computadoras, que forma parte del equipamiento básico de sus oficinas para el cumplimiento de sus funciones; sin embargo existe un 16% de los equipos de cómputo que no cuenta con un acceso a internet (Ministerio del Interior [MININTER], 2025). Además, la flota vehicular, para uso administrativo y operativo, no se encuentra en óptimas condiciones ni en la cantidad adecuada para satisfacer toda la demanda requerida por las unidades policiales en el ámbito nacional, tanto en vehículos menores, livianos y pesados, aeronaves, embarcaciones y talleres de mantenimiento (Policía Nacional del Perú, 2021); lo que demuestra la importancia en la asignación adecuada de los pocos recursos disponibles con los que cuenta la institución policial encargada de luchar contra la delincuencia.

La asignación de recursos para el patrullaje, por ende, es un tema fundamental que ha cobrado una relevancia significativa en el contexto actual de seguridad ciudadana. Este aspecto se ve profundamente afectado por la inexistencia de un protocolo específico que regule su planificación, organización, ejecución y control (Campos & Quiroz, 2023). La falta de directrices claras no solo genera confusión entre los efectivos policiales, sino que también impacta negativamente en la eficiencia del servicio. A esto se suma el alarmante dato de que hasta un 30% de los patrulleros se encuentran inoperativos, lo que limita severamente la capacidad operativa de las fuerzas del orden (Ponce & Poma, 2023). Esta situación resalta la necesidad urgente de optimizar la utilización de los escasos recursos disponibles, garantizando así una respuesta más efectiva ante las demandas de seguridad de la población.

En cuanto al uso de inteligencia artificial en las fuerzas policiales, diversos estudios han abordado su aplicación en áreas como el reconocimiento automático de matrículas (Morón & Rodriguez, 2023) y el desarrollo de modelos espacio-temporales que permiten predecir delitos a partir de datos históricos sobre criminalidad (Barragán et al., 2023). Otros estudios han explorado su potencial para identificar tendencias delictivas (Sueldo & Peña, 2022). Sin embargo, a pesar del avance tecnológico y las diversas aplicaciones documentadas en la literatura, persiste una brecha significativa respecto a cómo estas herramientas pueden impactar específicamente en la asignación eficiente de recursos para el patrullaje.

La implementación del aplicativo tiene implicaciones significativas para la estrategia policial. Al proporcionar datos precisos sobre dónde y cuándo es más probable que ocurran delitos, se permite al personal policial asignar recursos más estratégicamente. Esto no solo contribuye a una mayor eficiencia en el uso del tiempo y los vehículos disponibles, sino que también puede resultar en una disminución general en las tasas de criminalidad al aumentar la presencia policial en áreas identificadas como problemáticas.

La presente tesis está dividida en las siguientes partes:

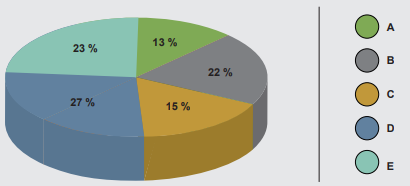
* En el capítulo I, muestra la visión del proyecto, donde se describen y detallan los antecedentes que se tomaron como referencia, la problemática a definir, así como el objetivo general y específicos. Seguidamente se menciona la justificación del tema y el alcance a desarrollarse.

# CAPÍTULO I: VISIÓN DEL PROYECTO

## 1.1 Antecedentes del Problema

### 1.1.1 El Negocio

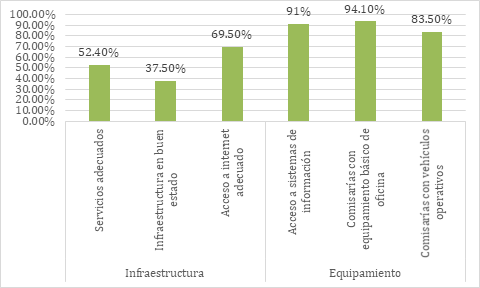
Las comisarías de la Policía Nacional del Perú son la célula base de esta institución y suman un total de 1,318 a nivel nacional. Según su tipo pueden clasificarse en A, B, C, D, E de acuerdo a su capacidad de efectivos policiales, área construida y cobertura de habitantes. La A representa la de mayor nivel y en su contraparte se encuentra la E. Así según su tipo en el Perú el 13 % es de tipo A, 22% de tipo B, 15% de tipo C, 27% de tipo D y 23% de tipo E (Ministerio del Interior [MININTER], 2025).



***Figura 1.1.*** *Porcentaje de comisarías PNP por tipo*

***Fuente:*** *(MININTER, 2025)*

En cuanto a la infraestructura de las comisarías básicas a nivel nacional se tiene que un 52,4% cuenta con servicios adecuados, el 37,5% cuenta con infraestructura en buen estado, el 69,5% con acceso a internet adecuado. Además, en lo que respecta a equipamiento el 91% cuenta con acceso a sistemas de información, el 94.1% con equipamiento básico de oficina y el 83.5% con vehículos operativos. (Policía Nacional del Perú, 2021)



***Figura 1.2.*** *Estado de la infraestructura y equipamiento de las comisarías PNP a nivel nacional*

***Fuente:*** *(MININTER, 2025)*

La situación anterior no ayuda a luchar contra la comisión de delitos, en el Perú, que según el Ministerio del Interior, las extorsiones en el primer trimestre del 2025 superan a las registradas en los años 2022, 2023 y 2024, en ese mismo periodo. A su vez, en Lima cada 24 minutos se registra una nueva extorsión que conlleva a que uno de cada cuatro personas haya sido víctima de este delito. Lo anterior podría explicar porque en nuestro país sucede un homicidio en cuatro horas.

Sabemos entonces que la mayoría de las comisarías básicas de la Policía Nacional del Perú no cuentan con internet adecuado; y no solo eso, sino que también la cantidad de vehículos inoperativos es preocupante. Por ello herramientas tecnológicas innovadoras permiten incrementar la eficiencia, es decir lograr los objetivos, en este caso de prevención del delito, a pesar de los reducidos recursos, mediante su uso óptimo.

Una de las propuestas representa el Aplicativo Predictivo de Criminalidad basado en Machine Learning para mejorar la asignación de recursos policiales y de esta manera contrarrestar los altos índices de extorsiones que se vienen dando en el país.

### 1.1.2 Procesos del Negocio

Diagrama

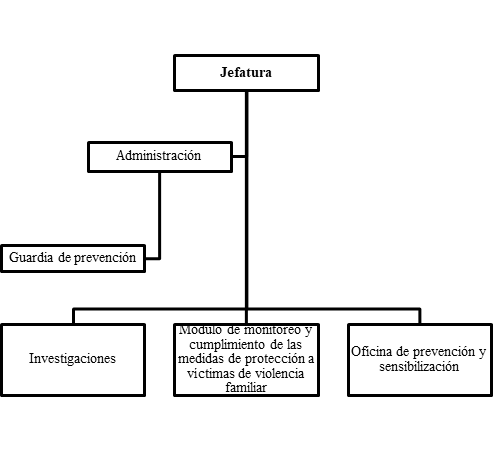
El contenido generado por IA puede ser incorrecto.

***Figura 1.3.*** *Proceso de asignación de vehículos*

***Fuente:*** *(Elaboración propia, 2025)*

### 1.1.3 Organigrama

A continuación se muestra un organigrama general de las comisarías PNP.

****

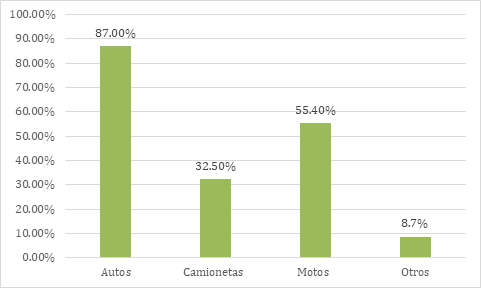
***Figura 1.4.*** *Organigrama de una comisaría PNP*

***Fuente:*** *(Elaboración propia, 2025)*

## 1.2 Formulación del Problema

### 1.2.1 Realidad Problemática

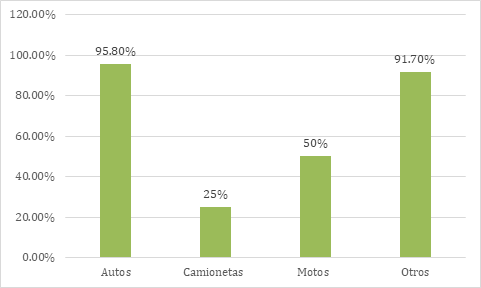
Los recursos en la institución policial resultan relevantes para mejorar la seguridad ciudadana, pero cuando se habla de vehículos policiales o también conocidos como patrulleros, se tiene la lamentable cifra de unidades que no pueden ser utilizadas debido a que se encuentran inoperativas. En ese sentido existe un 87% de autos inoperativos, un 32,5% de camionetas inoperativas, 55,4% motos inoperativas y 8,7% otros (Ministerio del Interior [MININTER], 2025).



***Figura 1.5.*** *Vehículos inoperativos de las comisarías PNP a nivel nacional*

***Fuente:*** *(MININTER, 2025)*

Ahora bien, en Lima se agrava la situación porque existe un 95.8% de autos inoperativos, un 25% de camionetas inoperativas, 50% motos inoperativas y 91,7% otros (Ministerio del Interior [MININTER], 2025).



***Figura 1.6.*** *Vehículos inoperativos de las comisarías PNP a nivel Lima*

***Fuente:*** *(MININTER, 2025)*

### 1.2.2 Descripción del Problema

#### 1.2.2.1 Problema Principal

En las comisarías de la Policía Nacional del Perú los recursos policiales, especialmente en lo que respecta a los vehículos, presentan una serie de deficiencias. Esto se debe principalmente a la ausencia de Machine Learning en el proceso de asignación de recursos (100% manual), lo que genera una asignación ineficiente de los mismos. Además, se observa una ausencia del 60% de datos predictivos en la hoja de ruta de patrulleros, lo que afecta la planificación y eficiencia de las patrullas. También existe un déficit del 40% de protocolo de asignación de patrulleros. Todo esto genera en conjunto la ineficiente asignación de recursos policiales en Lima Metropolitana, ya que la falta de datos predictivos y protocolos establecidos impacta directamente en la eficiencia del patrullaje. Como resultado, se realiza un patrullaje con un 60% de información imprecisa, lo que incrementa el 45% en la incidencia delictiva y provoca la asignación ineficiente del 40% de patrulleros, que se distribuye de manera desorganizada, sin considerar las áreas de mayor necesidad.

Variable 1: Número de denuncias registradas por extorsión. (Valor 7,779), según el Ministerio del Interior MININTER.

Variable 2: Número de patrullajes realizados en zonas de alta probabilidad delictiva. (Valor 40), según el Ministerio del Interior MININTER.

#### 1.2.2.2 Problema Secundario

* Asignación de recursos con información imprecisa.
* Incremento de incidencia delictiva.
* Asignación de recursos ineficiente.

## 1.3 Objetivos del Proyecto

### 1.3.1 Marco Lógico

#### 1.3.1.1 Árbol del Problemas:

***Figura 1.4.*** *Árbol de problemas determinados*

***Fuente:*** *(Elaboración propia, 2025)*

#### 1.3.1.2 Árbol de Objetivos:



***Figura 1.5.*** *Árbol de objetivos determinados*

***Fuente:*** *(Elaboración propia, 2025)*

### 1.3.2 Objetivo General

Desarrollar e implementar un aplicativo predictivo de criminalidad con Machine Learning funcional al 100%. Este sistema permitirá que el jefe de patrullaje forme una hoja de ruta de datos predictivos al 100%, tomando decisiones basadas en información precisa con el objetivo de mejorar la asignación de recursos policiales en Lima Metropolitana. Además, se implementará un protocolo actualizado para la asignación de patrulleros al 100%, asegurando una distribución más eficiente de los recursos disponibles y optimizando el patrullaje en la ciudad, logrando así la eficiencia en la asignación de recursos policiales de Lima Metropolitana, esto permitirá que el 100% de los patrullajes se basen en datos predictivos, lo que incrementará en un 60% el uso eficiente de recursos policiales y logrará una reducción del 50% en la incidencia delictiva en zonas de alto riesgo.

Variable 1: Número de denuncias registradas por extorsión. ( Valor 1,640), según el Ministerio del Interior MININTER

Variable 2: Número de patrullajes realizados en zonas de alta probabilidad delictiva. (Valor 220),según el Ministerio del Interior MININTER

### 1.3.3 Objetivos Específicos

* Diseñar e implementar un aplicativo predictivo basado en Machine Learning
* Optimizar la toma de decisiones en asignación de recursos policiales en base a información predictiva.
* Conocer el lugar y tiempo donde ocurrirán los delitos.
* Definir los módulos de implementación del aplicativo predictivo.
* Determinar los criterios de un aplicativo predictivo para una comisaría básica de Lima Metropolitana.

## 

## 1.4 Justificación del Proyecto

### 1.4.1 Justificación Académica

Se tiene conocimiento que según el ranking de Competitividad Mundial publicado por el Instituto Internacional para el Desarrollo Gerencial de Suiza (2024), pilar eficiencia del gobierno, el Perú se encuentra en el puesto 60 en el año 2024, situación contraria al año 2008 en el que se ubicaba en el puesto 26.

Gráfico, Gráfico de cajas y bigotes

El contenido generado por IA puede ser incorrecto.

***Figura 1.5.*** *Eficiencia del Gobierno peruano del 2008 al 2024*

***Fuente:*** *(COMEXPERÚ, 2025)*

En esa línea, el incremento en el uso de tecnologías permite el uso eficiente de los recursos, pero existe una brecha en la literatura respecto a la aplicación del Machine Learning en la asignación de recursos, por lo que la presente tesis pretende llenar ese vacío existente en el Perú.

En cuanto a la eficiencia debido a que las comisarías de básicas de la Policía Nacional del Perú ubicadas en Lima Metropolitana, según el Ministerio del interior tiene porcentajes considerables de vehículos, aportaría con el uso eficiente de los recursos permitiendo abarcar mayores puntos de patrullaje con menores vehículos e impactando positivamente en la reducción de hechos delictivos porque se actuaría predictivamente, sabiendo dónde y cuándo podría ocurrir un delito.

### 1.4.2 Beneficios Tangibles

* Reducción de tiempo en asignación de recursos.
* Uso eficiente de vehículos policiales para el patrullaje.

### 1.4.3 Beneficios Intangibles

* Satisfacción del ciudadano.
* Bienestar en el efectivo policial por cumplimiento efectivo de su labor.

## 1.5 Alcance del Proyecto

La presente tesis constituye una investigación del empleo de tecnologías como machine learning en la asignación de recursos vehiculares para lograr el uso eficiente de los escasos recursos con los que cuenta la Policía Nacional del Perú, permitiendo a los encargados de la asignación de la hoja de ruta un insumo objetivo para cumplir de manera eficaz su función.

La elección del recurso vehicular es porque representa un medio que abarca mayor área geográfica, creando una sensación de seguridad y disuadiendo la comisión de hechos delictivos, ya que la delincuencia ante la presencia policial evita su accionar.

# CAPÍTULO II: MARCO TEÓRICO

# CAPÍTULO III: ESTADO DEL ARTE

3.1. **Artículos**

### 3.1.1 Crime Prediction and Monitoring in Porto, Portugal, Using Machine Learning, Spatial and Text Analytics

### Predicción y seguimiento de la delincuencia en Oporto (Portugal) mediante aprendizaje automático, análisis espacial y textual

(Saraiva et al., 2022)

**(DOI:10.3390/ijgi11070400)**

**Citado 25 veces en Scopus**

En el presente artículo los autores explican que desde hace tiempo las respuestas tradicionales para combatir la delincuencia no son suficientes por sí mismas. Por eso, es que el paradigma policial ha migrado de la reacción hacia la prevención, y de analizar únicamente al agresor a los factores sociales contextuales. Eso ha obligado a las agencias de seguridad a aplicar enfoques parapredecir y prevenir los incidentes, ayudados en el uso de tecnologías espaciales, minería de datos y aprendizaje automático**.**

La tendencia actual, sugiere un enfoque en base a los principios de la criminología ambiental que se basan en tres ideas principales. En primer lugar, la conducta influenciada por la naturaleza del espacio con características individuales. En segundo lugar, la distribución de los patrones delictivos con el modelo Support Vector Machine (SVM) que no es aleatoria sino una consecuencia de las condiciones del territorio y que varían en espacio y tiempo. En último lugar, al cambiar esas características se podría reducir la inseguridad y el uso eficiente de recursos.

En el mencionado artículo el caso de estudio es la ciudad de Oporto donde se analizó los puntos críticos teniendo en cuenta el espacio-tiempo. De esa forma con la agrupación de clústeres y la herramienta de aprendizaje automático se identificó los patrones naturales de los datos. Luego se agrupó de manera espacial la delincuencia en relación con otros datos.

La investigación no debe terminar con la correlación de variables, si bien es un conocimiento importante, el que permitirá propuestas de soluciones es la explicativa. El motivo es que, en primer lugar, la correlación no significa causalidad y, en segundo lugar, porque, las ubicaciones a micro escala son sistemas urbanos y sociales que se relacionan con cuestiones personales y podrían no computarse en lo absoluto. Eso se puede mitigar con una modelización y comprensión espacial más profunda.

**B) Metodología**

Inicia con la comprensión del patrón puntual de los delitos registrados, seguido del análisis de aprendizaje automático basados en métodos supervisados para determinar la influencia de los factores contextuales urbanos, morfológicos y socioeconómicos. Luego, mediante la regresión Lasso se utiliza una penalización de nivel 1 para seleccionar un subconjunto de predictores que son los más importantes en términos de delincuencia. Tener menos predictores con un mayor poder predictivo reduce el error de predicción y minimiza el tiempo y los recursos computacionales, además de evitar que el modelo de predicción se sobreajuste. Después, para crear el modelo de clasificación la búsqueda en cuadrículas con validación cruzada sobre una variedad de hiper parámetros. Finalmente, para analizar la dimensión de la actividad social y la opinión con respecto a la delincuencia, se recopilaron tuits de Twitter utilizando la biblioteca Snscrape, un programa de búsqueda de servicios de redes sociales en Python.

C) **Etapas**

Etapa 1: comprensión del patrón.

Etapa 2: Regresión Lasso.

Etapa 3: Clasificación con aprendizaje automático:

- Logistic Regression

- Decision Tree

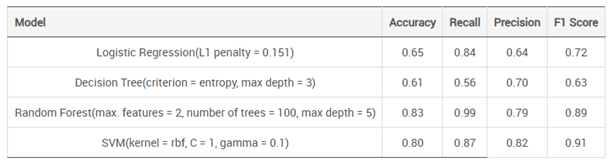
- Random Forest

- SVM

Etapa 4: análisis de sentimientos.

**Tabla 1**

*Comparación del rendimiento del modelo de clasificación de aprendizaje automático*

******

***Fuente:*** *(Saraiva, Matijošaitienė, Mishra, & Amante, 2022)*

**Utilidad del artículo para el proyecto de tesis**

Este artículo es una referencia fundamental para la tesis propuesta, **contribuyendo a que se comprenda los enfoques predictivos basados en aprendizaje automático son efectivos para el análisis y predicción de patrones delictivos en contextos urbanos complejos, lo cual proporciona una alternativa metodológica sólida para el desarrollo del sistema predictivo propuesto para la asignación eficiente de recursos policiales en Lima Metropolitana**. En ese sentido este artículo se enfoca en los conceptos clave como aprendizaje automático, esencial para crear predicciones confiables; criminología ambiental, para comprender la influencia del entorno sobre los delitos; espacio-tiempo, que permite identificar patrones delictivos dinámicos según lugar y momento con un modelo de predicción Support Vector Machine (SVM), por su alto desempeño predictivo, son elementos fundamentales que aplicaremos directamente en nuestra tesis.

### 3.1.2 Crime Prediction using Machine Learning

### Predicción de delitos mediante aprendizaje automático

### (Sridharan et al., 2024)

### (DOI: 10.4108/eetiot.5123)

**Citado 4 veces en Scopus**

En el presente artículo los autores explican que el proceso que permite investigar las tendencias delictivas para identificar causas y posibles soluciones en la prevención de delitos se conoce como análisis delictivo. Así también establecen que este tipo de análisis también implica la creación de modelos predictivos que utilizan datos previos.

Al predecir el delito, se presentan varios retos uno de los que resalta es que los delitos violentos no se pueden proyectar con certeza debido a que son sistemáticos o aleatorios. En ese mismo sentido, es muy difícil predecir quién podría cometer un delito o quien podría ser la víctima, pero si es posible es posible el lugar donde ocurra en un espacio de tiempo determinado.

En el presente artículo, se proporciona algoritmos como la regresión lineal y el bosque aleatorio utilizando datos del año 2001 al 2016 que permiten proyecciones para cada estado de la India, con el objeto de predecir la delincuencia en el periodo 2017 al 2020, eligiendo el marco denominado la teoría de las ventanas rotas que obliga la inclusión de datos demográficos.

Mediante lo anterior, las autoridades policiales pueden asignar recursos de manera eficiente y enfocar sus iniciativas con la finalidad de reducir la delincuencia e incrementar la seguridad. De esta forma, se ofrece una posibilidad de reingeniería de procesos para dirigir los recursos a las zonas que tendrían alta probabilidad, en el futuro, de presentar riesgo de delitos.

1. **Metodología**

Mediante el presente modelo se emplean algoritmos de clasificación, agrupamiento y regresión: algoritmos k-Nearest Neighbors (kNN), Naive Bayes y de regresión que pueden describir relaciones funcionales entre variables demográficas y geográficas mediante análisis de patrones. En el artículo primero se realizó el procesamiento cuando los datos se transforman en clústeres mediante algoritmos de agrupamiento para identificar zonas con mayor riesgo, se representan gráficamente en un mapa. Segundo, el procesamiento en el que se combina el tipo y momento del delito. Los miembros de los clusters se utilizan para categorización: los que tienen alta densidad se convierten en focos de delincuencia y los que menos tienen, se descartan. La técnica de agrupamiento popular es K-medias. Es elegida por que la naturaleza de los delitos varía mucho, por eso, se deja de lado a la técnica supervisada: clasificación. Tercero, el modelo se examinó, creó y utilizó para muestrear el conjunto de datos y entrenar al algoritmo, con la teoría de la ventana rota, y las técnicas bosque aleatorio y Naive Bayes se localizó el área criminal. Por último, el resultado fue que la mejor precisión en la predicción es proporcionada por el modelo ajustado de aprendizaje profundo. Es posible predecir las tasas de criminalidad empleando las técnicas de clasificación y regresión de aprendizaje automático.

B) **Etapas**

Etapa 1: procesamiento.

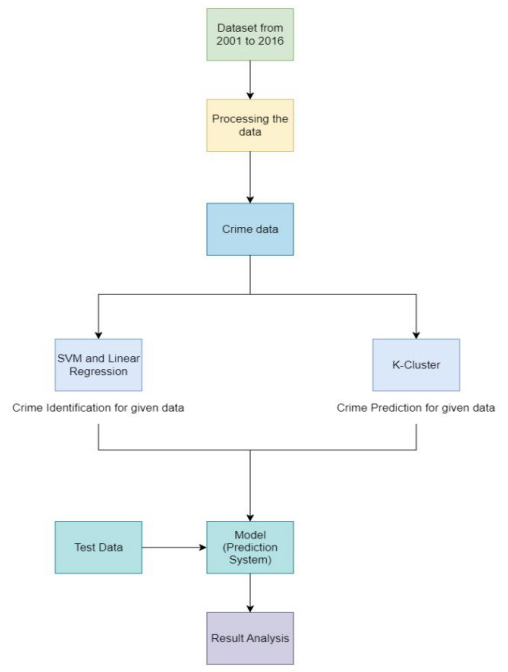
Etapa 2: datos procesados.

Etapa 3: modelo de predicción.

Etapa 4: análisis de los resultados.

Figura 1

*Arquitectura del sistema*



Nota. Fuente: (Sridharan, Srish, Vigneswaran, & Santhi, 2024)

**Utilidad del artículo para el proyecto de tesis**

Este artículo aporta significativamente a nuestro proyecto de tesis sobre la implementación de un aplicativo predictivo de criminalidad basado en Machine Learning, al destacar la relevancia de construir algoritmos de clasificación y agrupación modelos predictivos para la prevención de delitos. Se enfatiza la importancia de considerar el espacio de tiempo para contextualizar correctamente cuándo y dónde pueden ocurrir los delitos, mejorando así la capacidad de anticipación del sistema. La aplicación de técnicas estadísticas como la regresión lineal demuestra ser efectiva para modelar tendencias delictivas a partir de datos históricos, permitiendo realizar proyecciones confiables. Este enfoque brinda un marco robusto para que nuestro modelo pueda identificar con mayor precisión las zonas de alto riesgo y los periodos críticos, lo que es fundamental para tomar decisiones preventivas. Además, al integrar estos modelos, es posible asignar recursos de manera eficiente, focalizando la acción policial en los sectores con mayor probabilidad de ocurrencia de delitos. Esto no solo optimiza la gestión de los recursos disponibles, sino que también potencia la efectividad de las estrategias de seguridad y prevención. En consecuencia, la metodología presentada fortalece nuestra tesis al aportar fundamentos y técnicas que pueden ser adaptadas para el contexto específico de Lima Metropolitana, contribuyendo a la mejora de la seguridad ciudadana a través de sistemas predictivos más precisos y eficientes.

### 3.1.3. A study on the application of data mining-based crime prediction models in criminal justice Estudio sobre la aplicación de modelos de predicción de delitos basados ​​en minería de datos en la justicia penal.

(Zhao, 2024)

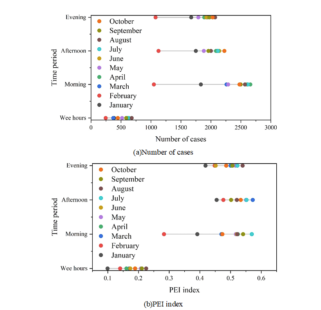
**(DOI: 10.61091/jcmcc123-40)**

**Citado 1 veces en Scopus**

En el presente artículo de Liupeng Zhao se enfoca en la aplicación de modelos de predicción de delitos basados en minería de datos dentro del ámbito de la justicia penal. A medida que las actividades criminales se especializan y se vuelven más complejas, la necesidad de predecir y prevenir estos delitos se ha vuelto crucial. Zhao presenta una metodología que integra redes neuronales gráficas (GAE) con el aprendizaje de características espaciales y temporales, mejorando la precisión de la predicción de crímenes al incluir mecanismos de atención espacial. La investigación se centra en construir un modelo predictivo de delitos utilizando una red neuronal recurrente de auto codificación gráfica (GAERNN) que combina datos espaciales y temporales, optimizando la capacidad del modelo para predecir comportamientos delictivos a partir de datos complejos.

El modelo propuesto, GAERNN, se construye mediante la incorporación de la codificación espacial con atención, lo que permite al modelo enfocarse en áreas y características clave relacionadas con el crimen. El artículo también aborda el uso de funciones de pérdida innovadoras y algoritmos recurrentes como el GRU, con el fin de mejorar la precisión en la predicción temporal. A través de comparaciones experimentales, se muestra que el GAERNN supera otros métodos tradicionales en cuanto a la precisión de predicciones, especialmente al manejar datos que contienen variaciones espaciales y temporales de alta frecuencia.

El estudio también profundiza en cómo el uso de datos ambientales y características del contexto local (como el tipo de área, instalaciones de interés, etc.) puede mejorar la exactitud de las predicciones. El análisis de riesgo especial, que evalúa las características de los entornos de los crímenes pasados, es crucial para prevenir futuros incidentes. Se concluye que la integración de estos datos con la predicción de delitos no solo aumenta la efectividad de la asignación de recursos, sino que también ofrece una forma más precisa de identificar zonas de alto riesgo para la prevención.



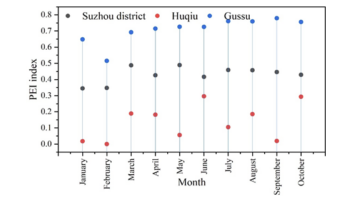
***Figura 3.1.*** *Estadísticas sobre el número de casos de la invasión*

***Fuente:*** *(Zhao, L. (2024)*

**B) Metodología**

Los autores desarrollaron su investigación en base al uso de técnicas avanzadas de minería de datos y aprendizaje automático, particularmente redes neuronales gráficas (GAE) y redes neuronales recurrentes (GRU). Los pasos principales de la metodología incluyen:

1. **Recopilación de Datos Espaciales y Temporales:** Se utilizan datos históricos de delitos y características espaciales, como puntos de interés (POI) y factores ambientales de la región.
2. **Modelado con GAERNN:** Este modelo combina las capacidades de auto-codificación gráfica y redes neuronales recurrentes para predecir la ocurrencia de delitos en futuras ventanas temporales.
3. **Optimización del Modelo:** Se aplican funciones de pérdida avanzadas y mecanismos de atención para mejorar la precisión de las predicciones en escenarios complejos.
4. **Evaluación y Comparación de Resultados:** Se comparan los resultados de GAERNN con otros modelos tradicionales utilizando métricas como RMSE, MSE y MAPE para evaluar la exactitud de las predicciones.



***Figura 3.2.*** *Índice De Previsión De Delitos Regionales Diferentes PEI*

***Fuente:*** *(Zhao, L. (2024)*

**Utilidad del artículo para el proyecto de tesis**

Este artículo representa una contribución valiosa para el desarrollo del proyecto de tesis, al evidenciar cómo la integración de datos espaciales y temporales mediante técnicas avanzadas de minería de datos y aprendizaje automático, específicamente redes neuronales gráficas recurrentes (GAERNN), mejora significativamente la precisión en la predicción de delitos. Estos aportes se enfocan enormemente en la integración de minería de datos, que nos permite analizar y procesar grandes volúmenes de información criminal de manera eficiente. La capacidad de combinar datos espaciales y temporales en el análisis es fundamental para capturar la dinámica real de los patrones delictivos, lo que ayuda a contextualizar mejor la ocurrencia de delitos en diferentes lugares y momentos específicos. La utilización de redes neuronales gráficas (GAE) representa un avance significativo, al permitir la modelación de relaciones complejas entre diferentes variables espaciales y temporales, mejorando la precisión del modelo predictivo. Además, el enfoque del artículo en la capacidad para predecir y prevenir delitos mediante el uso de tecnologías avanzadas es directamente aplicable a nuestra tesis, ya que permite anticipar comportamientos delictivos y diseñar estrategias efectivas para la intervención policial. Son cruciales para focalizar el análisis en áreas de mayor riesgo, optimizando los recursos disponibles y mejorando la toma de decisiones en tiempo real. Asimismo, la mejora en la precisión en la predicción temporal garantiza que las alertas y acciones preventivas se realicen oportunamente, reduciendo la incidencia delictiva.

### 3.1.4. Comparing Regression Models to Predict Property Crime in High-Risk Lima Districts Comparación de modelos de regresión para predecir delitos contra la propiedad en distritos de alto riesgo de Lima.

(Escobedo et al., 2024)

**(DOI:** [**10.14569/IJACSA.2024.0150307**](https://dx.doi.org/10.14569/IJACSA.2024.0150307)**)**

**Citado 6 veces en Scopus**

En el presente artículo, el autor explica que el objetivo de la predicción de delitos en el futuro es anticipar el lugar y el momento en que ocurrirán los crímenes, utilizando modelos de regresión supervisada. Este enfoque ha ganado atención por parte de diversas instituciones encargadas de la seguridad pública y la vigilancia predictiva.

Debido a esto, los investigadores han empleado varios modelos de regresión como XGBoost, Extra Tree, Support Vector, entre otros, para analizar datos de crímenes en Lima y predecir delitos contra la propiedad. Sin embargo, el autor destaca que aún existe una falta de evaluaciones rigurosas sobre la precisión y efectividad de estos modelos para prevenir crímenes

En dicho artículo se proporciona una revisión de las técnicas de regresión supervisada utilizadas en la predicción aprendizaje automático de crímenes y se destacan tres cuestiones clave sobre el uso de estos modelos algorítmicos, las cuales son las siguientes.

Primero, los supuestos sobre los cuales se basan estos algoritmos y cómo los datos utilizados para entrenarlos pueden afectar la precisión de las predicciones.

Segundo, aunque se han realizado algunas evaluaciones, aún se necesitan más estudios rigurosos para comprender mejor el impacto de estos modelos en la distribución de recursos policiales.

Finalmente, la capacidad de estos algoritmos de predecir zonas de alta incidencia de delitos sin perjudicar la capacidad de los agentes de policía o de los organismos encargados de la seguridad.

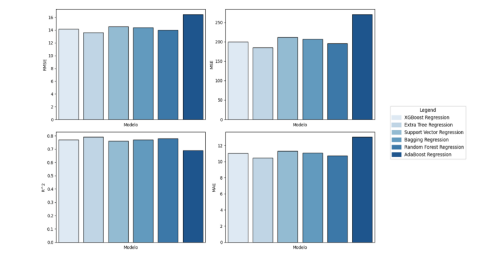
**A) Modelo Subyacente de Predicción de Delitos Contra la Propiedad**

Consiste en un ciclo de actividades y puntos de decisión: recopilación de datos, análisis, operaciones policiales, respuesta criminal y regreso a la recopilación de datos. En cada etapa del ciclo, se toman decisiones sobre, por ejemplo, los tipos de datos a recopilar como hora del día, día de la semana, ubicación geográfica y frecuencia de crímenes contra la propiedad, la frecuencia de la recopilación y actualización de datos, las herramientas analíticas a emplear como los modelos de regresión supervisada, y las variables dependientes en las que centrarse, tales como patrones espaciales y temporales de los crímenes.

**B) Metodología**

Los autores desarrollaron su investigación utilizando técnicas de minería de datos y aprendizaje automático, particularmente con modelos de regresión supervisada. Los pasos principales de la metodología incluyen:

1. **Recopilación de Datos Espaciales y Temporales:** Se utilizaron datos históricos de crímenes contra la propiedad y características espaciales de Lima, como la ubicación geográfica de los crímenes y variables temporales (hora del día, día de la semana). Estos datos permiten identificar patrones de criminalidad y predecir la ocurrencia futura de delitos en áreas específicas.
2. **Modelado con Modelos de Regresión Supervisada:** Se aplicaron modelos de regresión supervisada, como XGBoost, Extra Tree, y Random Forest, para predecir la cantidad de crímenes en función de variables como la ubicación y el tiempo. El modelo Extra Tree Regression demostró ser el más efectivo, alcanzando un R 2 de 0.79.
3. **Optimización del Modelo:**  
    Se utilizó la optimización de hiperparámetros mediante GridSearchCV para mejorar el desempeño de los modelos, minimizando errores como el MSE, RMSE y MAE, y asegurando la precisión de las predicciones.
4. **Evaluación y Comparación de Resultados:** Los resultados obtenidos de los modelos fueron evaluados utilizando métricas como R 2, MSE, RMSE y MAE. Además, se compararon los resultados de Extra Tree, XGBoost y otros modelos de regresión para determinar cuál ofrecía las predicciones más precisas para el crimen en Lima.



***Figura 3.4*** *Métricas de Modelos de Regresión*

***Fuente:*** *Escobedo, M., Tapia, C., Gutierrez, J., & Ayma, V. (2024).*

**Utilidad del artículo para el proyecto de tesis:**

Este artículo aportará significativamente a nuestro proyecto de tesis al proporcionar una revisión de las técnicas de regresión supervisada aplicadas en la predicción de crímenes. Los modelos empleados en el artículo, como XGBoost, Extra Tree, y Support Vector, permiten analizar grandes volúmenes de datos de crímenes y predecir delitos contra la propiedad. Además, la revisión de las técnicas de aprendizaje automático utilizadas en el análisis de crímenes fortalecerá las predicciones y la asignación de recursos para prevenir crímenes. Este enfoque se centra en optimizar la precisión de los modelos, utilizando evaluaciones rigurosas para mejorar los resultados y garantizar la efectividad de los modelos predictivos.En este artículo, usaré los modelos de regresión supervisada como XGBoost, Extra Tree, y Support Vector para analizar datos de crímenes en Lima y predecir delitos contra la propiedad. Además, la optimización de los modelos mediante aprendizaje automático será clave para mejorar la precisión de las predicciones de delitos y para ayudar en la asignación eficiente de recursos policiales, focalizando los esfuerzos en las áreas con mayor probabilidad de ocurrencia de delitos. Los enfoques y modelos descritos en el artículo ayudarán a prevenir crímenes y a mejorar la seguridad pública mediante la vigilancia predictiva.

### 3.1.5. Crime Prediction Using Machine Learning: A Comparative Analysis

### Predicción de delitos mediante aprendizaje automático: un análisis comparativo

(Alsubayhin et al., 2023)

**(DOI: 10.3844/jcssp.2023.1170.1179)**

**Citado 10 veces en Scopus**

En este artículo, los autores analizan el uso del aprendizaje automático como una herramienta fundamental para predecir delitos y fortalecer la prevención del crimen, posicionándose como un recurso valioso dentro de las estrategias de seguridad ciudadana. A través de un análisis comparativo de 51 estudios académicos, se identificaron los principales algoritmos de Machine Learning aplicados a este campo para recursos policiales.

Los autores explican que este tipo de herramientas tecnológicas busca facilitar la toma de decisiones estratégicas en seguridad, al aprovechar los patrones históricos de criminalidad. Resaltan que el aprendizaje supervisado es el enfoque más utilizado, siendo el algoritmo Random Forest el que ha demostrado mayor precisión en la predicción del delito. A pesar de los buenos resultados, se enfatiza la necesidad de validar estos modelos en escenarios reales, especialmente para apoyar en la asignación eficiente de recursos policiales y mejorar la cobertura del patrullaje preventivo.

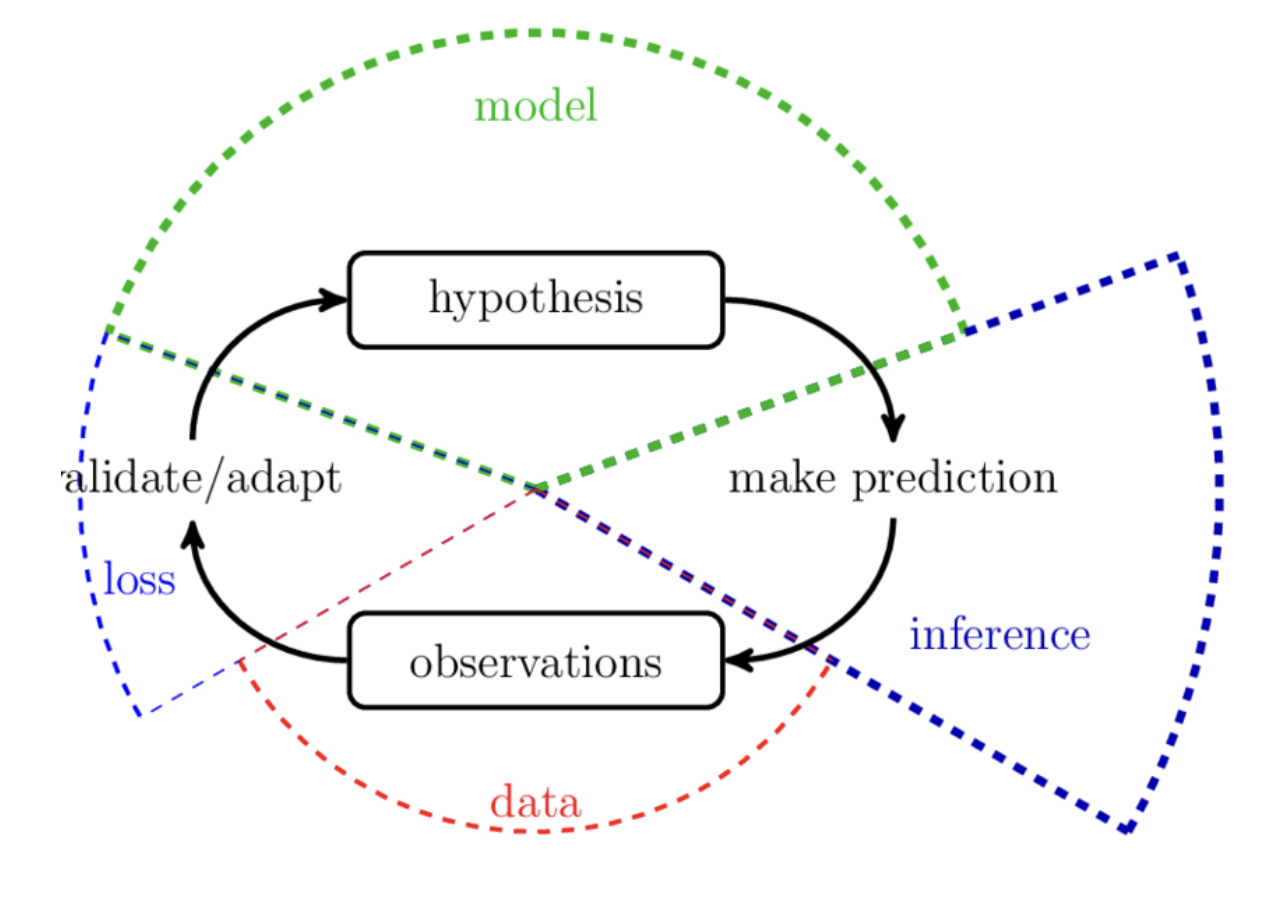
Asimismo, se identifican desafíos frecuentes en la implementación de estos sistemas, como la dificultad para acceder a datos limpios y estructurados, el alto costo computacional, y la variabilidad de resultados según el contexto urbano. Por ello, se propone avanzar hacia modelos más interpretables y adaptables que puedan aplicarse eficazmente en instituciones públicas con limitaciones operativas.

**A) Referentes Teóricos**

Los autores destacan que el Machine Learning es una rama de la inteligencia artificial que ha demostrado gran potencial en áreas como la predicción del crimen. Este enfoque se basa en el análisis de grandes volúmenes de datos para identificar patrones que permitan anticipar hechos delictivos.

* **Naïve Bayes:** Algoritmo basado en probabilidades condicionales, útil para clasificar eventos según su ocurrencia pasada.
* **Árboles de Decisión:** Estructuras jerárquicas que permiten tomar decisiones a partir de condiciones lógicas en los datos.
* **Redes Neuronales Artificiales (ANN):** Modelos que simulan el aprendizaje humano ajustando pesos entre conexiones para minimizar errores.
* **Random Forest:** Técnica de ensamble que combina múltiples árboles de decisión para mejorar la exactitud de las predicciones.
* **Regresión Logística:** Método estadístico para clasificar eventos en categorías binarias o múltiples.

Se menciona además el aprendizaje por ensamble, que combina varios modelos para mejorar la precisión en escenarios complejos como el análisis del crimen urbano.

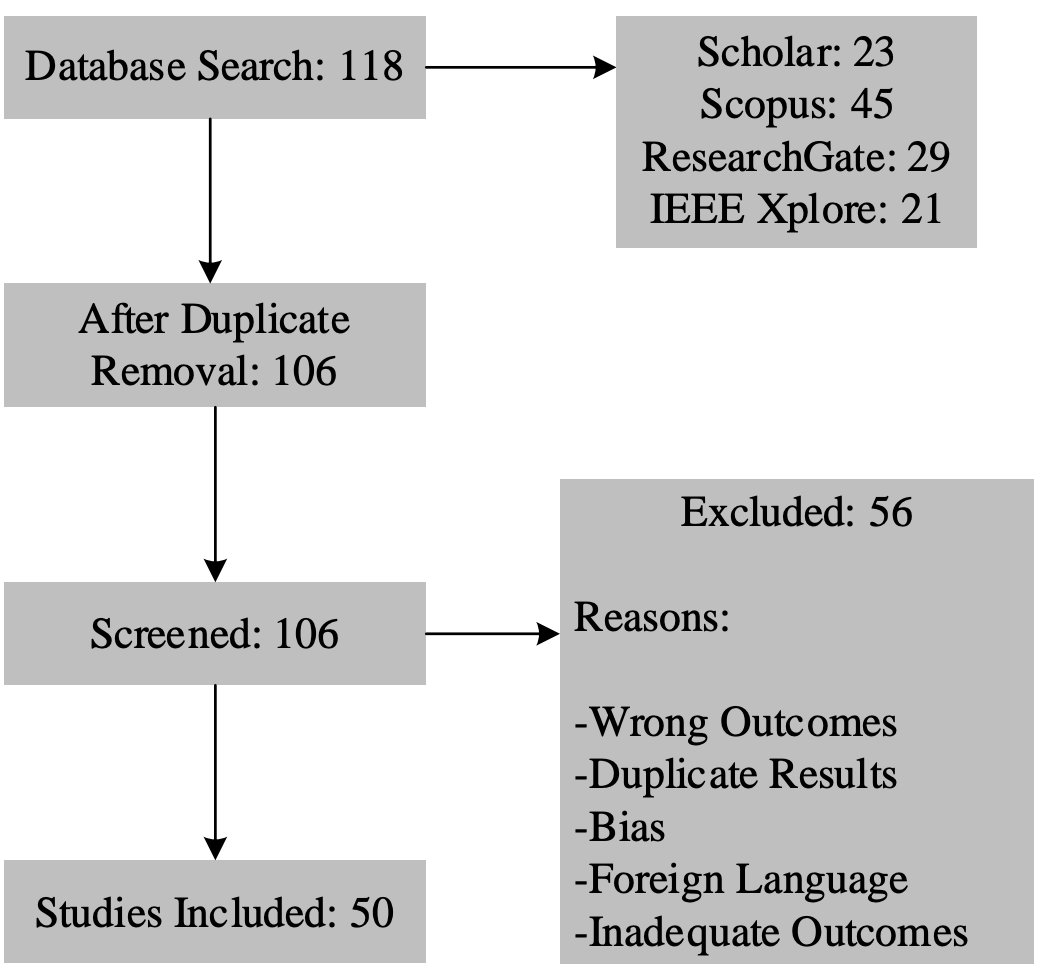


**Figura 3.5:** Esquema general de un modelo de aprendizaje automático

**B) Metodología**

La investigación siguió un enfoque cualitativo, centrado en una revisión sistemática de literatura. Para ello, se definió una estrategia de búsqueda que abarcó bases de datos académicas como Google Scholar, Scopus, ResearchGate e IEEE Xplore.

* **Fase 1:** Definición de criterios de inclusión (estudios recientes, revisados por pares y centrados en la aplicación de ML al crimen).
* **Fase 2:** Exclusión de artículos duplicados, sin resultados claros o fuera del enfoque temático.
* **Fase 3**: Revisión de 51 estudios seleccionados.
* **Fase 4:** Análisis comparativo de los algoritmos utilizados.
* **Fase 5:** Identificación de retos comunes en la implementación real de estas soluciones tecnológicas.



**Figura 3.6:** Esquema de la estrategia de búsqueda y de la inclusión/exclusión de estudios

**Utilidad del artículo para el proyecto de tesis**

Gracias al aporte de este artículo, se puede comprender que el uso del aprendizaje automático es una herramienta fundamental para predecir delitos y fortalecer la prevención del crimen, posicionándose como un recurso valioso dentro de las estrategias de seguridad ciudadana. A través de un análisis comparativo de 51 estudios académicos, se identificaron los principales algoritmos de Machine Learning aplicados a este campo para recursos policiales. En este artículo, utilizaré los principales algoritmos de Machine Learning identificados, tales como Random Forest, los cuales han demostrado mayor precisión en la predicción del delito. Estos modelos me permitirán facilitar la toma de decisiones estratégicas en seguridad, aprovechar los patrones históricos de criminalidad y optimizar la asignación de recursos policiales. Además, se utilizarán técnicas para validar los modelos en escenarios reales, con el objetivo de mejorar la precisión en la predicción de crímenes y optimizar la cobertura del patrullaje preventivo.

### 3.1.6 Crime Sensing With Big Data: The Affordances and Limitations of Using Open-source Communications to Estimate Crime Patterns

### Detección de delitos con macrodatos: posibilidades y limitaciones del uso de comunicaciones de código abierto para estimar patrones delictivos

(Williams et al.)

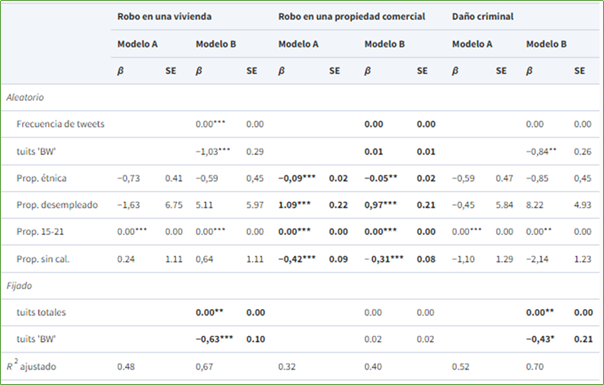
**(DOI: 10.1093/bjc/azw031)**

**Citado 10 veces en Scopus**

En el presente artículo los autores informan sobre un experimento metodológico con Big data en el campo de la criminología. En particular, proporciona un examen crítico basado en datos de las posibilidades y limitaciones de las comunicaciones de código abierto recopiladas de las interacciones de las redes sociales para el estudio del crimen y el desorden. El experimento realizado fue de naturaleza exploratoria y utilizó métodos nacientes de criminología computacional para recolectar, transformar, vincular y analizar éticamente 'grandes datos sociales' para abordar el problema clásico de la estimación de patrones delictivos. Los resultados presentados forman una base preliminar para la discusión crítica de estas nuevas formas de datos y para el posterior análisis de confirmación que se llevará a cabo. El objetivo del experimento fue construir modelos estadísticos de Big data que desarrollen un trabajo predictivo previo utilizando las redes sociales. Por ejemplo, Tumasjan et al. (2010) midió el sentimiento de Twitter en relación con los candidatos en las elecciones generales alemanas y concluyó que esta fuente de datos era tan precisa para predecir los patrones de votación como las encuestas. Asur y Huberman (2010) correlacionaron la frecuencia de las publicaciones y el sentimiento relacionado con las películas en Twitter con sus ingresos, afirmando que este método de predicción era más preciso que el de la Bolsa de Valores de Hollywood. Sakaki et al.(2010)descubrió que el análisis de los datos de Twitter produjo estimaciones de los epicentros de los terremotos con mayor precisión que los métodos de sensores geológicos convencionales. Estos estudios ilustran cómo las redes sociales generan datos socialmente relevantes "que ocurren naturalmente" que se pueden usar para complementar y aumentar los datos curados convencionales para estimar la ocurrencia de fenómenos fuera de línea. En nuestro experimento, llevamos a cabo un análisis ecológico del crimen en Londres utilizando datos de Twitter como predictor para probar la hipótesis de que los tweets relacionados con el crimen y el desorden están asociados con las tasas reales de crimen policial. Nuestros resultados proporcionan evidencia tentativa de que los modelos estadísticos basados ​​en datos de redes sociales pueden proporcionar una fuente alternativa de información sobre el problema de estimación del patrón delictivo. Este documento se suma a la base de evidencia y al debate en el campo emergente de la criminología computacional de cuatro maneras: (1) estima la utilidad de los datos de las redes sociales para explicar la variación en los patrones delictivos fuera de línea y compara los resultados con indicadores convencionales (variables del censo); (2) proporciona la primera evidencia de la estimación de patrones delictivos fuera de línea utilizando una medida de ventanas rotas encontradas en el contenido textual de las comunicaciones de las redes sociales; (3) prueba específicamente si el sesgo presente en las percepciones fuera de línea y los informes de delincuencia y desorden (que se encuentran entre las áreas de baja y alta delincuencia) está presente en las redes sociales; y (4) utiliza los resultados de estos experimentos para participar críticamente en los debates sobre big data y estimación del crimen.

**A)Resultados**

La frecuencia de los tuits se asoció positivamente con el robo en una vivienda, los daños criminales, la violencia contra la persona y el robo en tiendas, lo que corrobora trabajos anteriores que argumentan que los marcadores de geolocalización en los datos de Twitter son útiles para estimar los patrones delictivos (Bendler et al. 2014; Malleson y Andresen 2015 ) . ). Al igual que Malleson y Andresen, este estudio encuentra la relación positiva entre la frecuencia de las publicaciones en Twitter y la violencia contra la persona cuando se eliminan las posibles fuentes de sesgo mediante el control de las características ecológicas estables (observadas y no observadas). Estos resultados contradicen el trabajo de Bendler et al. (2014) quienes encontraron una relación negativa entre la frecuencia de tuits y la violencia. Sin embargo, en los modelos que toman el mes y no la hora como escala temporal, es probable que la frecuencia de los tweets actúe como un indicador de la densidad de población y no de la población móvil. Esto explicaría la relación positiva con los delitos que tienden a ocurrir en ausencia de transeúntes (robo y daños criminales). Como la frecuencia de los tweets no es una variable clave de interés en este artículo, no es una deficiencia fundamental de este estudio exploratorio.



*Tabla 2 . Modelos de efectos aleatorios y fijos para todos los distritos*

*Fuente:*  (Williams et al, 2017)

**Utilidad del artículo para el proyecto de tesis**

Este artículo aporta significativamente a nuestro proyecto de tesis al proporcionar un enfoque metodológico para el uso de Big Data en la criminología computacional. Se enfoca en cómo los modelos estadísticos basados en Big Data pueden ser útiles para predecir patrones delictivos. Además, los datos de redes sociales como Twitter se consideran como una herramienta adicional para analizar patrones de crimen y desorden social, generando datos socialmente relevantes. Esto es clave para abordar la prevención del crimen y mejorar la asignación de recursos policiales, focalizando los esfuerzos en las zonas de mayor riesgo.

La metodología presentada permite complementar los métodos tradicionales con los modelos predictivos de redes sociales, lo que contribuye a una gestión eficiente de recursos en el ámbito de seguridad pública. Además, se incorporan modelos estadísticos de Big Data para desarrollar un trabajo predictivo previo utilizando las redes sociales como predictor, lo que facilita la estimación de las tasas reales de crimen policial. Este artículo establece que el análisis de redes sociales puede ser una herramienta eficaz para predecir delitos en áreas de alta reincidencia delictiva y para reducir la criminalidad.

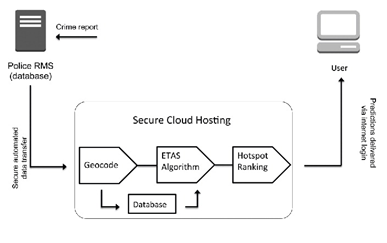
### 3.1.7. Randomized controlled field trials of predictive policing Ensayos de campo controlados aleatorios de vigilancia predictiva

(Mohler et al., 2015)

**(DOI: 10.1080/01621459.2015.1077710)**

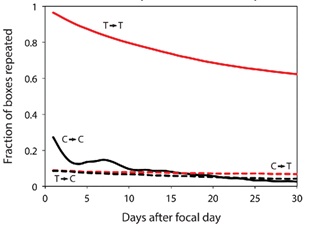
**Citado 8 veces en Scopus**

En este resumen, los autores indican que la concentración de los recursos policiales en puntos críticos de delincuencia estables ha demostrado ser eficaz para reducir la delincuencia, pero se desconoce hasta qué punto la policía puede interrumpir los puntos críticos de delincuencia que cambian dinámicamente. La policía debe poder anticipar la ubicación futura de los puntos críticos dinámicos para interrumpirlos. Los autores informan los resultados de dos ensayos controlados aleatorios de predicción de delitos de secuencia de réplicas de tipo epidémico (ETAS). Investigaron hasta qué punto los modelos ETAS de riesgo delictivo a corto plazo superan las mejores prácticas existentes de mapas de puntos críticos producidos por analistas de delitos dedicados, los oficiales de policía en el campo pueden patrullar dinámicamente los puntos críticos previstos dados los recursos limitados, y el crimen puede reducirse mediante algoritmos policiales predictivos bajo limitaciones realistas de recursos de aplicación de la ley. Si bien los experimentos anteriores de control de puntos críticos corrigen los puntos críticos de tratamiento y control durante todo el período experimental, utilizaron un diseño experimental novedoso para permitir que los puntos críticos de tratamiento y control cambien dinámicamente durante el transcurso del experimento. Los resultados muestran que los pronósticos ETAS predicen entre 1,4 y 2,2 veces más delitos en comparación con un analista de delitos dedicado que utiliza la inteligencia criminal existente y la práctica de mapeo de puntos críticos. Las patrullas policiales que utilizaron pronósticos ETAS llevaron a una reducción promedio del 7,4 % en el volumen de delitos en función del tiempo de patrulla, mientras que las patrullas basadas en las predicciones de los analistas no mostraron un efecto significativo.

****

***Figura 44 .*** *El modelo ETAS*

***Fuente:***  (Mohler et al, 2018)



***Figura 46 .*** *Fracción de cuadros de predicción repetidos en el día N después de un tratamiento focal o un día de control*

***Fuente:***  (Mohler et al, 2018)

**Utilidad del artículo para el proyecto de tesis**

Este artículo aportará significativamente a nuestro proyecto de tesis, ya que evaluará la implementación de la predicción de delitos de secuencia de réplicas de tipo epidémico (ETAS), un enfoque innovador para predecir patrones delictivos. En el artículo, se menciona que los modelos ETAS de riesgo delictivo a corto plazo superan las mejores prácticas existentes de mapas de puntos críticos, lo que es fundamental para mejorar la asignación de recursos policiales en áreas de alto riesgo. Esto permitirá anticipar de manera más efectiva los puntos críticos de criminalidad y reducir la criminalidad de manera eficiente.

Además, los resultados del artículo indican que los modelos ETAS predicen entre 1,4 y 2,2 veces más delitos en comparación con las prácticas actuales, lo que proporciona evidencia sólida de la efectividad de los algoritmos policiales predictivos. También se destaca que las patrullas basadas en pronósticos ETAS llevaron a una reducción promedio del 7,4 % en el volumen de delitos, un dato que será clave para mejorar la vigilancia predictiva y la gestión eficiente de recursos en la prevención de delitos.

### 3.1.8. A study on the application of data mining-based crime

### prediction models in criminal justice

### Un estudio sobre la aplicación de modelos de predicción de delitos basados ​​en minería de datos en la justicia penal.

### (Gonzalez & LeBoulluec, 2019).

**(DOI: 10.61091/jcmcc123-40)**

**Citado 25 veces en Scopus**

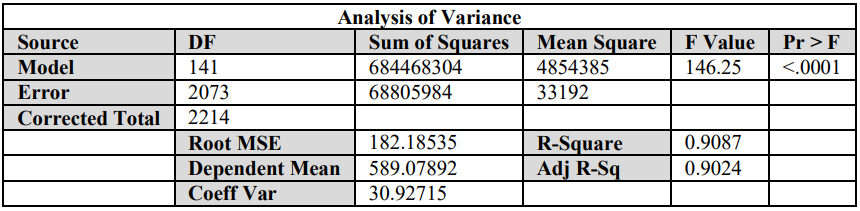
En el presente artículo, se explica que el objetivo de la predicción de crímenes es anticipar el lugar y el momento en que ocurrirán, utilizando algoritmos de regresión basados en aprendizaje automático. Este enfoque ha sido adoptado por diversas instituciones encargadas de la seguridad pública para mejorar la asignación de recursos.

En este estudio, se utilizan cuatro modelos de regresión: Regresión Lineal Múltiple, Bosque Aleatorio, Regresión Bayesiana y Redes Neuronales, con el Bosque Aleatorio mostrando el mejor rendimiento al predecir crímenes violentos en Lima, obteniendo un valor de R² cercano a 1. El artículo destaca tres puntos clave sobre estos modelos.

Primero, cómo los supuestos en los que se basan los algoritmos y los datos de entrenamiento afectan la precisión de las predicciones.

Segundo, la necesidad de más estudios rigurosos para evaluar el impacto de estos modelos en la distribución de recursos policiales.

Tercero, la capacidad de estos algoritmos de predecir zonas de alta incidencia delictiva sin perjudicar la efectividad de los agentes de policía.



***Tabla 1:*** *ANOVA para el modelo de regresión lineal múltiple con 141 variables predictoras*

***Fuente:***  *Gonzalez, J. J., & LeBoulluec, A. (2019))*

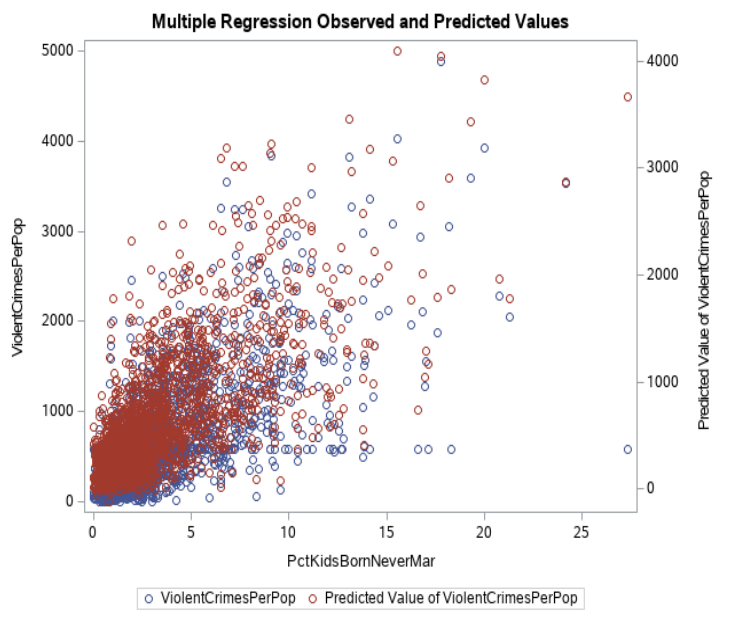
1. **Modelo Subyacente de estudio Comparativo de Algoritmos de Regresión**

Consiste en un ciclo de actividades y puntos de decisión: recopilación de datos, análisis, operaciones policiales, respuesta criminal y regreso a la recopilación de datos. En cada etapa del ciclo, se toman decisiones sobre, por ejemplo, los tipos de datos a recopilar como ingreso per cápita, nivel educativo, porcentaje de desempleo, número de inmigrantes, ubicación geográfica y frecuencia de crímenes violentos. También se decide la frecuencia con la que se recopilan y actualizan los datos, las herramientas analíticas a emplear, como los modelos de regresión supervisada (como Regresión Lineal Múltiple, Bosque Aleatorio, Redes Neuronales y Regresión Bayesiana), y las variables dependientes en las que centrarse, tales como patrones espaciales y temporales de los crímenes, y factores socioeconómicos.

**B) Metodología**

Los autores desarrollaron su investigación utilizando técnicas de minería de datos y aprendizaje automático, particularmente con modelos de regresión supervisada. Los pasos principales de la metodología incluyen:

1. **Recopilación de Datos Socioeconómicos y de Criminalidad**:  
   Se utilizaron datos históricos sobre crímenes violentos y características sociodemográficas de las áreas, como ingreso per cápita, nivel educativo, tasa de desempleo, y porcentaje de inmigrantes. Estos datos permiten identificar patrones de criminalidad y predecir la ocurrencia futura de crímenes violentos en zonas específicas.
2. **Modelado con Modelos de Regresión Supervisada**:  
   Se aplicaron modelos de regresión supervisada, como Regresión Lineal Múltiple, Random Forest, Regresión Bayesiana y Redes Neuronales, para predecir la cantidad de crímenes violentos en función de factores socioeconómicos y otras variables clave. El modelo Random Forest demostró ser el más efectivo, alcanzando un R 2 de 0.999999.
3. **Optimización del Modelo**:  
   Se utilizó la optimización de hiperparámetros mediante GridSearchCV para mejorar el desempeño de los modelos, minimizando errores como el MSE, RMSE y MAE, y asegurando la precisión de las predicciones realizadas.
4. **Evaluación y Comparación de Resultados**:  
   Los resultados obtenidos de los modelos fueron evaluados utilizando métricas como R 2, MSE, RMSE y MAE. Además, se compararon los resultados de Random Forest, XG Boost y otros modelos de regresión para determinar cuál ofrecía las predicciones más precisas para la ocurrencia de crímenes violentos basados en factores socioeconómicos.

****

***Figura 2:*** *Regresión Múltiple - Crimen Observado y Predicho vs PctKidsBornNeverMar.*

***Fuente:***  *Gonzalez, J. J., & LeBoulluec, A. (2019))*

**Utilidad del artículo para el proyecto de tesis**

Por medio de este artículo de investigación, se tomará en cuenta la metodología de predicción de crímenes utilizando modelos de regresión supervisada, como Regresión Lineal Múltiple, Random Forest, Redes Neuronales y Regresión Bayesiana, los cuales analizan datos históricos de crímenes violentos y características sociodemográficas de las áreas, como ingreso per cápita, nivel educativo, desempleo y porcentaje de inmigrantes, para hacer predicciones precisas sobre la ocurrencia de crímenes violentos. Este enfoque permite identificar las áreas más vulnerables a crímenes violentos, lo que facilita la asignación eficiente de recursos y la toma de decisiones estratégicas para combatir la criminalidad. Además, se considera la optimización de los modelos mediante la optimización de hiperparámetros utilizando modelos de regresión: Regresión Lineal Múltiple, Bosque Aleatorio, Regresión Bayesiana y Redes Neuronales, lo cual mejora la precisión de las predicciones, permitiendo predecir con mayor exactitud las áreas y momentos de mayor riesgo. La integración de factores socioeconómicos y demográficos ayuda a priorizar las zonas de mayor incidencia delictiva y optimizar las políticas públicas de seguridad, lo que es fundamental para la prevención de crímenes y el fortalecimiento de la seguridad en las comunidades.

### 3.1.9 Improving the Fairness of Deep-Learning, Short-term Crime Prediction with Under-reporting-aware Models

### Mejorar la imparcialidad de la predicción de delitos a corto plazo mediante aprendizaje profundo con modelos que tienen en cuenta la falta de denuncias

(Wu & Frias-Martinez, 2024)

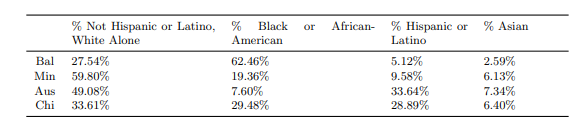
**(DOI: 10.48550/arXiv.2406.04382)**

**Citado 5 veces en Scopus**

En el presente artículo, se explora cómo mejorar la equidad en la predicción de crímenes a corto plazo mediante el uso de aprendizaje profundo, enfocándose en los sesgos causados por el subregistro de crímenes en los datos históricos. El objetivo es desarrollar un modelo predictivo que no solo corrija estos sesgos, sino que también mantenga la precisión de las predicciones. Para ello, los autores proponen una arquitectura innovadora que combina métodos de preprocesamiento e inprocesamiento, integrando un mecanismo de convolutional gate para ajustar las predicciones basadas en las tasas de reporte de crímenes.

El estudio se lleva a cabo en cuatro ciudades de EE. UU. (Baltimore, Minneapolis, Austin y Chicago), utilizando datos públicos sobre crímenes y movilidad humana. Los resultados muestran que el modelo propuesto mejora significativamente la equidad de las predicciones en comparación con modelos tradicionales sin corrección de sesgo, aunque se observa una ligera disminución en la precisión.

El artículo subraya tres puntos clave: el uso de modelos de aprendizaje profundo para corregir sesgos en los datos de crímenes, la importancia de mejorar la equidad en los sistemas predictivos, especialmente en lo relacionado con la asignación de recursos policiales, y el desafío de equilibrar precisión y equidad en la predicción de crímenes a corto plazo.



***Tabla 1****: Porcentaje de población por raza y etnia en las cuatro ciudades según la Encuesta Comunitaria Americana (estimaciones de 5 años del ACS 2019).*

***Fuente:*** *American Community Survey (2019 ACS 5-year estimates).*

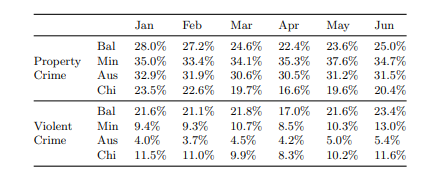
**Modelo Subyacente de Estudio sobre Predicción de Crímenes a Corto Plazo y Equidad en el Aprendizaje Profundo**

El modelo de estudio se basa en un ciclo de actividades en el cual se recopilan datos sobre crímenes y factores socioeconómicos, se analizan patrones históricos y se realizan predicciones sobre la ocurrencia de crímenes en el futuro. En cada fase del ciclo, se toman decisiones clave, como qué datos recopilar (por ejemplo, tasas de criminalidad, tasas de movilidad humana, características socioeconómicas), qué herramientas analíticas utilizar (como el aprendizaje profundo) y cómo corregir los sesgos derivados del subregistro de crímenes en los datos. El objetivo es crear un modelo que no solo sea preciso, sino también justo, al corregir los sesgos presentes en los datos de crímenes, que a menudo afectan desproporcionadamente a los grupos minoritarios.

**B) Metodología**

Los autores desarrollan su investigación utilizando aprendizaje profundo para predecir crímenes a corto plazo y mejorar la equidad en las predicciones al abordar el subregistro de crímenes. La metodología empleada se describe a continuación:

1. **Recopilación de Datos**: Se utilizan datos históricos de crímenes y patrones de movilidad humana basados en grandes bases de datos de teléfonos móviles. Los factores socioeconómicos, como la pobreza, la tasa de desempleo y el porcentaje de inmigrantes, también son incluidos en los datos, ya que se ha demostrado que influyen en la ocurrencia y el reporte de crímenes.
2. **Modelado con Aprendizaje Profundo**: Se aplican redes neuronales profundas, utilizando un modelo que incorpora un mecanismo de corrección de sesgo, denominado "convolutional gate". Este modelo ajusta las predicciones para corregir los efectos del subregistro de crímenes, un problema conocido que afecta la equidad en las predicciones. El modelo se entrena para predecir no sólo los crímenes reportados, sino también para estimar el número real de crímenes que ocurrieron, independientemente de si fueron reportados o no.
3. **Optimización del Modelo**: Se optimizan los hiperparámetros del modelo para maximizar la precisión de las predicciones y mejorar la equidad en las estimaciones de crímenes. Se emplean técnicas de ajuste de hiperparámetros y regularización para minimizar el error y asegurar que el modelo sea capaz de corregir los sesgos inherentes a los datos.
4. **Evaluación y Comparación de Resultados**: Los resultados de los modelos se evalúan utilizando métricas como la precisión (F1), la equidad (medida a través de la estadística de paridad y los errores de predicción), y la mejora en la equidad respecto a modelos sin corrección de sesgos. Se comparan los modelos propuestos con enfoques previos que no consideran el subregistro de crímenes, demostrando que el modelo propuesto mejora la equidad de las predicciones sin comprometer significativamente la precisión.



***Tabla 2****: Densidad mensual de ocurrencia de crímenes en las cuatro ciudades en 2020: Baltimore (Bal), Minneapolis (Min), Austin (Aus) y Chicago (Chi) de enero a junio.*

**Utilidad del artículo para el proyecto de tesis**

Este artículo aporta a nuestro proyecto de tesis al proporcionar una arquitectura innovadora que utiliza modelos de aprendizaje profundo para mejorar la equidad en la predicción de crímenes a corto plazo. El artículo propone un enfoque para corregir los sesgos causados por el subregistro de crímenes en los datos históricos, lo que nos permitirá optimizar las predicciones de crímenes y asegurar que los modelos predictivos sean más precisos. Esta arquitectura innovadora incluye un mecanismo de convolutional gate, que ajusta las predicciones basadas en las tasas de reporte de crímenes, lo que ayudará a mejorar la precisión y la equidad en las predicciones de crímenes violentos. En el contexto de nuestra tesis, utilizaremos la metodología del artículo para ajustar los modelos predictivos, aprovechando el uso de modelos de aprendizaje profundo para identificar patrones delictivos en los datos históricos de crímenes y características socioeconómicas. La aplicación de métodos de corrección de sesgos y el uso del mecanismo de convolutional gate nos permitirá optimizar la asignación de recursos policiales y mejorar la prevención de crímenes en las zonas más vulnerables.

### 3.1.10 A Comparison Between Human and Generative AI Decision-Making Attributes in Complex Health Services

### Comparación entre los atributos de toma de decisiones humanas y de la IA Generativa en servicios de salud complejos

### (Doraiswamy, n.d.2024)

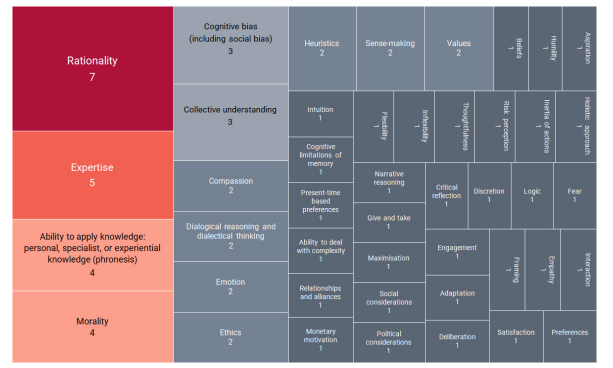
**DOI: 10.48550/arXiv.2505.08360**

**Citado 5 veces en Scopus**

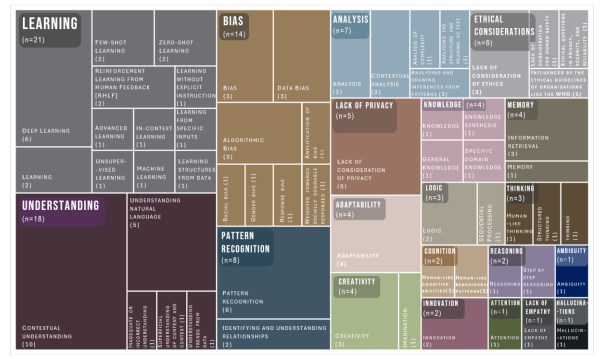
El artículo sobre la comparación entre los atributos de toma de decisiones humanas y de la IA Generativa en servicios de salud complejos es altamente relevante para el proyecto de tesis, ya que proporciona un análisis profundo de las fortalezas y limitaciones de ambos agentes en contextos de decisión compleja. El enfoque del artículo permite identificar atributos únicos, útiles y contextuales que influyen en la toma de decisiones, ayudando a fundamentar la importancia de considerar tanto las capacidades humanas (como la moralidad, empatía y juicio ético) como las ventajas de la IA (como el procesamiento rápido de datos y el reconocimiento de patrones).

En el marco de la tesis, esta investigación proporciona bases conceptuales para evaluar cómo humanos e IA pueden cooperar, competir o converger en la toma de decisiones en los servicios de salud, permitiendo así diseñar sistemas híbridos que optimicen la eficiencia y calidad de las decisiones clínicas, administrativas y regulatorias. Asimismo, el análisis de atributos contextuales y los desafíos éticos ofrece un marco para desarrollar estrategias de formación y diseño de sistemas que potencien las habilidades humanas únicas, mientras integran de manera efectiva la IA, garantizando roles significativos para ambos en el futuro del sector salud.

De esta forma, el artículo fundamenta la relevancia y viabilidad de explorar la complementariedad humano-IA, aportando una perspectiva crítica para la mejora continua en la gestión de decisiones complejas en el sector de salud.



***Figura 1:***  *Muestra los atributos que influyen en la toma de decisiones humanas en servicios de salud complejos, según lo reportado en la literatura.*

**

***Figura 2:*** *Muestra los atributos que influyen en la toma de decisiones de la IA Generativa en servicios de salud complejos, según lo reportado en la literatura.*

**A) Modelo Comparativo de Atributos en la Toma de Decisiones en Servicios de Salud Complejos**

El estudio se basa en la comparación de dos conjuntos de atributos que influyen en la toma de decisiones: los humanos y la inteligencia artificial generativa (IA Generativa). Para ello, se utilizan dos revisiones previas que identifican y clasifican estos atributos según su utilidad, impacto y contexto. El análisis contempla atributos únicos, compartidos y contrapuestos entre humanos y IA, evaluando cómo cada grupo procesa información, aplica juicio ético, maneja la complejidad y responde a sesgos o limitaciones cognitivas. El objetivo es comprender cómo estas diferencias pueden influir en la efectividad y calidad de las decisiones en entornos de salud complejos, y qué modelos de interacción (competencia, cooperación o convergencia) pueden surgir para optimizar la toma de decisiones futuras.

**B) Metodología** Los autores realizan una investigación comparativa basada en revisiones de literatura para identificar y clasificar los atributos que influyen en la toma de decisiones en servicios de salud complejos, tanto en humanos como en IA Generativa.

* **Recolección y análisis de atributos:** Se utilizan dos revisiones previas: una exploratoria para atributos humanos y otra rápida para atributos de IA Generativa, identificando características útiles, perjudiciales y contextuales.
* **Clasificación y comparación:** Los atributos se agrupan en únicos, compartidos o contrapuestos entre humanos e IA, y se analizan en función de su impacto en la toma de decisiones efectivas.
* **Evaluación narrativa:** Se realiza una comparación narrativa para determinar las fortalezas y debilidades relativas, y para explorar posibles modelos de interacción futura: competencia, cooperación o convergencia.
* **Síntesis de hallazgos:** Se sintetizan los resultados para entender cómo cada conjunto de atributos puede complementar al otro en la toma de decisiones en contextos complejos de salud.

**Utilidad del artículo para el proyecto de tesis**

Este artículo contribuye significativamente a nuestro proyecto de tesis sobre la implementación de un aplicativo predictivo de criminalidad basado en Machine Learning, al destacar la relevancia de construir algoritmos de clasificación y agrupación modelos predictivos para la prevención de delitos. Se enfatiza la importancia de considerar el espacio de tiempo para contextualizar correctamente cuándo y dónde pueden ocurrir los delitos, mejorando así la capacidad de anticipación del sistema. La aplicación de técnicas estadísticas como la regresión lineal demuestra ser efectiva para modelar tendencias delictivas a partir de datos históricos, permitiendo realizar proyecciones confiables. Este enfoque, al integrar modelos de IA Generativa, brinda un marco robusto para que nuestro modelo pueda identificar con mayor precisión las zonas de alto riesgo y los períodos críticos, lo que es fundamental para tomar decisiones preventivas. Además, al integrar estos modelos, diseñar sistemas híbridos es posible, lo cual mejora la eficiencia en la asignación de recursos, focalizando la acción policial en los sectores con mayor probabilidad de ocurrencia de delitos. Esto no solo optimiza la gestión de los recursos disponibles, sino que también potencia la efectividad de las estrategias de seguridad y prevención.

En consecuencia, la metodología presentada fortalece nuestra tesis al aportar fundamentos y técnicas que pueden ser adaptadas para el contexto específico de Lima Metropolitana, contribuyendo a la mejora de la seguridad ciudadana a través de sistemas predictivos más precisos y eficientes. Además, el análisis de atributos contextuales ayuda a perfeccionar la toma de decisiones y a optimizar el uso de recursos.

### 3.1.11 Perceived Information Revisited II Information-Theoretical Analysis of Deep-Learning Based Side-Channel Attacks

### 

### Información percibida revisada II Análisis teórico de la información de ataques de canal lateral basados ​​en aprendizaje profundo

### (Akira Ito, Rei Ueno, Naofumi Homma2024)

### (DOI: 10.46586/tches.v2025.i1.450-474)

**Citado 1 vez en Scopus**

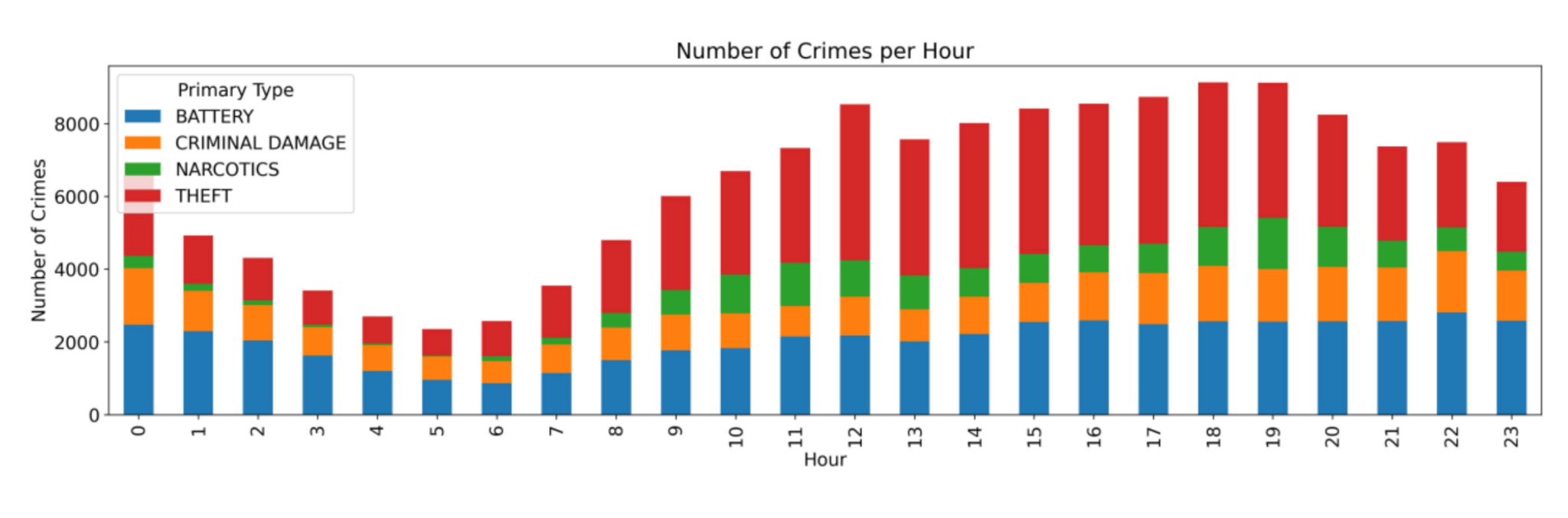
### En el presente artículo, los autores analizan el potencial del aprendizaje profundo (Deep learning) en la predicción del delito, evaluando modelos recientes y comparando su rendimiento frente a enfoques tradicionales de aprendizaje automático supervisado. Se resalta que el uso de redes neuronales profundas permite capturar relaciones complejas en los datos delictivos, incluyendo variables como ubicación, tiempo y características contextuales, lo cual es clave para una prevención más eficaz.

### Los autores proponen un marco experimental detallado donde evalúan distintos modelos predictivos con datos reales de criminalidad, identificando que los enfoques basados en deep learning superan en rendimiento a algoritmos clásicos como Random Forest o Naïve Bayes. Destacan también la necesidad de considerar la calidad del dataset, la representación temporal y la interpretabilidad del modelo, especialmente en aplicaciones de seguridad ciudadana.

### La investigación enfatiza la importancia de usar estos modelos no solo para anticipar delitos, sino también para asignar recursos estratégicamente, optimizando las acciones preventivas de las autoridades policiales mediante sistemas inteligentes de apoyo a la decisión.

### 

**Figura 1**: Número de crímenes por área



**Figura 2:** Número de crímenes por hora

**A) Referentes Teóricos**

El artículo se apoya en los fundamentos del aprendizaje profundo, una rama avanzada del Machine Learning, que emplea redes neuronales artificiales con múltiples capas ocultas capaces de detectar patrones complejos. También se hace mención a técnicas como:

* **Redes neuronales convolucionales (CNN):** Eficientes para extraer características espaciales de los datos.
* **Redes recurrentes (RNN) y LSTM:** Aptas para capturar secuencias y dependencias temporales en series de tiempo delictivas.
* **Modelos híbridos:** Que combinan diferentes arquitecturas para mejorar la capacidad de predicción.

Este enfoque se alinea con marcos de referencia actuales sobre inteligencia artificial aplicada a la seguridad pública, y demuestra cómo la tecnología puede integrarse en la gestión preventiva del crimen.

**B) Metodología**

Los autores implementaron un enfoque experimental, utilizando datasets reales de criminalidad urbana con variables geo espaciales y temporales. Las fases del estudio fueron:

* **Fase 1:** Preprocesamiento de datos y construcción de variables relevantes (fecha, lugar, tipo de crimen, frecuencia).
* **Fase 2:** Implementación de modelos predictivos con arquitecturas como CNN, RNN y LSTM.
* **Fase 3:** Entrenamiento de modelos con distintas combinaciones de hiper parámetros.
* **Fase 4:** Evaluación del rendimiento con métricas como precisión, recall y F1-score.
* **Fase 5:** Comparación con algoritmos tradicionales como Random Forest y análisis de su viabilidad operativa.

**Utilidad del artículo para el proyecto de tesis**

Este artículo contribuye directamente al desarrollo del proyecto de tesis al demostrar cómo la implementación de modelos de aprendizaje profundo (deep learning) mejora sustancialmente la predicción del delito, especialmente cuando se compara con enfoques clásicos de aprendizaje automático supervisado. Se destaca que estos modelos permiten una prevención más eficaz, el estudio pone en evidencia la importancia de considerar la calidad del dataset, así como la adecuada representación temporal y la interpretabilidad del modelo, factores cruciales en aplicaciones de seguridad ciudadana. Finalmente, el valor de utilizar estas tecnologías para asignar recursos estratégicamente, optimizando el despliegue policial en función de los patrones delictivos detectados. Se refuerza la idea de que los modelos de aprendizaje automático pueden constituirse en una herramienta efectiva de apoyo a la estrategia policial, facilitando una respuesta más rápida y eficiente ante la criminalidad, incluso en contextos donde los recursos disponibles son limitados.

### 3.1.12 Artificial intelligence & crime prediction: A systematic literature review

### Inteligencia artificial y predicción del delito: una revisión sistemática de la literatura

### (Fatima Dakalbab, et al.2023)

### (D[oi.org/10.1016/j.ssaho.2022.100342](https://doi.org/10.1016/j.ssaho.2022.100342))

**Citado 1 vez en Scopus**

### 

### En este artículo, los autores presentan una revisión sistemática de la literatura enfocada en la predicción del delito mediante inteligencia artificial (IA) y aprendizaje automático (Machine Learning), recopilando 120 investigaciones publicadas entre 2008 y 2021. La revisión cubre aspectos como los tipos de crimen analizados, los algoritmos utilizados, métricas de rendimiento, limitaciones y herramientas empleadas.

### 

### La mayoría de estudios evaluados emplearon algoritmos como Random Forest, Naïve Bayes y Árboles de Decisión. También se analizaron más de 40 datasets, siendo los más utilizados los de Chicago, India y Estados Unidos. Se señala que la elección del modelo depende en gran medida de las características del conjunto de datos, lo cual implica que no existe un algoritmo universalmente superior, sino que debe adaptarse al tipo de crimen y entorno.

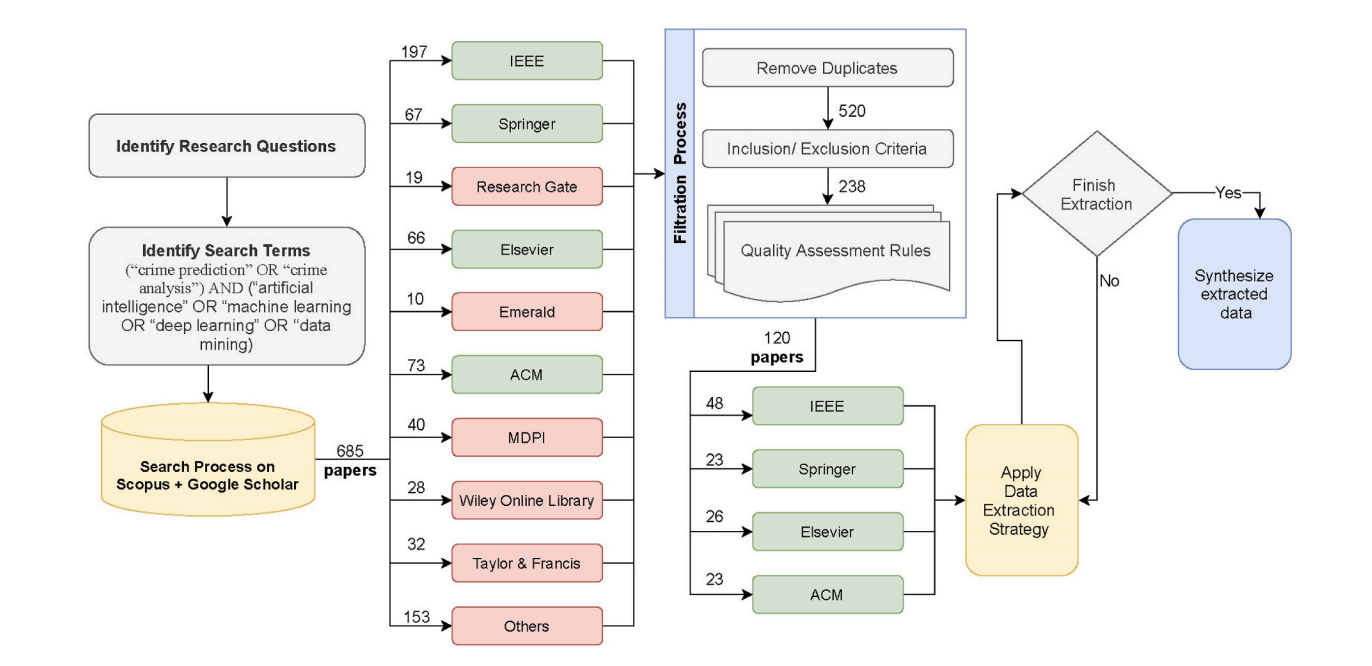
### 

### Los autores destacan que el uso de herramientas como Python, R, Weka y sistemas GIS ha facilitado el análisis de variables espaciales y temporales, optimizando así la detección de zonas calientes del crimen (hotspots). También se identifican retos como la disponibilidad de datos, el alto costo computacional de los modelos complejos y la necesidad de interpretabilidad, especialmente en contextos donde los resultados deben ser comprensibles para la toma de decisiones.

**A) Referentes Teóricos**

El estudio considera múltiples dimensiones para el análisis del crimen:

* **Análisis del crimen y el vecindario:** 43% de los artículos utilizaron densidad delictiva y demografía.
* **Análisis espacial:** 16% de los estudios aplicaron clustering o correlación espacial.
* **Análisis del comportamiento humano y redes sociales:** Algunos estudios utilizaron tweets y expresiones faciales para inferir riesgos.

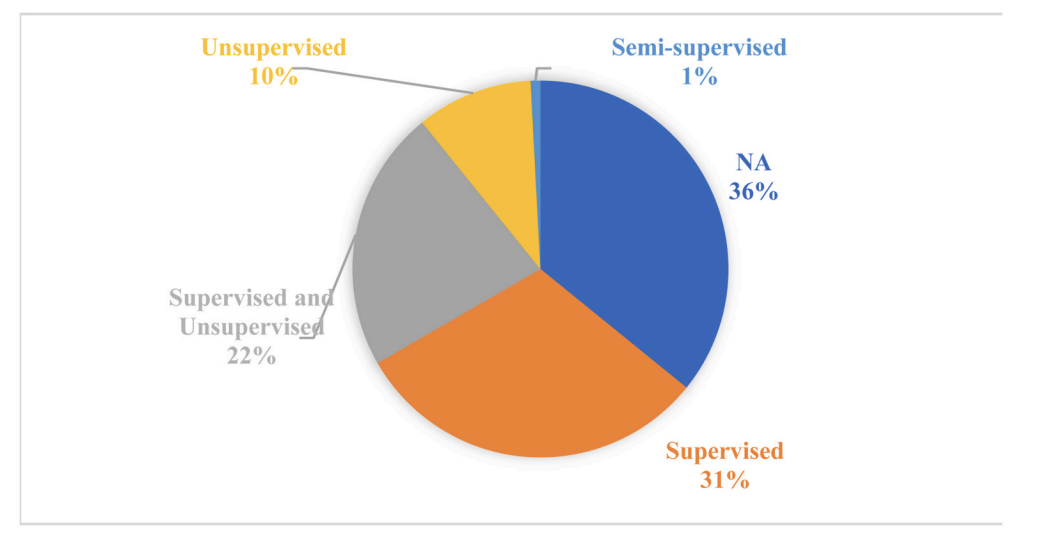


**Figura 1:** Modelo de identificación

**B) Metodología**

La investigación se desarrolló mediante una revisión sistemática de literatura, organizada en cuatro fases:

* **Fase 1:** Se definieron los objetivos del estudio y las preguntas de investigación relacionadas con tipos de crimen, algoritmos aplicados, métricas utilizadas y limitaciones encontradas.
* **Fase 2:** Se recopilaron artículos académicos desde bases de datos como Scopus, IEEE, Springer, ACM y Elsevier, seleccionando 120 estudios publicados entre 2008 y 2021.
* **Fase 3:** Se identificaron los crímenes más estudiados, algoritmos más utilizados, herramientas empleadas, y se evaluaron las métricas aplicadas y los principales desafíos técnicos.
* **Fase 4:** Se organizaron los hallazgos según las preguntas planteadas, permitiendo conocer el estado actual de la predicción del delito con aprendizaje automático y sus oportunidades de mejora.



**Figura 2:** Algoritmo de machine learning

**Utilidad del artículo para el proyecto de tesis**

Este artículo es de gran utilidad para el proyecto de tesis, ya que ofrece una visión amplia sobre la predicción del delito mediante inteligencia artificial (IA) y aprendizaje automático (Machine Learning). Gracias al análisis de más de 100 estudios, se destacan algoritmos como Random Forest, Naïve Bayes y Árboles de Decisión.

Además, se resalta la importancia del uso de herramientas como Python, R y GIS, esenciales para integrar variables espaciales y temporales que mejoren la precisión del modelo. Esta perspectiva respalda el desarrollo de un sistema que permita una asignación estratégica de recursos policiales, optimizando el patrullaje preventivo. Finalmente, el artículo subraya la necesidad de modelos con alta interpretabilidad, lo cual es clave en contextos de seguridad ciudadana, donde las decisiones deben ser comprendidas y justificadas ante autoridades y ciudadanía ya que sintetiza las técnicas de aprendizaje automático supervisado y su efectividad en la predicción del delito en diferentes contextos.

### 3.1.13 WordNet based Implicit Aspect Sentiment Analysis for Crime Identification from Twitter

### Análisis de sentimiento de aspecto implícito basado en WordNet para la identificación de delitos en Twitter

### (El Hannach, H., & Benkhalifa, M. 2018)

**(DOI:** [**10.14569/IJACSA.2018.091222**](https://dx.doi.org/10.14569/IJACSA.2018.091222)**)**

**Citado 21 veces en Scopus**

En el presente estudio, se propuso un modelo híbrido para el análisis de aspectos implícitos en el análisis de sentimientos en Twitter, aplicado a la detección de crímenes. La idea principal fue mejorar el proceso de clasificación mediante el uso de relaciones semánticas de WordNet y un esquema de ponderación de términos, para optimizar el proceso de detección y clasificación de oraciones relacionadas con crímenes implícitos.

La investigación se basó en tres fases principales: la detección de oraciones con aspectos implícitos (IASD), la identificación de estos aspectos (IAI) y esquema de ponderación TF-ICF de los sentimientos (positivo o negativo) en relación con cada aspecto. Los experimentos demostraron que el uso de relaciones de sinónimos de WordNet, combinado con el esquema de ponderación TF-ICF, mejoró el rendimiento de tres clasificadores populares: Naïve Bayes Multinomial (MNB), Máquina de Vectores de Soporte (SVM) y Random Forest (RF). Estos clasificadores fueron evaluados sobre tres conjuntos de datos relacionados con crímenes obtenidos de Twitter.

Se encontró que el modelo propuesto, al integrar sinónimos y relaciones definitorias de WordNet, permitió mejorar el rendimiento en la tarea de detección y clasificación de crímenes implícitos, especialmente cuando se utilizaron términos como verbos, que mostraron ser clave para representar mejor los aspectos implícitos en las oraciones.La investigación concluyó que el uso de este enfoque híbrido mejora la precisión y la fiabilidad de la clasificación de crímenes, especialmente en conjuntos de datos desbalanceados, lo cual es crucial para la predicción de delitos a partir de datos no estructurados de plataformas como Twitter.

**B) Metodología**

El estudio presenta un modelo híbrido para analizar aspectos implícitos en los sentimientos de tweets sobre crímenes utilizando aprendizaje automático supervisado. La metodología se divide en tres fases: detección de oraciones con aspectos implícitos (IASD), identificación de aspectos implícitos (IAI), y clasificación de sentimientos. Se extraen adjetivos y verbos usando relaciones semánticas de WordNet y un esquema de ponderación de términos (TF-ICF) en lugar de TF-IDF, mejorando el rendimiento en conjuntos de datos desbalanceados. Los clasificadores Naïve Bayes, SVM y Random Forest (RF) se utilizan para evaluar el modelo, demostrando que la integración de sinónimos y definiciones de WordNet mejora la precisión en la detección e identificación de aspectos implícitos y en la clasificación de sentimientos.

**C) Etapas**

Etapa 1: Recopilación de datos de Twitter.  
Etapa 2: Detección de oraciones con aspectos implícitos (IASD).

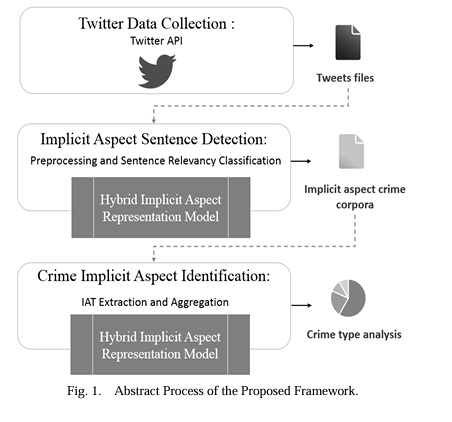
Etapa 3: Identificación de aspectos implícitos (IAI):

* Extracción de términos implícitos:
* Agregación de términos

Etapa 4: Análisis de sentimientos.

***Figura 1***

***Proceso Abstracto del Marco Propuesto.***

******

***Nota. Fuente: El Hannach, H., & Benkhalifa, M. (2018).***

**Utilidad del artículo para el proyecto de tesis**

Gracias a los aportes de este artículo, se comprende con mayor claridad que, para predecir delitos y fortalecer la seguridad pública, es fundamental aplicar el análisis de sentimiento de aspecto implícito (IASA) a partir de la información recopilada en Twitter. Este enfoque permite identificar patrones delictivos latentes en los mensajes publicados, resaltando elementos implícitos como adjetivos y verbos. En este sentido, el modelo propuesto, que incorpora relaciones semánticas a través de WordNet y un esquema de ponderación de términos, mejora significativamente la clasificación de oraciones vinculadas a actividades delictivas.

En ese sentido, este artículo se centra en conceptos clave como WordNet y sus relaciones semánticas, indispensables para enriquecer el vocabulario de delitos implícitos; el análisis de sentimiento basado en aspectos implícitos, que permite descubrir indicios criminales ocultos en los tuits; y el esquema de ponderación TF-ICF junto con los clasificadores Naïve Bayes, SVM y Random Forest, que elevan la precisión predictiva del modelo; todos ellos son elementos fundamentales que aplicaremos directamente en nuestra tesis.

### 3.1.14 Integrating Machine Learning Techniques for Enhanced Safety and Crime Analysis in Maryland

**Integración de Técnicas de Machine Learning para Mejorar el Análisis y la Seguridad Pública en Maryland**

(Bandpey, Z., Piri, S., & Shokouhian, M., 2025)  
 **(DOI:10.3390/app15094642)  
 Citado 1 vez en Scopus**

En el presente estudio, se propuso un modelo avanzado de Machine Learning (ML) para mejorar el análisis de criminalidad en Maryland, aprovechando técnicas sofisticadas adaptadas a las dinámicas urbanas, suburbanas y rurales del estado. El principal objetivo fue optimizar la identificación de patrones delictivos mediante la integración de modelos de ML como Random Forest, Gradient Boosting, XGBoost y métodos avanzados de ensamblaje. Estos modelos fueron afinados para abordar las variaciones locales, mejorando significativamente la capacidad de capturar las tendencias criminales de manera precisa.

La investigación se centró en tres fases clave: la recopilación de datos de informes policiales y bases de datos de crímenes, el procesamiento espacial y temporal de estos datos, y la implementación de los modelos de ML optimizados. La combinación de estos enfoques permitió una evaluación detallada de las dinámicas delictivas en diversas regiones de Maryland, proporcionando una base sólida para predecir las tendencias y la ubicación de los puntos calientes de crimen.

Los resultados mostraron que la integración de múltiples modelos de ML, combinados con un enfoque de validación rigurosa, mejoró notablemente el rendimiento en comparación con métodos tradicionales. Este modelo propuesto no solo mejora la precisión de la predicción de crímenes, sino también la capacidad de adaptar las estrategias de seguridad pública a las características únicas de cada región.

### B) Metodología

El estudio inicia con la recopilación de datos sobre crímenes y factores contextuales en Maryland, analizando las variaciones en áreas urbanas, suburbanas y rurales. Se emplean modelos de Machine Learning como Random Forest, Gradient Boosting y XGBoost para identificar patrones delictivos. Los datos son preprocesados para garantizar calidad, normalizándolos y eliminando valores atípicos. Posteriormente, se ajustan los modelos mediante validación cruzada para evitar sobreajustes y mejorar su precisión. Además, se realizan ajustes de hiperparámetros para optimizar el rendimiento. Se incluye el análisis de datos de redes sociales, especialmente de Twitter, usando la biblioteca Snscrape en Python. Este enfoque permite correlacionar las percepciones sociales con los patrones delictivos, proporcionando una visión integral para mejorar la seguridad pública en Maryland.

### C) Etapas

Etapa 1: Análisis de patrones delictivos.

Etapa 2: Preprocesamiento de datos y normalización.

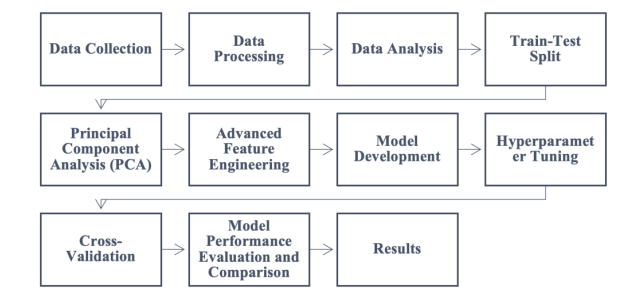
Etapa 3: Clasificación con técnicas de Machine Learning:

* Random Forest
* Gradient Boosting
* XGBoost
* Métodos de ensamblaje avanzados

Etapa 4: Análisis de datos de redes sociales mediante Snscrape.

***Figura 1***

*Diagrama de flujo para el desarrollo de un modelo predictivo en el análisis de delitos.*



***Nota. Fuente:*** *Bandpey, Z., Piri, S., & Shokouhian, M. (2025).*

**Utilidad del artículo para el proyecto de tesis**

Este artículo aporta significativamente a nuestro proyecto de tesis sobre la implementación de un modelo predictivo de criminalidad basado en Machine Learning, destacando la importancia de construir algoritmos de clasificación y agrupación de modelos predictivos para la prevención de delitos. Se resalta la relevancia de considerar el espacio y tiempo para contextualizar adecuadamente cuándo y dónde pueden ocurrir los crímenes, mejorando la capacidad de anticipación del sistema. La implementación de técnicas avanzadas de Machine Learning, como Random Forest, Gradient Boosting y XGBoost, demuestra ser eficaz para identificar patrones de criminalidad a partir de grandes volúmenes de datos históricos, lo que permite predecir de manera precisa los puntos calientes de crímenes futuros. Este enfoque brinda un marco robusto que optimiza la asignación de recursos policiales, permitiendo identificar las zonas de alto riesgo y los períodos críticos, claves para tomar decisiones preventivas más efectivas. Además, el uso de estos modelos proporciona una base sólida para asignar recursos de manera eficiente, focalizando la acción en las áreas con mayor probabilidad de ocurrencia de delitos. Este modelo no solo mejora la gestión de la seguridad pública, sino también la efectividad de las políticas de prevención. En consecuencia, la metodología presentada refuerza la capacidad de análisis predictivo, contribuyendo a la mejora de la seguridad ciudadana mediante sistemas más precisos y eficaces en el contexto de Maryland.

### 3.1.15 Multi-UAV Allocation Framework for Predictive Crime Deterrence and Data Acquisition (Miyano et al., 2020)

### Marco de Asignación de UAVs Múltiples para la Disuasión Predictiva del Crimen y Adquisición de Datos

### (DOI: [10.1016/j.iot.2020.100205](https://doi.org/10.1016/j.iot.2020.100205))

### Citado 15 veces en Scopus

### Los autores desarrollan un marco innovador para la predicción del delito y su prevención utilizando aprendizaje automático y vehículos aéreos no tripulados (UAVs), integrados en un sistema de vigilancia inteligente. Este sistema propone la asignación óptima de UAVs a zonas con alto riesgo delictivo en función de datos derivados del entorno, como nivel de iluminación, tránsito peatonal y sonidos ambientales.

### 

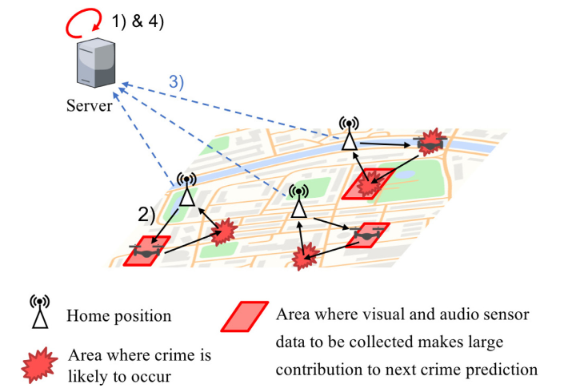
### El marco se basa en un modelo predictivo que guía el despliegue de los UAVs no solo para disuadir incidentes delictivos, sino también para recolectar datos históricos relevantes que retroalimentan y mejoran la eficacia futura del sistema. Esta estrategia permite una optimización continua en la asignación de recursos de vigilancia, lo cual se alinea con la necesidad de una estrategia policial más eficiente ante la escasez de personal.

**B) Metodología**

La metodología implementada por los autores se estructura en torno a un sistema compuesto por un servidor central que controla múltiples vehículos aéreos no tripulados (UAVs), los cuales se encargan de recolectar *datos históricos* en zonas urbanas específicas. Cada UAV está equipado con sensores visuales y acústicos que capturan información relevante del entorno, como la densidad peatonal, niveles de iluminación y sonidos anómalos.

* **Fase 1:** Diseño de arquitectura del sistema basado en sensores visuales y acústicos.
* **Fase 2:** Integración de variables como oscuridad, ruido y densidad peatonal para identificar zonas de alto riesgo.
* **Fase 3:** Aplicación de redes neuronales para la *predicción del delito* en bloques urbanos.
* **Fase 4:** Evaluación de la cobertura mediante dos métricas: disuasión y calidad del dato recolectado.
* **Fase 5:** Simulación realista en la ciudad de Chicago con UAVs comerciales.

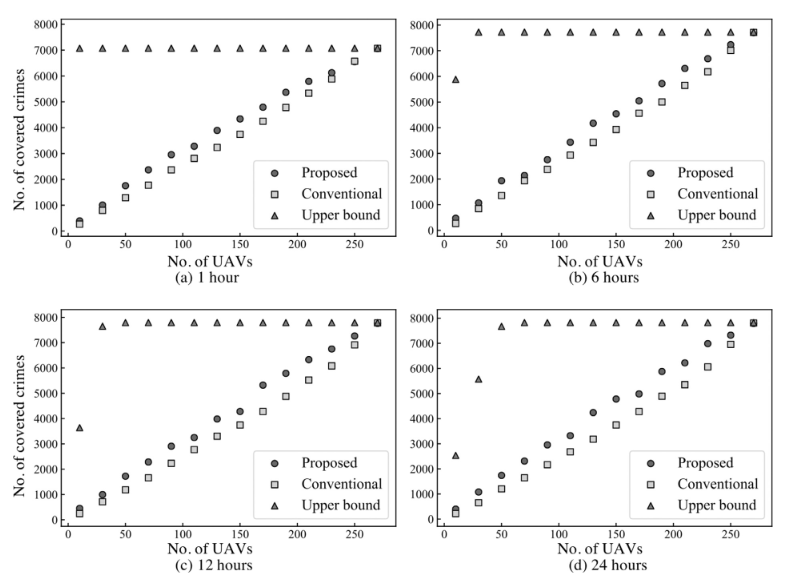
**Figura 1:** Arquitectura del sistema asumida en este estudio.



**C) Etapas**

* **Etapa 1:** Recolección de datos de zonas urbanas.
* **Etapa 2:** Entrenamiento del modelo con esos datos.
* **Etapa 3:** *Asignación de recursos* mediante rutas óptimas.
* **Etapa 4:** Medición del impacto y eficiencia energética del despliegue.

**Figura 2:** Número total de delitos cubiertos por vehículos aéreos no tripulados vs. número de vehículos aéreos no tripulados.



**Utilidad del artículo para el proyecto de tesis**

El modelo planteado en este estudio representa un referente metodológico clave para diseñar un *sistema predictivo* en contextos de escasez de personal o recursos aplicado a la distribución dinámica de UAVs en función de datos obtenidos en tiempo real, demuestra cómo la tecnología puede anticipar necesidades y mejorar la respuesta ante situaciones de alta variabilidad.

El uso de aprendizaje automático para la optimización y priorización de zonas críticas puede extrapolarse a la asignación de recursos médicos, donde se requiere anticipar picos de demanda y mejorar la predicción del delito. Asimismo, la integración de datos históricos permite mejorar continuamente la toma de decisiones, asegurando una respuesta más eficaz y preventiva en el ámbito de la seguridad reduciendo riesgos y mejorando la planificación operativa general.

### 3.1.16 Predictive criminology: a near future or a fiction in the making? (Pérez Salazar, 2024)

### La criminología predictiva: ¿un futuro próximo o una ficción en lontananza?

### (10.14718/NovumJus.2024.18.3.13)

### El artículo ofrece una profunda reflexión sobre la evolución del enfoque de criminología predictiva y su implementación en la seguridad ciudadana mediante el uso de inteligencia artificial y aprendizaje automático, modelos estadísticos multivariantes. El autor destaca cómo estos sistemas han sido implementados por la Policía Nacional de Colombia bajo el modelo AISec, el cual integra múltiples fuentes de datos históricos para anticipar comportamientos criminales, especialmente homicidios.

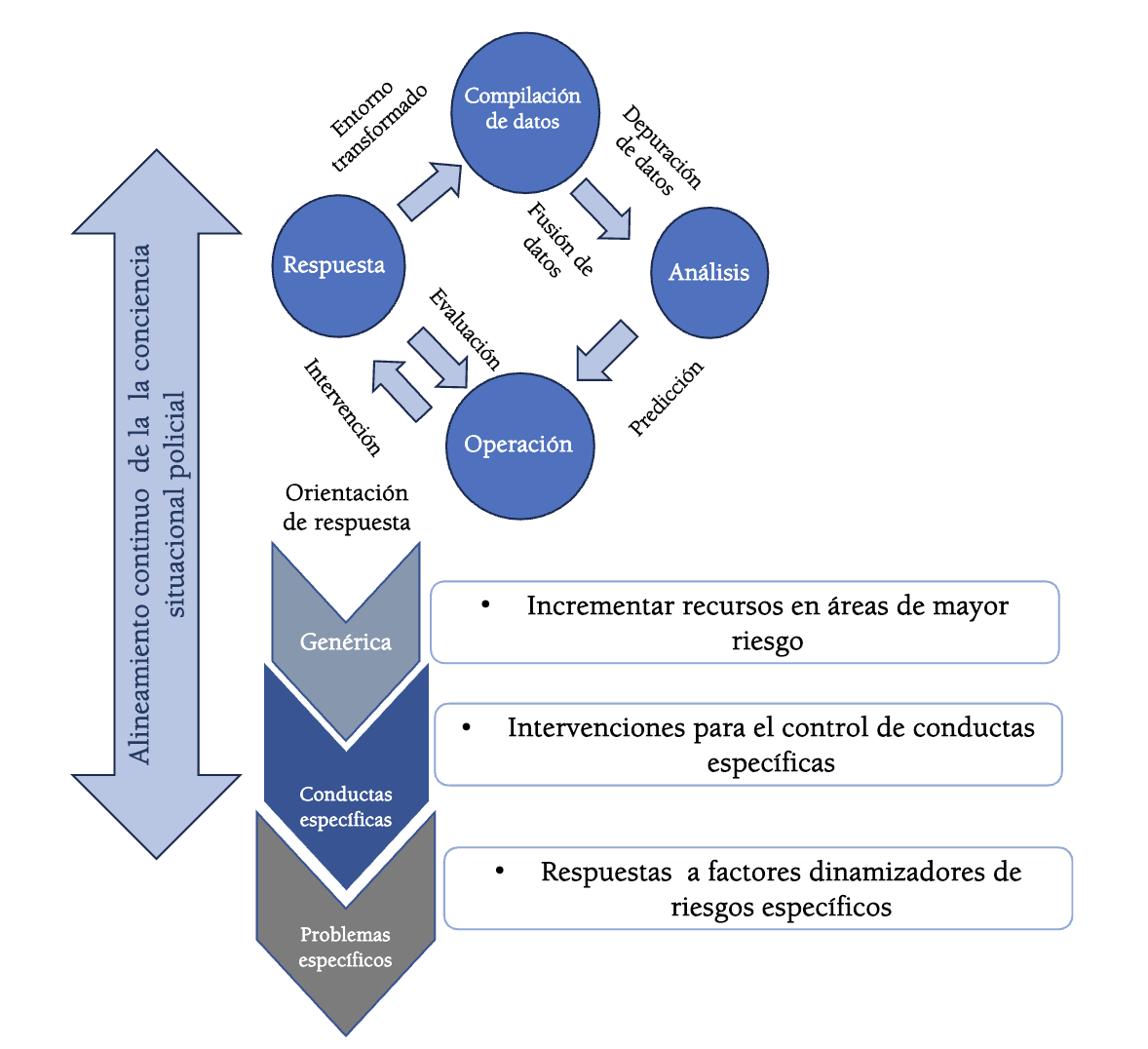
### Pérez Salazar señala que si bien estos sistemas ofrecen beneficios operativos y contribuyen a la prevención del delito, presentan riesgos en cuanto a sesgos algorítmicos y falta de interpretación contextual. Aun así, se plantea que el uso de estos modelos puede fortalecer la capacidad del Estado para desarrollar una estrategia policial más informada y adaptativa.

**B) Metodología**

El autor desarrolla una investigación de tipo documental, basada en una revisión crítica de literatura publicada entre los años 2013 y 2023, centrada en el uso de herramientas de aprendizaje automático aplicadas a la predicción del delito y a las transformaciones de la estrategia policial. A partir del análisis de artículos académicos, informes institucionales y estudios de caso, se examinan las plataformas predictivas más reconocidas, como PredPol en Estados Unidos y AISec en Colombia.

La metodología incluye el estudio detallado de variables sociales, geográficas y delictivas utilizadas por estos sistemas, y evalúa el impacto de los modelos predictivos en la prevención del crimen, particularmente en el contexto colombiano. El análisis también incorpora una dimensión crítica, orientada a visibilizar los riesgos éticos y políticos asociados a la automatización del control social, tales como los sesgos algorítmicos, la falta de transparencia y la dependencia excesiva de los datos históricos sin interpretación cualitativa.

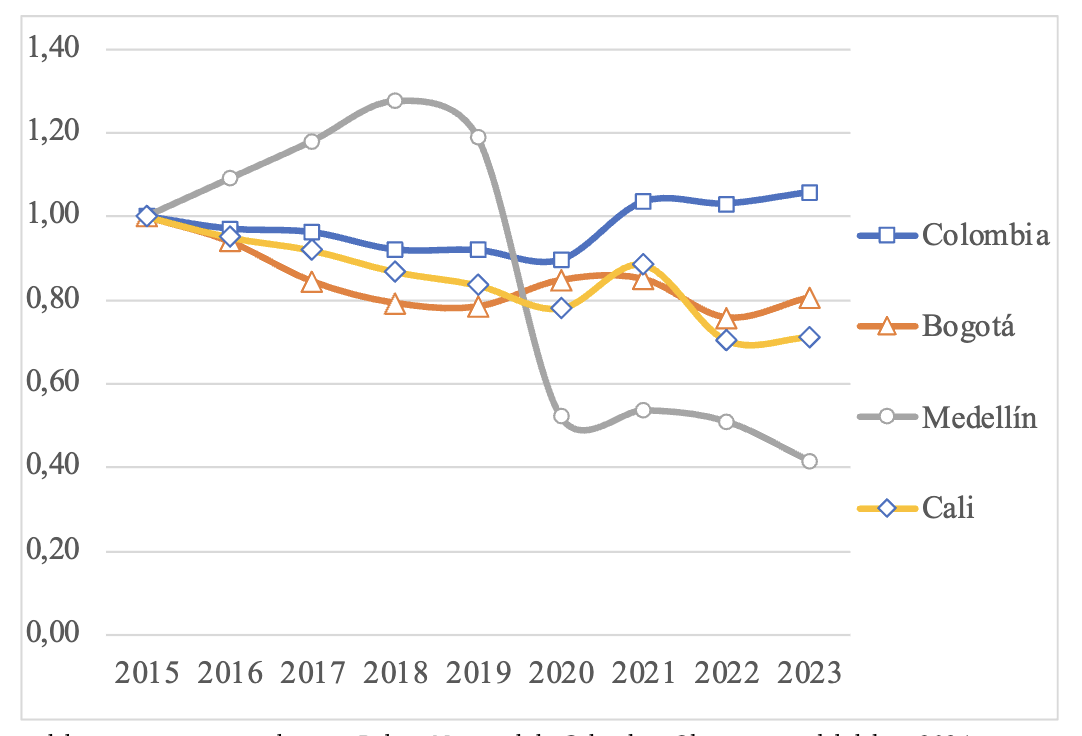
**Figura 1:** Ciclo de gestión del conocimiento en el enfoque de policía predictiva



**C) Etapas**

* **Etapa 1:** Sistematización de modelos predictivos existentes.
* **Etapa 2:** Evaluación de resultados en la reducción del delito.
* **Etapa 3:** Crítica ética a los riesgos del uso automatizado.
* **Etapa 4:** Propuestas de gobernanza de evidencias y modelos interpretables.

**Figura 2:** Colombia y tres ciudades principales. Tendencias del homicidio 2015-2023



**Utilidad del artículo para el proyecto de tesis**

Este estudio aporta bases conceptuales sólidas sobre cómo integrar inteligencia artificial en estrategias públicas de prevención y gestión. Al analizar la evolución de modelos predictivos en seguridad ciudadana, el autor demuestra que su efectividad depende no solo de la calidad del algoritmo, sino también del uso ético, responsable y transparente. Las reflexiones presentadas son aplicables al desarrollo de herramientas en otros sectores que también requieren planificación con recursos limitados, especialmente en contextos donde es necesario anticipar la demanda mediante análisis multivariable.

La importancia de evaluar la calidad de los datos y su uso ofrece un respaldo teórico crucial para comprender cómo los sistemas de inteligencia artificial y aprendizaje automático pueden ser utilizados en la formulación de políticas públicas orientadas a la prevención y la toma de decisiones estratégicas. La insistencia en la optimización basada en evidencia y la evaluación crítica de la criminología predictiva sirven como guía para la implementación de soluciones tecnológicas complejas en sistemas que exigen una estrategia policial más ágil, precisa y orientada al bien común.

# REFERENCIA BIBLIOGRÁFICA

Alsubayhin, A., Ramzan, M., & Alzahrani, B. (2023). Crime Prediction Using Machine Learning: A Comparative Analysis. *Journal of Computer Science*, *19*(9), 1170–1179. https://doi.org/10.3844/JCSSP.2023.1170.1179

Doreswamy, N. (n.d.). *A Comparison Between Human and Generative AI Decision-Making Attributes in Complex Health Services Short title: Comparing Human and Generative AI Decision-Making Attributes*.

Escobedo, M., Tapia, C., Gutierrez, J., Ayma, V., Member, S., De Lima, U., Lima, P., & Pacifico, D. (2024). Comparing Regression Models to Predict Property Crime in High-Risk Lima Districts. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 15, Issue 3). https://doi.org/10.48550/arXiv.2505.08360

Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., & Brantingham, P. J. (2015). Randomized Controlled Field Trials of Predictive Policing. *Journal of the American Statistical Association*, *110*(512), 1399–1411. https://doi.org/10.1080/01621459.2015.1077710

Saraiva, M., Matijošaitienė, I., Mishra, S., & Amante, A. (2022). Crime Prediction and Monitoring in Porto, Portugal, Using Machine Learning, Spatial and Text Analytics. *ISPRS International Journal of Geo-Information*, *11*(7). https://doi.org/10.3390/ijgi11070400

Sridharan, S., Srish, N., Vigneswaran, S., & Santhi, P. (2024). Crime Prediction using Machine Learning. *EAI Endorsed Transactions on Internet of Things*, *10*. https://doi.org/10.4108/eetiot.5123

Utsha, R. B., Alif, M. N., Rayhan, Y., Hashem, T., & Ali, M. E. (2024). *Deep Learning Based Crime Prediction Models: Experiments and Analysis*. http://arxiv.org/abs/2407.19324

Williams, M. L., Burnap, P., & Sloan, L. (2017). Crime sensing with big data: The affordances and limitations of using open-source communications to estimate crime patterns. *British Journal of Criminology*, *57*(2), 320–340. https://doi.org/10.1093/bjc/azw031

Wu, J., & Frias-Martinez, V. (2024). *Improving the Fairness of Deep-Learning, Short-term Crime Prediction with Under-reporting-aware Models*. http://arxiv.org/abs/2406.04382

Zhao, L. (2024). A Study on the Application of Data Mining-based Crime Prediction Models in Criminal Justice. *Journal of Combinatorial Mathematics and Combinatorial Computing*, *123*, 559–575. https://doi.org/10.61091/jcmcc123-40