



**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 consolidándose 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

<b>DEDICATORIA</b>	<b>2</b>
<b>AGRADECIMIENTO</b>	<b>3</b>
<b>EPÍGRAFE</b>	<b>4</b>
<b>RESUMEN</b>	<b>5</b>
<b>ABSTRACT</b>	<b>6</b>
<b>ÍNDICE</b>	<b>7</b>
<b>INTRODUCCIÓN</b>	<b>1</b>
<b>CAPÍTULO I: VISIÓN DEL PROYECTO</b>	<b>3</b>
<b>1.1 Antecedentes del Problema</b>	<b>3</b>
1.1.1 El Negocio	3
1.1.2 Procesos del Negocio	5
1.1.3 Organigrama	6
<b>1.2 Formulación del Problema</b>	<b>6</b>
1.2.1 Realidad Problemática	6
1.2.2 Descripción del Problema	8
<b>1.3 Objetivos del Proyecto</b>	<b>9</b>
1.3.1 Marco Lógico	9
1.3.2 Objetivo General	10
1.3.3 Objetivos Específicos	11
<b>1.4 Justificación del Proyecto</b>	<b>11</b>
1.4.1 Justificación Académica	11
1.4.2 Beneficios Tangibles	12
1.4.3 Beneficios Intangibles	12
<b>1.5 Alcance del Proyecto</b>	<b>12</b>
<b>CAPÍTULO II: MARCO TEÓRICO.....</b>	<b>13</b>
<b>CAPÍTULO III: ESTADO DEL ARTE.....</b>	<b>16</b>
<b>3.1. Artículos.....</b>	<b>16</b>
<b>CAPÍTULO IV: MODELADO DE NEGOCIO .....</b>	<b>20</b>
<b>4.1 Reglas del Negocio .....</b>	<b>26</b>
<b>CAPÍTULO V : REQUERIMIENTOS DEL PROYECTO</b>	
<b>5.1 Relación de Requerimientos Funcionales del Sistema .....</b>	<b>28</b>
<b>REFERENCIA BIBLIOGRÁFICA .....</b>	<b>30</b>

## 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 & Rodríguez, 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:

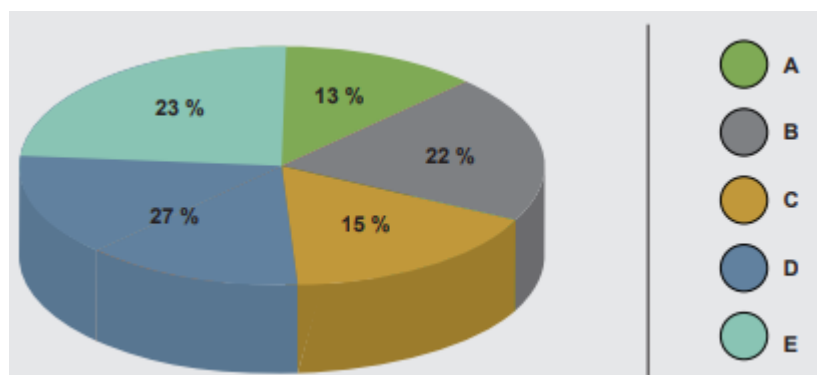
- En el capítulo I, muestra la visión del proyecto, donde se describen y detallan los antecedentes que se tomaron como referencia, la problemática a definir, así como el objetivo general y específicos. Seguidamente se menciona la justificación del tema y el alcance a desarrollarse.

## CAPÍTULO I: VISIÓN DEL PROYECTO

### 1.1 Antecedentes del Problema

#### 1.1.1 El Negocio

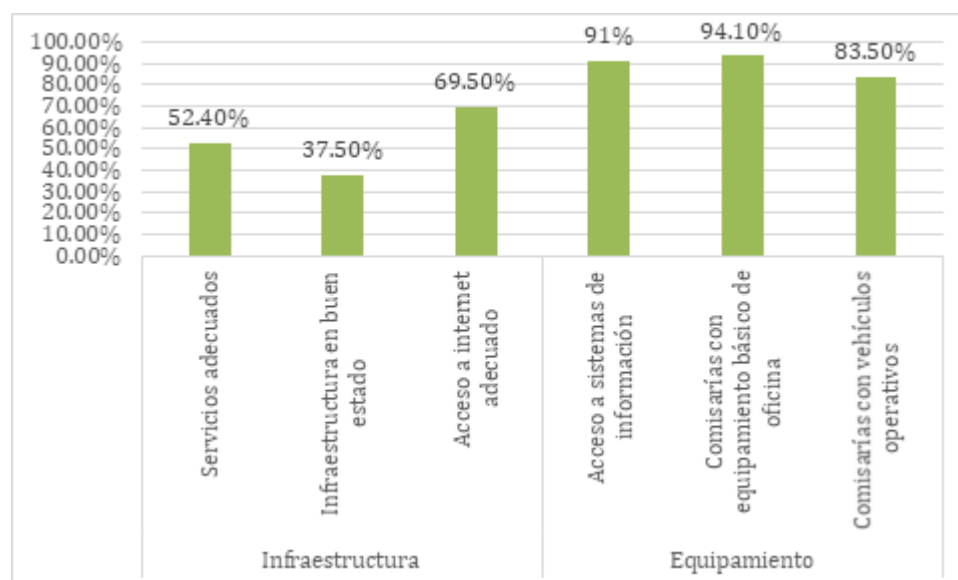
Las comisarías de la Policía Nacional del Perú son la célula base de esta institución y suman un total de 1,318 a nivel nacional. Según su tipo pueden clasificarse en A, B, C, D, E de acuerdo a su capacidad de efectivos policiales, área construida y cobertura de habitantes. La A representa la de mayor 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.** 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 2.** Estado de la infraestructura y equipamiento de las comisariías PNP a nivel nacional

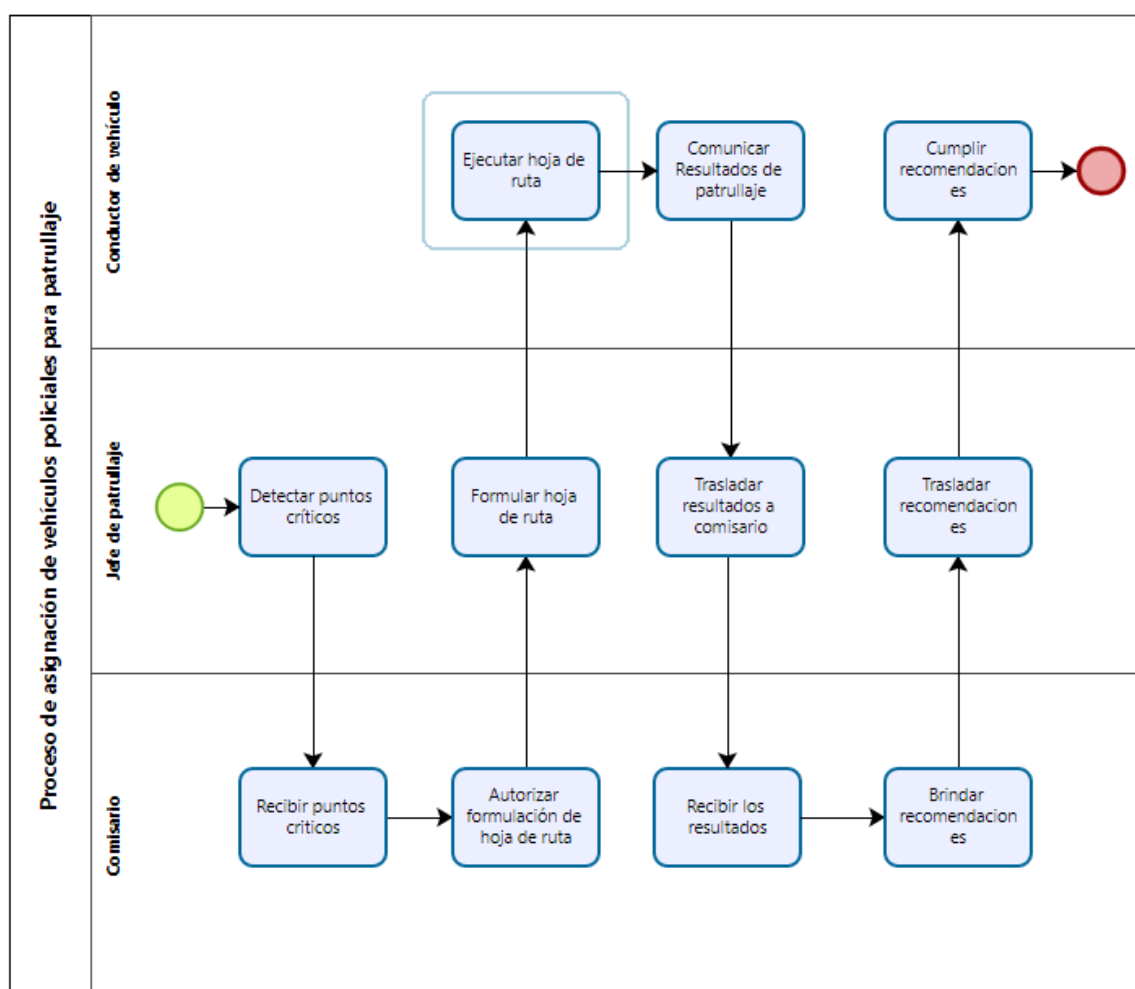
**Fuente:** (MININTER, 2025)

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 comisariías 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.

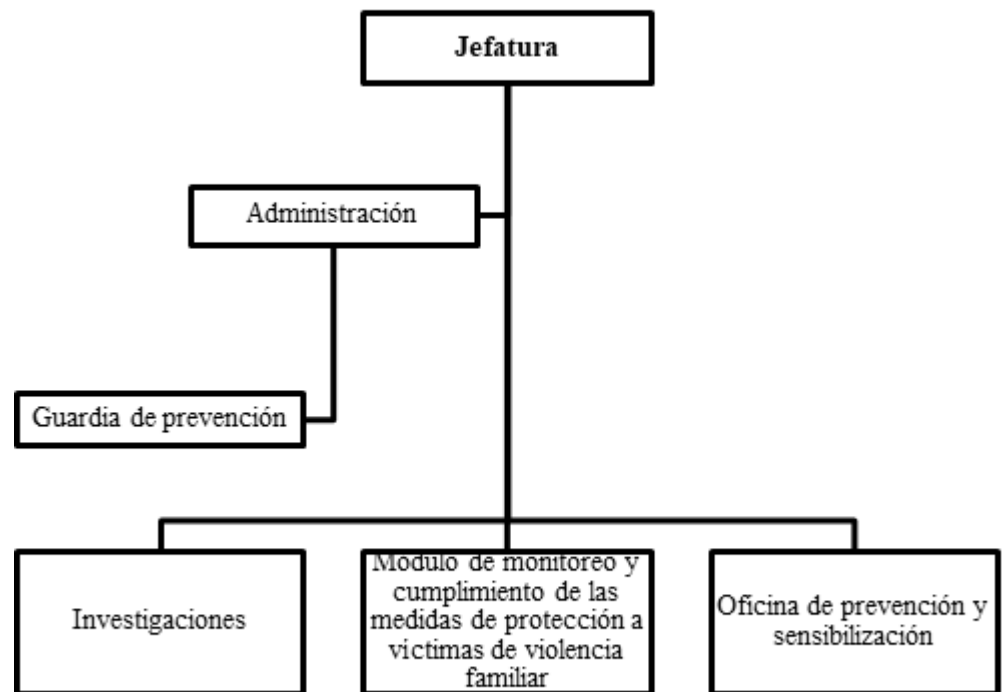
### 1.1.2 Procesos del Negocio



**Figura 3.** Diagrama de AS-IS de asignación de recurso policiales

### 1.1.3 Organigrama

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

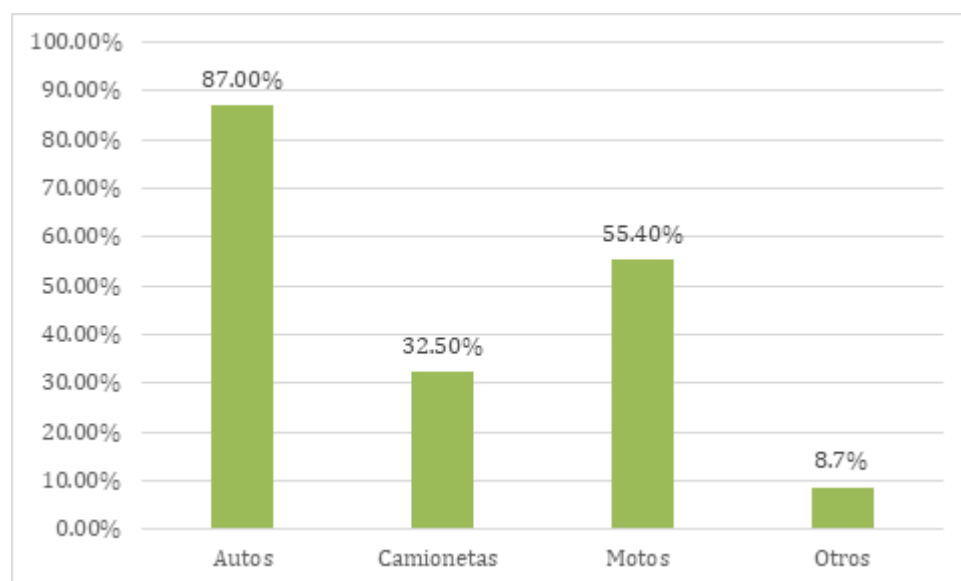


*Figura 4. Organigrama de una comisaría PNP*

## 1.2 Formulación del Problema

### 1.2.1 Realidad Problemática

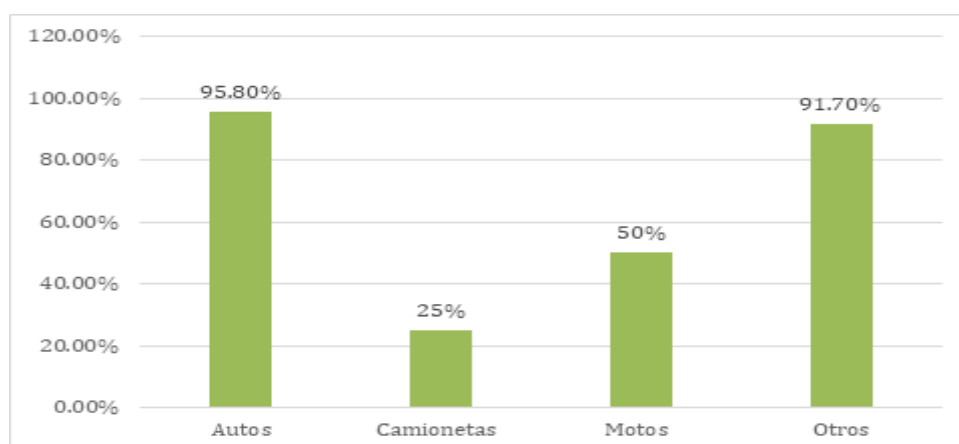
Los recursos en la institución policial resultan relevantes para mejorar la seguridad ciudadana, pero cuando se habla de vehículos policiales o también conocidos como patrulleros, se tiene la lamentable cifra de unidades que no pueden ser utilizadas debido a que se encuentran inoperativas. En ese sentido 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 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 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.

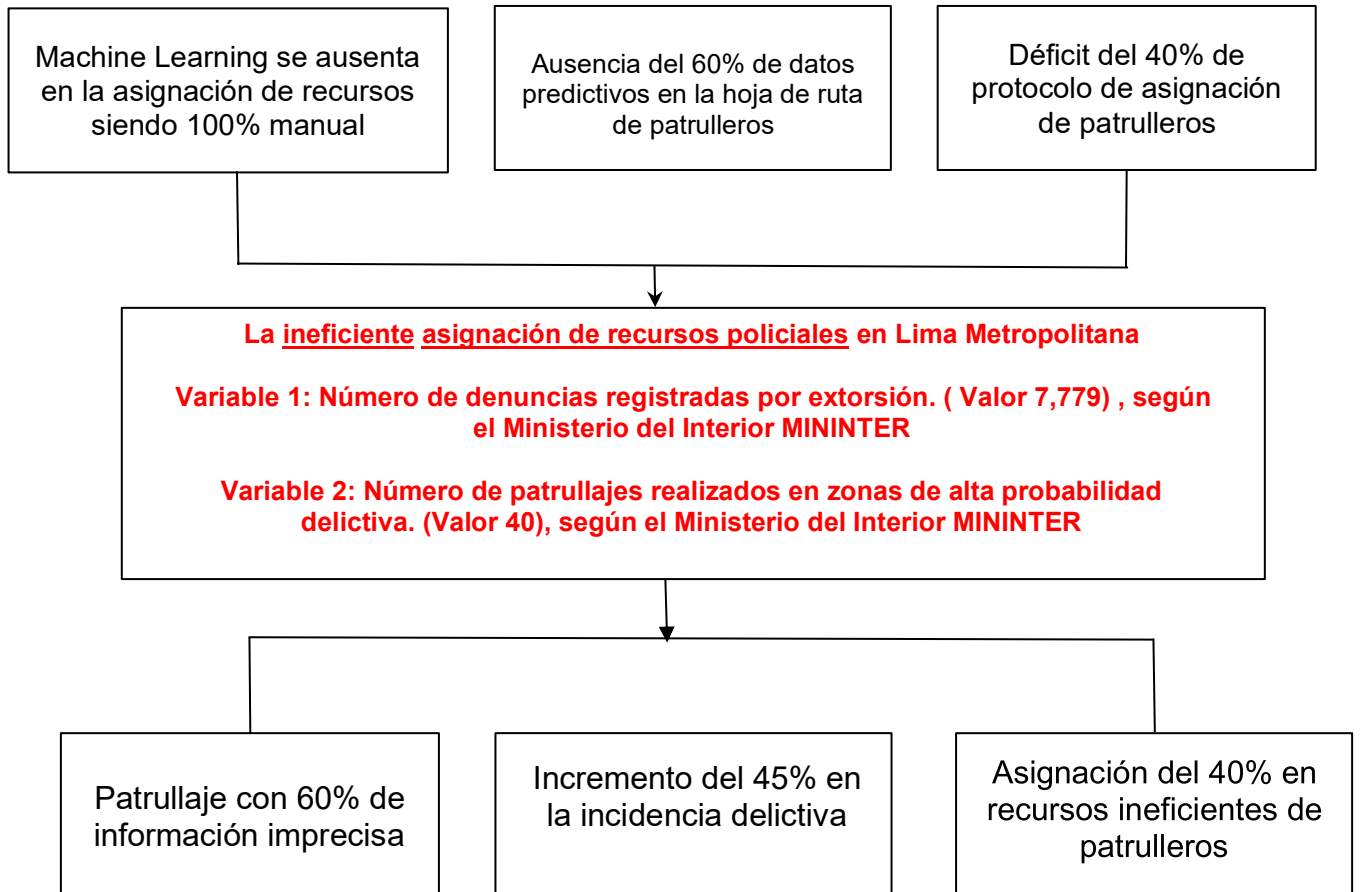
### 1.2.2.2 Problema Secundario

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

### 1.3 Objetivos del Proyecto

#### 1.3.1 Marco Lógico

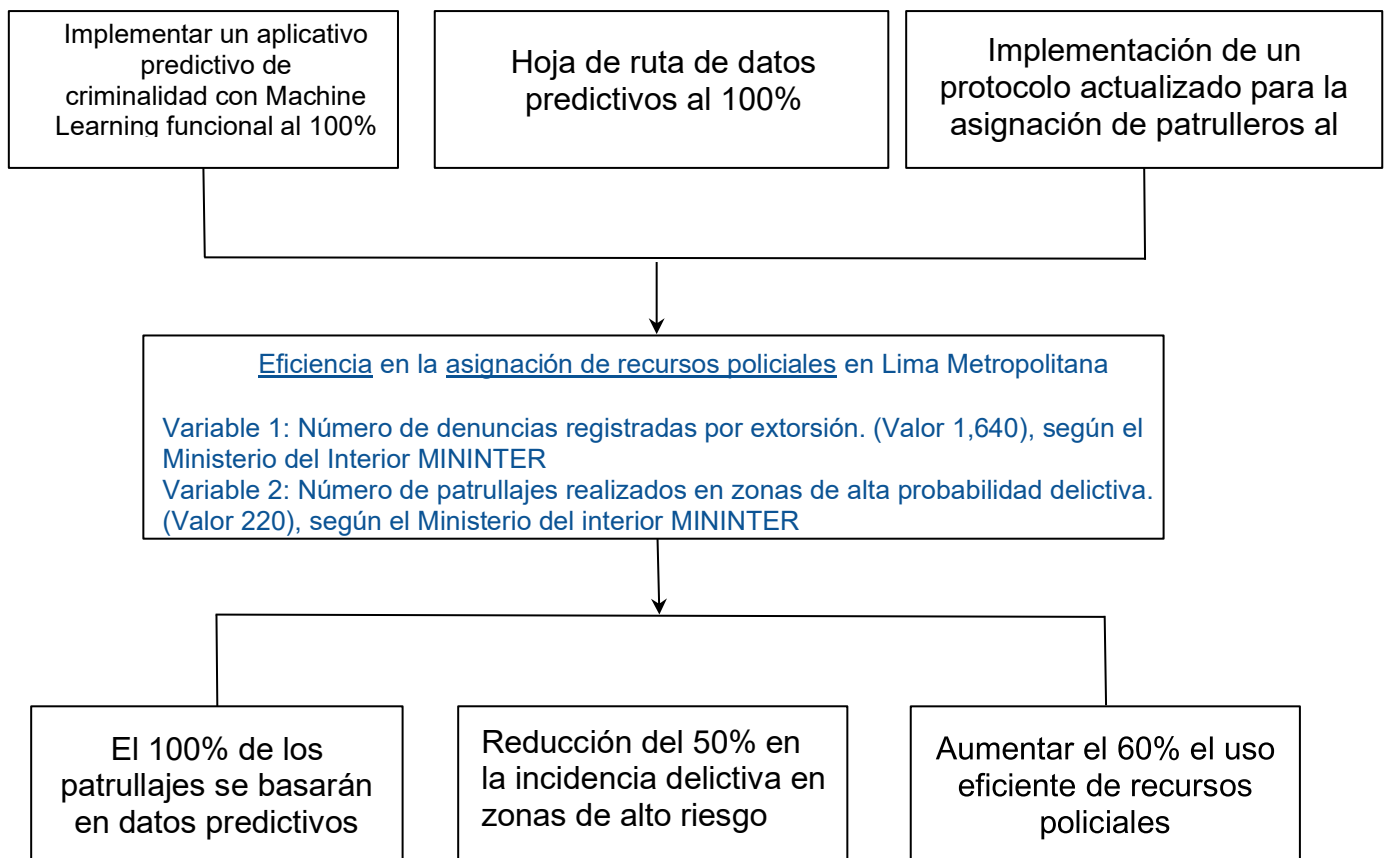
##### 1.3.1.1 Árbol del Problemas:



**Figura 7.** Árbol de problemas determinados



### 1.3.1.2 Árbol de Objetivos:



**Figura 8.** Árbol de objetivos determinados

### 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.

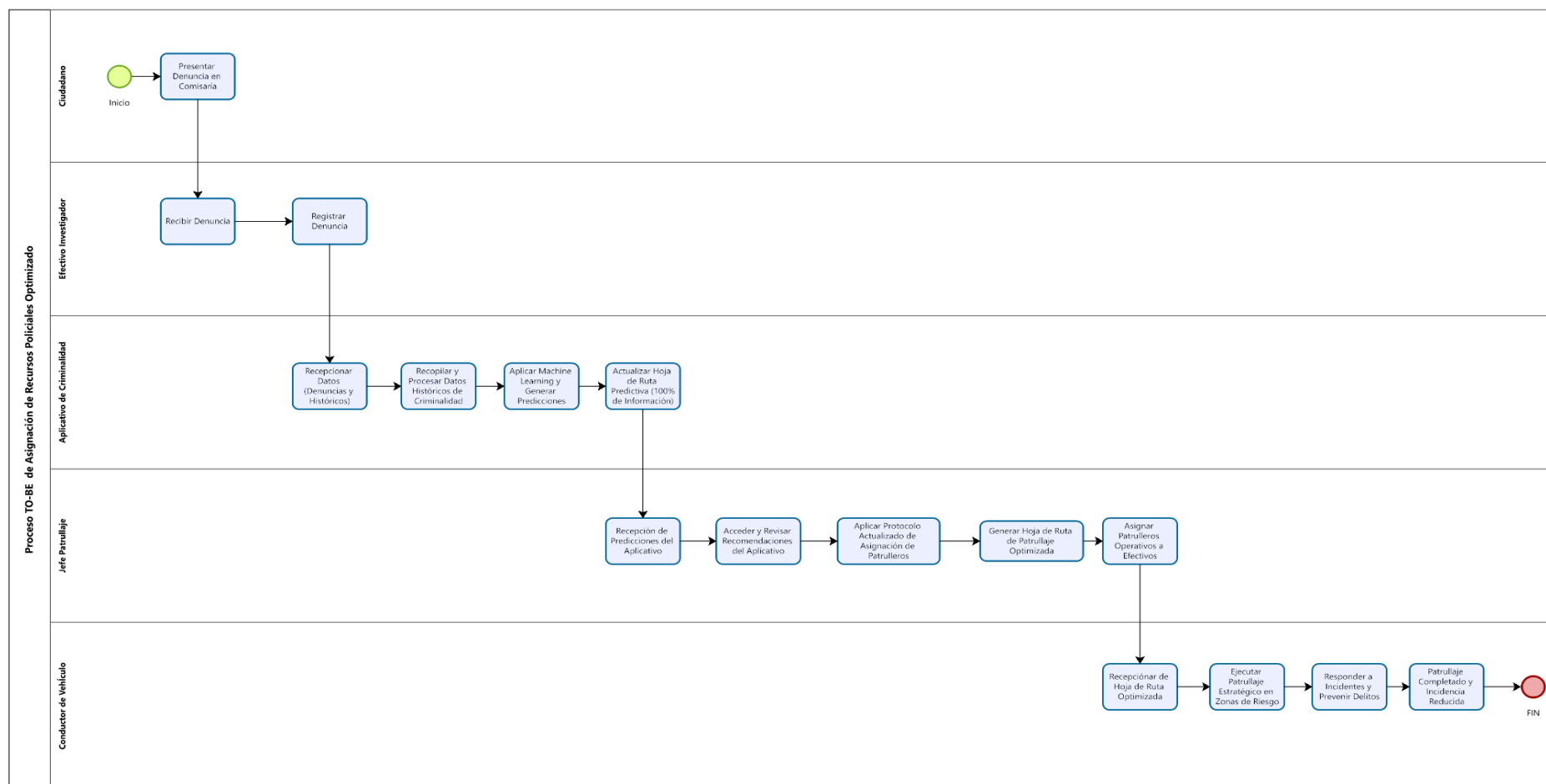
Variable 1: Número de denuncias registradas por extorsión. ( Valor 1,640), según el Ministerio del Interior MININTER

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.

### 1.3.3 Modelado del Proceso To-Be



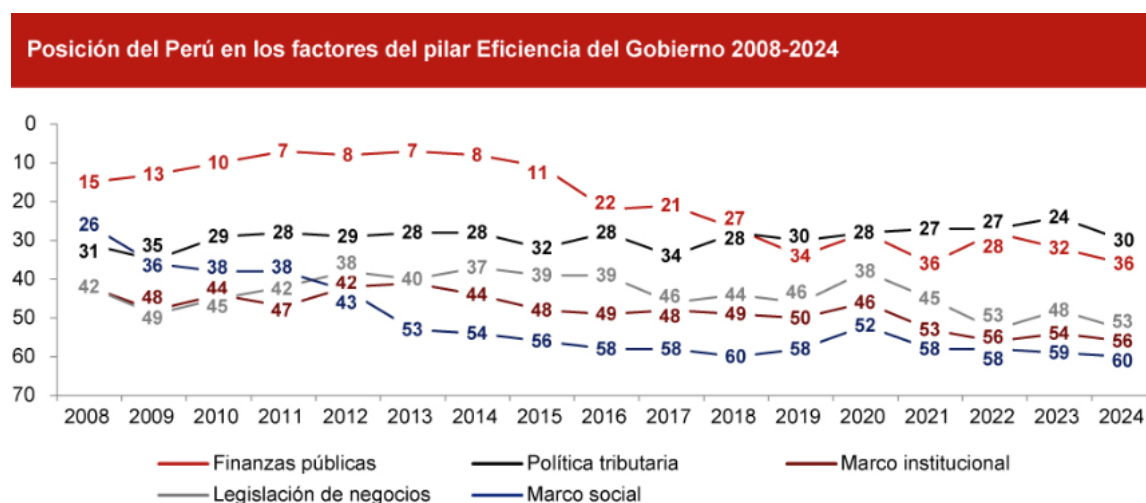
Powered by  
bpmn.io  
Modeler

**Figura 9 – Diagrama de proceso (To-Be) de la asignación de recurso policiales**

## 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.



**Figura 10.** 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

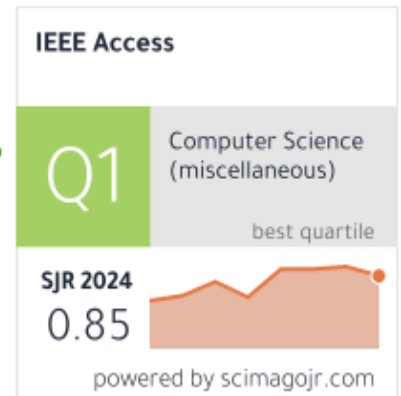
#### 3.1.1 Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques

**Análisis empírico para la predicción y el pronóstico de delitos mediante técnicas de aprendizaje automático y aprendizaje profundo**

(Safat et al., 2021)

(DOI: [10.1109/ACCESS.2021.3078117](https://doi.org/10.1109/ACCESS.2021.3078117))

**Citado 100 veces en Scopus, Cuartil Q1**

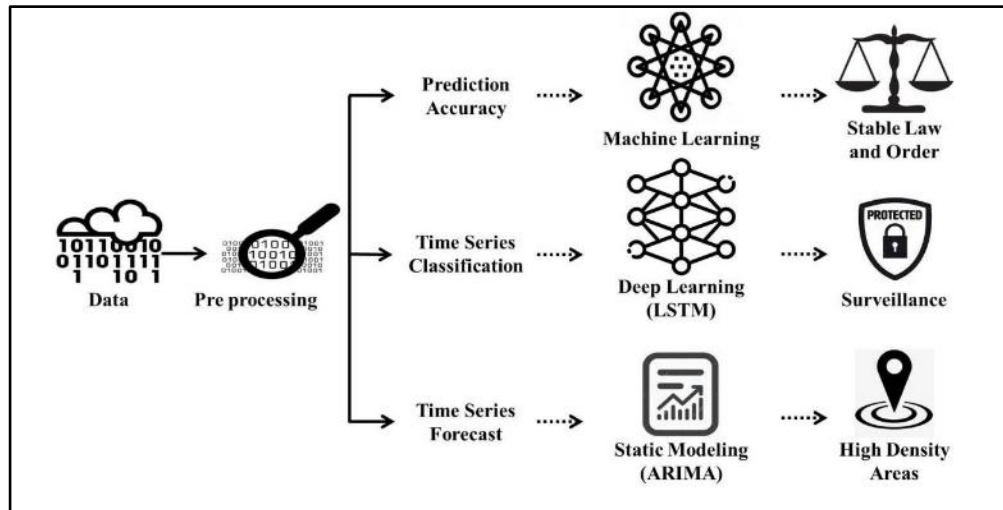


Este artículo aborda la predicción y el pronóstico de delitos utilizando técnicas de aprendizaje automático y aprendizaje profundo. El objetivo principal es mejorar la precisión de la predicción del crimen y la identificación de las zonas de alta criminalidad, lo que puede ayudar a las autoridades policiales a asignar recursos de manera más eficiente. La problemática que trata se centra en la dificultad de procesar grandes cantidades de datos y cómo esto limita la capacidad para predecir los crímenes de forma precisa. A través de un análisis de series temporales y la implementación de algoritmos como el LSTM (Long Short-Term Memory) y ARIMA (AutoRegressive Integrated Moving Average), el artículo mejora la precisión en comparación con estudios previos.

La metodología empleada incluye el uso de algoritmos de aprendizaje automático, como la regresión logística, máquinas de soporte vectorial (SVM), Naïve Bayes, K-nearest neighbors (KNN), árboles de decisión, redes neuronales multicapa (MLP), Random Forest y eXtreme Gradient Boosting (XGBoost). El análisis de series temporales con LSTM también se utiliza para prever las tendencias futuras del crimen. Además, se realiza un análisis exploratorio de los datos, lo que permite identificar patrones y zonas de alta criminalidad.

Las conclusiones indican que los resultados obtenidos con los modelos propuestos (LSTM y ARIMA) permiten predecir con mayor precisión las zonas de mayor criminalidad y las tendencias futuras. Los modelos también muestran que el índice de

criminalidad en Chicago está aumentando moderadamente, mientras que en Los Ángeles se prevé una disminución.



*Figura 11: Metodología propuesta y marco de estudio.*

## B) Metodología

El enfoque metodológico de este estudio es de tipo **investigación aplicada**, con énfasis en el uso de técnicas de **aprendizaje automático** y **aprendizaje profundo**. La investigación está estructurada en dos partes principales: la predicción de crímenes y la predicción de tendencias futuras de crímenes (pronóstico).

Los pasos principales de la metodología incluyen:

1. **Recopilación de datos:** Se utilizaron conjuntos de datos públicos sobre delitos en las ciudades de Chicago y Los Ángeles. Estos datos incluyen información sobre el tipo de delito, la ubicación, la fecha y otros factores contextuales.
2. **Preprocesamiento de datos:** Los datos fueron limpiados, con la eliminación de registros incompletos o erróneos, y luego se dividieron en conjuntos de entrenamiento y prueba (70% para entrenamiento y 30% para prueba).
3. **Análisis de algoritmos:** Se implementaron ocho algoritmos de aprendizaje automático para predecir los delitos en ambos conjuntos de datos. Los algoritmos utilizados fueron regresión logística, SVM, aive Bayes, KNN,

árbol de decisión, MLP, Random Forest y XGBoost.

4. **Análisis de series temporales con LSTM:** El modelo LSTM fue entrenado para predecir patrones de crímenes a lo largo del tiempo, capturando dependencias a largo plazo en los datos de crímenes pasados.
5. **Pronóstico con ARIMA:** Se utilizó el modelo ARIMA para prever las tendencias futuras del crimen y determinar las áreas de mayor riesgo en los años venideros.

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

Gracias a los aportes de este artículo, es de gran utilidad para nuestra tesis propuesta, ya que desarrollaremos un módulo de predicción temporal de criminalidad basado en análisis de series históricas de datos delictivos en Lima Metropolitana utilizando modelos avanzados de Deep Learning como Long-Short Term Memory (LSTM) y AutoRegressive Integrated Moving Average (ARIMA), que han demostrado alta precisión para captar patrones estacionales y tendencias evolutivas del crimen urbano. Otra aportación de gran utilidad ya que diseñaremos un módulo funcional para generar reportes predictivos visuales que identifiquen claramente los puntos calientes de criminalidad, detallando el comportamiento delictivo esperado por día, semana, mes o trimestre y por zonas geográficas específicas, entregando información precisa a las autoridades policiales para optimizar la asignación anticipada de recursos en los distritos más vulnerables de Lima Metropolitana. En este sentido el artículo, proporciona información valiosa sobre el aprendizaje automático y el aprendizaje profundo son fundamentales para mejorar la precisión de la predicción del crimen, ya que permiten procesar grandes volúmenes de datos y extraer patrones complejos. LSTM (Long Short-Term Memory) es un modelo de redes neuronales recurrentes que es particularmente útil para el análisis de series temporales, lo que permitirá capturar las dependencias a largo plazo en los datos históricos de criminalidad y prever los patrones futuros. Por otro lado, ARIMA (AutoRegressive Integrated Moving Average) es un modelo estadístico eficaz para realizar pronósticos de series temporales, que en nuestro caso se utilizará para predecir la evolución de la criminalidad en Lima Metropolitana a partir de los datos históricos disponibles. Además, el análisis exploratorio de los datos, como se menciona en el artículo, es crucial para identificar las tendencias y características clave de los crímenes en las distintas áreas de la ciudad, lo que facilitará la creación de mapas de calor y la identificación de puntos críticos para la asignación de recursos. Estos enfoques



contribuirán de manera significativa al éxito de nuestra tesis y al desarrollo de una herramienta eficiente para la prevención del crimen.

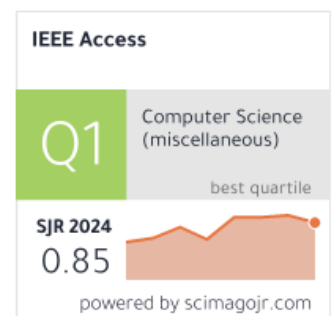
De acuerdo a la **figura 11**, que ilustra el marco metodológico y los procesos involucrados en la predicción de criminalidad, se indica **cómo se integran los modelos de predicción y clasificación de criminalidad a partir del preprocesamiento de los datos hasta su aplicación en la predicción de eventos delictivos y la asignación de recursos**. En este sentido, se tomará el **enfoque de aprendizaje automático con modelos como LSTM y ARIMA, los cuales son adecuados para desarrollar un sistema que procese las series temporales de criminalidad, permitiendo una predicción automática del comportamiento delictivo y la optimización de la asignación de recursos en Lima Metropolitana**.

### 3.1.2 Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions

#### **Predicción de delitos mediante aprendizaje automático y aprendizaje profundo: una revisión sistemática y futuras direcciones**

(Mandalapu et al., 2023)

(DOI: [10.1109/ACCESS.2023.3286344](https://doi.org/10.1109/ACCESS.2023.3286344))



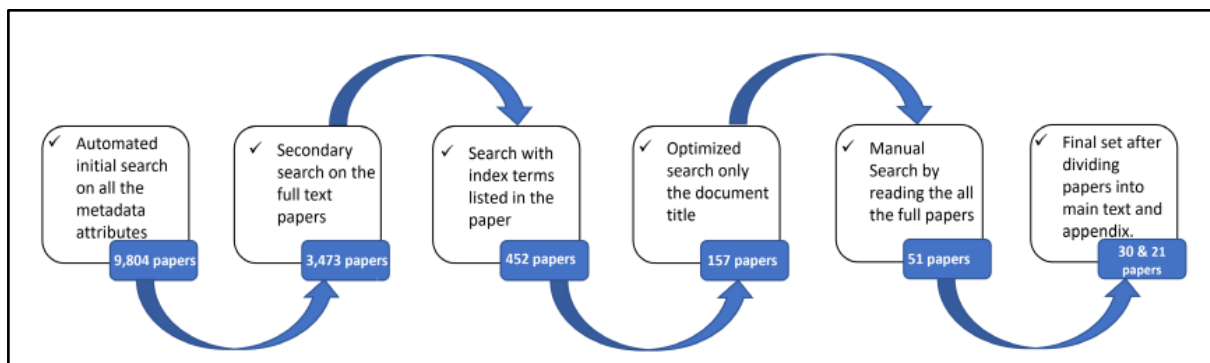
**Citado 59 veces en Scopus , Cuartil Q1**

El presente artículo realiza una revisión sistemática sobre las técnicas de aprendizaje automático (Machine Learning, ML) y aprendizaje profundo (Deep Learning, DL) aplicadas en la **predicción de delitos**. El objetivo es proporcionar una visión general de los enfoques y algoritmos más relevantes utilizados para predecir la ocurrencia de crímenes, identificar patrones y mejorar las estrategias de policía predictiva. A través del análisis de más de 150 artículos, el estudio destaca los principales algoritmos, conjuntos de datos y aplicaciones, además de identificar brechas y áreas de mejora en este campo.

La problemática que aborda el artículo radica en la necesidad de predecir y prevenir los crímenes de manera más precisa y efectiva. A pesar de los avances en el uso de técnicas de ML y DL, los modelos actuales enfrentan limitaciones debido a la calidad y la disponibilidad de los datos, así como a la complejidad de interpretar los modelos utilizados. Esta situación es particularmente crítica para las fuerzas de seguridad, que necesitan tomar decisiones rápidas y basadas en datos confiables.

Los resultados clave del artículo revelan que las técnicas de ML y DL pueden mejorar significativamente la precisión de las predicciones sobre patrones criminales. Sin embargo, el artículo también señala que todavía existen desafíos importantes, como la interpretabilidad de los modelos, la calidad de los datos y las implicaciones éticas del uso de estas tecnologías.

Las conclusiones principales incluyen la recomendación de mejorar los modelos actuales, integrando mejores datos, abordando las preocupaciones éticas y desarrollando métodos que permitan una mayor comprensión y transparencia en las predicciones. A largo plazo, esto podría contribuir a una mayor eficacia en la asignación de recursos y la prevención del crimen.



*Figura 12: Metodología de selección de artículos de investigación.*

## B) Metodología

El enfoque metodológico adoptado por el artículo es una investigación cualitativa de tipo revisión sistemática. La investigación se centra en examinar y analizar una serie de estudios previos sobre la predicción de delitos usando técnicas de ML y DL. Se utilizaron varias bases de datos académicas (como IEEE Xplore y ScienceDirect) para seleccionar artículos relevantes, utilizando criterios de búsqueda automatizada y manual para identificar los estudios más significativos.

### Pasos principales de la metodología:

- 1. Recopilación de datos:** Se recolectaron más de 450 artículos relacionados con la predicción de crímenes utilizando ML y DL. Estos artículos fueron filtrados mediante un proceso de selección basado en su relevancia, novedad y aplicabilidad.
- 2. Análisis cualitativo:** Los estudios seleccionados fueron analizados para identificar los algoritmos más utilizados, los conjuntos de datos aplicados y los métodos de

evaluación empleados. El artículo clasifica los enfoques en categorías como clasificación, regresión y detección de anomalías.

3. **Identificación de brechas y desafíos:** A través del análisis, se identificaron áreas de mejora, tales como la falta de interpretabilidad de los modelos, los problemas éticos relacionados con la predicción del crimen y la disponibilidad de datos confiables.

### **Fases del estudio**

- **Búsqueda automatizada:** Se utilizaron términos clave relacionados con "predicción de crímenes", "aprendizaje automático", "aprendizaje profundo", entre otros, para buscar artículos en las bases de datos académicas.
- **Selección de artículos:** Después de una búsqueda inicial, los artículos fueron seleccionados manualmente según su relevancia y calidad. Los estudios fueron clasificados según los enfoques de ML y DL aplicados y las métricas utilizadas para evaluar su efectividad.
- **Revisión de los resultados:** Tras la selección, se compararon los resultados de los modelos utilizados y se evaluaron en términos de precisión, escalabilidad y aplicabilidad en el mundo real.

### **Evaluación de los resultados**

Los resultados obtenidos muestran que los enfoques de aprendizaje profundo son especialmente útiles para el análisis de datos complejos, como imágenes de cámaras de seguridad o patrones de comportamiento en redes sociales, mientras que los modelos de aprendizaje automático tradicionales, como los árboles de decisión y las máquinas de soporte vectorial, son más fáciles de interpretar y requieren menos datos. Sin embargo, ambos enfoques enfrentan desafíos relacionados con la calidad de los datos y la falta de modelos explicativos. Los modelos híbridos, que combinan ML y DL, parecen ser más efectivos para abordar estas limitaciones.

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

Gracias a los aportes de este artículo, es de gran utilidad para nuestra tesis propuesta, ya que desarrollaremos un módulo de predicción de criminalidad que utilizará los modelos de Machine Learning y Deep Learning revisados en el estudio para analizar datos históricos de delitos y registros policiales de Lima Metropolitana. Específicamente, se aplicarán algoritmos como los árboles de decisión, SVM o redes neuronales para identificar patrones y tendencias de criminalidad en distintas zonas y horarios. Otra aportación de gran utilidad es que, diseñaremos un módulo de visualización de datos que generará reportes interactivos y mapas de calor que mostrarán las zonas de mayor riesgo y los tipos de delitos más frecuentes, permitiendo a la Policía Nacional del Perú en Lima Metropolitana optimizar la asignación de recursos policiales en áreas específicas y en momentos clave, mejorando así la capacidad de respuesta y la prevención del delito.

En ese sentido, el artículo proporciona información valiosa sobre los enfoques de aprendizaje automático (Machine Learning, ML) son fundamentales para analizar grandes volúmenes de datos y extraer patrones significativos que ayuden en la predicción de la criminalidad. Además, se destacan las capacidades de aprendizaje profundo (Deep Learning, DL) para trabajar con datos complejos, como imágenes o patrones temporales y espaciales, que pueden ser útiles en la detección de puntos críticos de criminalidad. Sin embargo, uno de los desafíos señalados en el artículo es la interpretabilidad de los modelos, ya que muchos de los modelos de ML y DL pueden ser complejos y difíciles de entender, lo que limita su aplicación en contextos prácticos. Asimismo, la calidad de los datos es crucial para obtener predicciones precisas; los modelos dependen de datos completos y actualizados para generar resultados confiables. Por otro lado, el artículo también subraya las implicaciones éticas del uso de estas tecnologías, especialmente en lo que respecta a la privacidad y el sesgo en los datos, que deben ser considerados cuidadosamente para evitar discriminación y proteger los derechos de los individuos.

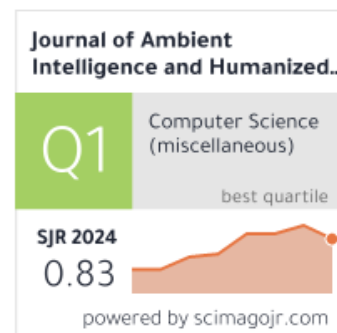
### 3.1.3 Machine learning in crime prediction

#### Aprendizaje automático en la predicción de delitos

(Jenga et al., 2023)

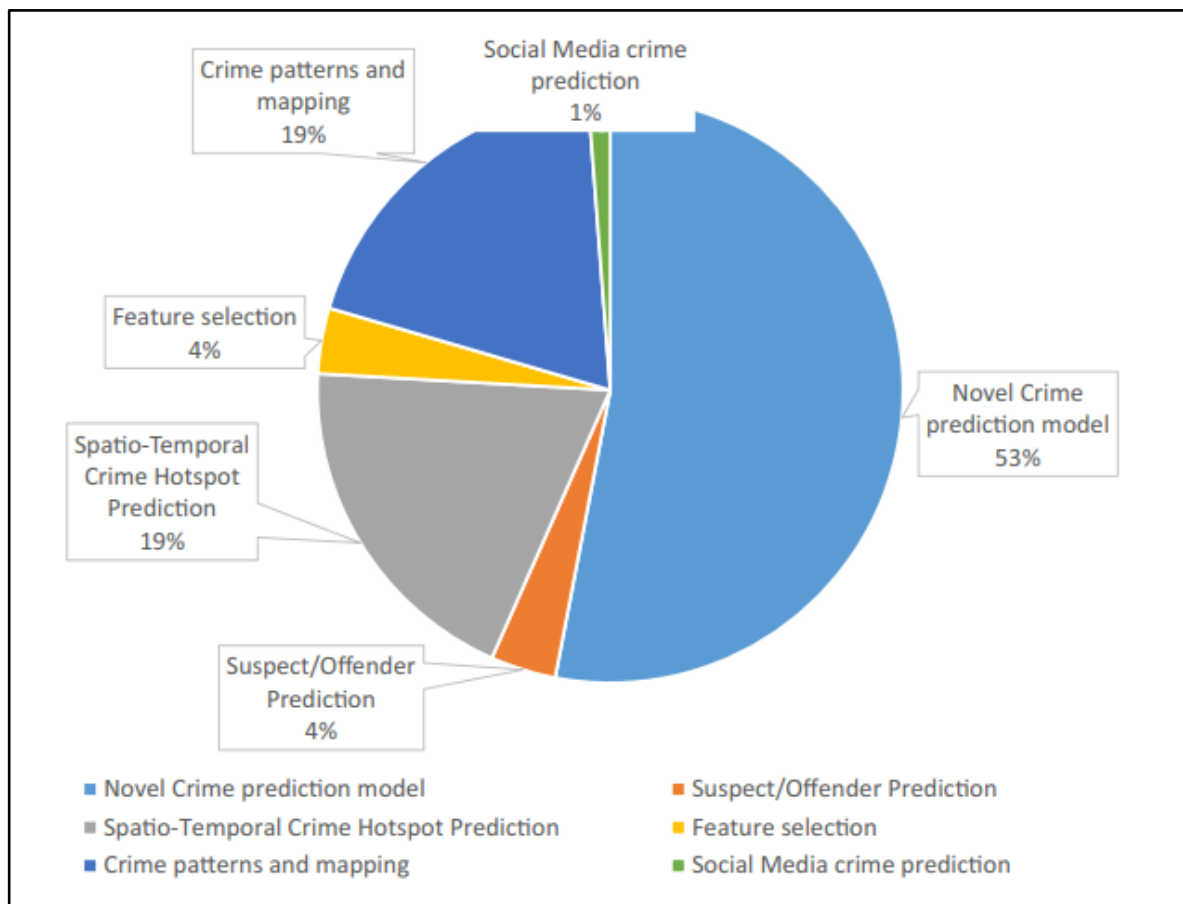
(DOI: [10.1007/s12652-023-04530-y](https://doi.org/10.1007/s12652-023-04530-y))

Citado 55 veces en Scopus, Cuartil Q1



El artículo “Machine Learning in Crime Prediction” aborda el uso de técnicas de minería de datos y aprendizaje automático para predecir y analizar el crimen. La predicción de crímenes antes de que ocurran puede salvar vidas y reducir pérdidas materiales, lo que hace esencial el desarrollo de modelos predictivos eficaces. El artículo tiene como objetivo evaluar las técnicas de predicción de crímenes más avanzadas de la última década, identificar los principales desafíos en el campo y proporcionar una visión sobre las futuras áreas de investigación. A pesar de que la mayoría de los trabajos se centran en predecir crímenes, los métodos utilizados y los conjuntos de datos varían ampliamente.

Una de las principales problemáticas que se aborda es la subutilización de grandes volúmenes de datos disponibles por parte de las agencias de policía. Aunque estos datos podrían ser extremadamente útiles para predecir y prevenir crímenes, la falta de herramientas computacionales efectivas para su análisis sigue siendo un obstáculo. El artículo se enfoca en una revisión sistemática de 68 estudios clave que emplean aprendizaje automático para la predicción del crimen. A través de esta revisión, se identifican las principales metodologías utilizadas, como los enfoques supervisados y no supervisados, y se analizan sus ventajas y limitaciones. En última instancia, el artículo propone que el uso adecuado de modelos predictivos, como redes neuronales y algoritmos de optimización, podría mejorar sustancialmente la asignación de recursos policiales y ayudar a prevenir futuros delitos.



**Figura 13: Distribución de objetivos**

## **B) Metodología:**

La metodología utilizada en este artículo es una Revisión Sistemática de la Literatura (SLR), una técnica ampliamente empleada en la investigación para sintetizar y analizar los estudios existentes sobre un tema específico. El enfoque SLR permite identificar patrones, evaluar los métodos utilizados en la predicción del crimen y resaltar las brechas de conocimiento en la literatura existente.

### **1. Enfoque metodológico:**

El estudio utiliza un enfoque cuantitativo y comparativo, que se centra en una revisión exhaustiva de la literatura científica publicada entre 2010 y 2022. A través de esta metodología, se recopilan datos de más de 60 artículos relevantes que aplican técnicas de aprendizaje automático para la predicción del crimen. La revisión se organiza en torno a ocho preguntas clave de investigación, que incluyen la evaluación de las fuentes de datos, los

algoritmos utilizados, la calidad de los modelos predictivos, y los desafíos encontrados por los investigadores.

## **2. Recopilación de datos**

Los datos fueron recopilados a partir de cinco bases de datos científicas confiables: ACM Digital Library, IEEE Xplore, Springerlink, Science Direct y Scopus. A través de búsquedas avanzadas con términos específicos como “crime prediction” y “machine learning”, se seleccionaron artículos que cumplieran con los criterios de inclusión definidos previamente. La búsqueda también incluyó una revisión de las introducciones y conclusiones de los artículos más recientes para identificar términos sinónimos y ampliar los parámetros de búsqueda.

## **3. Técnicas de análisis utilizadas**

Las técnicas de análisis en los estudios revisados se agruparon en dos enfoques principales: aprendizaje supervisado y aprendizaje no supervisado:

- Aprendizaje supervisado: Los algoritmos más comunes fueron Decision Trees, Naïve Bayes, Support Vector Machines (SVM) y regresión logística. Estos métodos requieren datos previamente etiquetados para entrenar los modelos y predecir el tipo de crimen o el perfil del criminal.
- Aprendizaje no supervisado: Se emplearon técnicas como K-Means para el clustering de patrones criminales. Este enfoque no requiere datos etiquetados y permite identificar áreas o grupos de delitos sin necesidad de categorías predefinidas.

## **4. Fases del estudio**

El proceso de la revisión se dividió en tres fases principales:

- Planeación de la revisión: Se identificaron los temas relevantes y se definieron las preguntas de investigación, como los objetivos de los estudios, las fuentes de datos utilizadas y las métricas de evaluación de los modelos.
- Realización de la revisión: Los artículos seleccionados fueron analizados y clasificados según los métodos de predicción utilizados, las fuentes de datos, y los enfoques de evaluación.
- Reporte de los resultados: Se sintetizaron los hallazgos y se respondieron las preguntas de investigación a través de la comparación de los estudios, con la ayuda

de gráficos, tablas y otros elementos visuales.

## 5. Evaluación y comparación

Se aplicaron varios criterios de calidad para seleccionar los estudios más relevantes, utilizando una escala de puntuación que evalúa aspectos como la claridad de los objetivos del estudio, la metodología utilizada, y la aplicabilidad de los resultados. Solo los estudios que superaron una puntuación mínima fueron incluidos en la revisión final. Además, se utilizaron métricas de rendimiento como Accuracy, Precision, Recall y F-value para comparar los resultados de los distintos algoritmos.

### Utilidad del artículo para el proyecto de tesis

Gracias a los aportes de este artículo, es de gran utilidad para nuestra tesis propuesta, ya que desarrollaremos un módulo de procesamiento de datos criminales que aplicará las técnicas de preprocesamiento y limpieza de datos discutidas en el estudio, como la gestión de datos faltantes y la eliminación de ruido, para asegurar la calidad de los datos históricos de delitos y registros policiales de Lima Metropolitana. Otra aportación de gran utilidad ya que, diseñaremos un módulo de modelado predictivo que emplea los algoritmos de Machine Learning supervisados más prometedores para la predicción de criminalidad, como Random Forest o SVM, para predecir hotspots de criminalidad en Lima Metropolitana. Además desarrollaremos módulo que generará reportes de mapas de calor interactivos que resaltan las áreas geográficas y temporales con mayor probabilidad de ocurrencia de delitos, proporcionando a las autoridades policiales información precisa para optimizar la asignación de recursos y estrategias de patrullaje preventivo en la capital.

En ese sentido, el artículo proporciona información valiosa que los modelos de minería de datos y el aprendizaje automático son fundamentales para predecir la ocurrencia de crímenes al analizar grandes volúmenes de datos históricos. La predicción de crímenes se basa en el uso de estos modelos para identificar patrones recurrentes en los delitos y anticipar su ocurrencia en el futuro. Además, el artículo destaca que el aprendizaje automático permite a los sistemas mejorar continuamente a medida que se alimentan con nuevos datos, lo que aumenta la precisión de las predicciones a lo largo del tiempo. En cuanto a los enfoques supervisados y no supervisados, se explican como dos de las principales técnicas utilizadas en la predicción del crimen. Los enfoques supervisados requieren datos etiquetados para entrenar los modelos y son particularmente útiles para clasificar tipos de



delitos, mientras que los enfoques no supervisados, como el clustering (por ejemplo, con K-Means), son efectivos para detectar patrones de criminalidad en áreas geográficas sin necesidad de etiquetas predefinidas.

De acuerdo a la figura 13, que ilustra la distribución de los objetivos en el estudio de la predicción de criminalidad, se indica cómo se abordan diversas áreas clave, como la predicción de hotspots criminales, la predicción de crímenes a través de modelos novedosos, y la selección de características relevantes. En este sentido, se tomará el enfoque de modelado predictivo con técnicas de Machine Learning supervisadas, como Random Forest y SVM, ya que estos modelos han demostrado ser efectivos para identificar patrones y predecir la ocurrencia de crímenes. Estos modelos serán aplicados específicamente para predecir hotspots de criminalidad en Lima Metropolitana, generando mapas de calor interactivos que resalten las áreas geográficas y temporales con mayor probabilidad de ocurrencia de delitos.

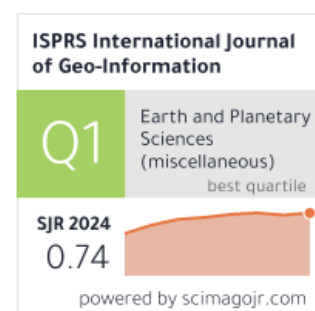
### 3.1.4 Crime Prediction and Monitoring in Porto, Portugal, Using Machine Learning, Spatial and Text Analytics

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, Cuartil Q1



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 para predecir 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 microescala 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 recopilamos tuits de Twitter utilizando la biblioteca Snsrape, un programa de búsqueda de servicios de redes sociales en Python.

## **C) Etapas**

Etapas 1: comprensión del patrón.

Etapla 2: Regresión Lasso.

Etapla 3: Clasificación con aprendizaje automático:

- Logistic Regression
- Decision Tree
- Random Forest
- SVM

Etapla 4: Análisis de sentimientos.

**Tabla 1**

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

Model	Accuracy	Recall	Precision	F1 Score
Logistic Regression(L1 penalty = 0.151)	0.65	0.84	0.64	0.72
Decision Tree(criterion = entropy, max depth = 3)	0.61	0.56	0.70	0.63
Random Forest(max. features = 2, number of trees = 100, max depth = 5)	0.83	0.99	0.79	0.89
SVM(kernel = rbf, C = 1, gamma = 0.1)	0.80	0.87	0.82	0.91

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

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

Gracias a los aportes de este artículo, es de gran utilidad para nuestra tesis propuesta, es por ello que [implementaremos un módulo de análisis espacial y de texto](#) que aplicará las metodologías de análisis espacial (como el Kernel Density Estimation) y análisis de texto (para extraer información de descripciones de delitos), tal como se emplean para la predicción y monitoreo de crímenes. Específicamente, este módulo

procesa datos geográficos de incidentes delictivos y las descripciones textuales de los registros policiales de Lima Metropolitana para identificar patrones espaciales y semánticos ocultos en la criminalidad urbana. Adicionalmente, implementaremos como utilidad un módulo de visualización interactiva de mapas que generará **mapas de calor de predicción de criminalidad** con capas de información espacial y textual (similar a las visualizaciones presentadas en el caso de estudio de Oporto), mostrando las áreas de alto riesgo, los tipos de delitos predominantes y factores contextuales relevantes en Lima Metropolitana, lo que permitirá a la Policía Nacional del Perú optimizar la asignación de recursos policiales de manera geográficamente *inteligente* y basada en el tipo de incidencia, mejorando la eficiencia y efectividad en la prevención y respuesta al crimen.

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.

### 3.1.5 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**

(Hannach & Benkhalifa, 2018)

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

**Citado 22 veces en Scopus, Cuartil Q3**



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

Etapla 1: Recopilación de datos de Twitter.

Etapla 2: Detección de oraciones con aspectos implícitos (IASD).

Etapla 3: Identificación de aspectos implícitos (IAI):

- Extracción de términos implícitos:
- Agregación de términos

Etapla 4: Análisis de sentimientos.

**Figura 14**

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

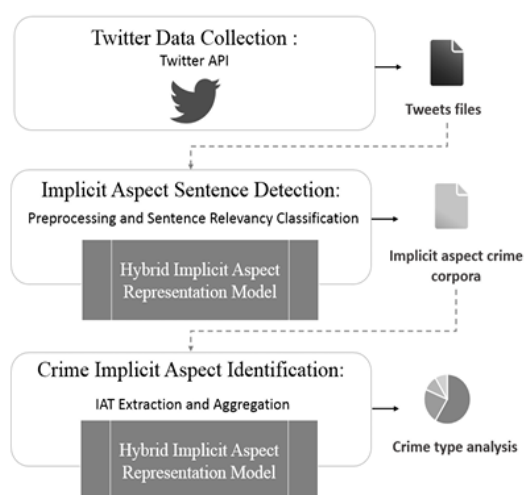


Fig. 1. Abstract Process of the Proposed Framework.

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

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

Gracias a los aportes de este artículo, es de gran utilidad para nuestra tesis propuesta, implementaremos un módulo de análisis de aspectos implícitos de criminalidad que procesa datos históricos de delitos en Lima Metropolitana aplicando el modelo híbrido WordNet con esquema TF-ICF que combina relaciones semánticas de sinónimos y definiciones para identificar patrones ocultos en registros policiales relacionados con delitos como homicidio, robo, secuestro y agresiones. El aplicativo integrará tres algoritmos de clasificación de machine learning (Multinomial Naïve Bayes, Support Vector Machine y Random Forest) Asimismo, incorporaremos como funcionalidad para la detección automática de aspectos implícitos criminales en los datos de incidencia delictiva y la clasificación posterior de tipos específicos de crímenes, permitiendo generar reportes predictivos de criminalidad por categorías que visualicen el porcentaje de aspectos implícitos detectados por tipo delictivo y zona geográfica, proporcionando a las autoridades policiales de Lima Metropolitana información procesada para optimizar la asignación preventiva de recursos. Adicionalmente, se implementará el módulo de preprocesamiento de datos criminales con técnicas de filtrado, extracción de características relevantes y ponderación de términos que permitan identificar patrones de criminalidad implícitos, creando una base de datos enriquecida que alimentará los modelos predictivos del aplicativo para mejorar la precisión en la predicción de zonas y tipos de crímenes potenciales en la capital.

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 Nave Bayes, SVM y Random Forest, que elevan la precisión predictiva del modelo; todos ellos son elementos fundamentales que aplicaremos directamente en nuestra tesis.

### 3.1.6 Survey on Crime Analysis and Prediction Using Data Mining and Machine Learning Techniques

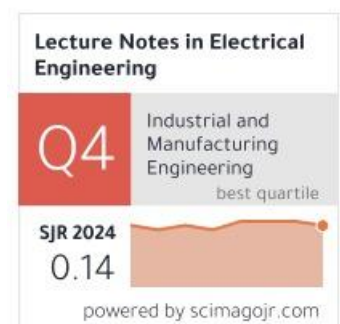
#### Encuesta sobre análisis y predicción de delitos mediante técnicas de minería de datos y aprendizaje automático

(Parthasarathy et al., 2020)

(DOI: [10.1007/978-981-15-7241-8\\_31](https://doi.org/10.1007/978-981-15-7241-8_31))

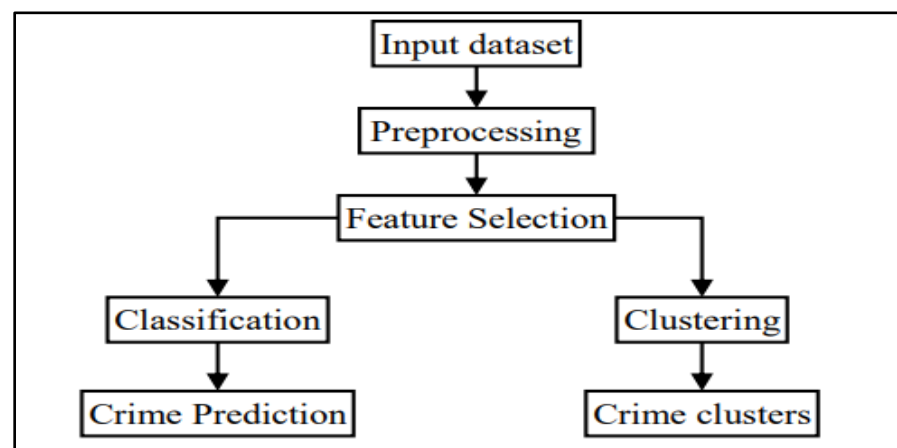
**Citado 34 veces en Scopus, Quartil Q4**

El artículo presenta una revisión exhaustiva sobre el uso de técnicas de minería de datos para el análisis y la predicción del crimen. Su objetivo principal es explorar



cómo los algoritmos de aprendizaje supervisado y no supervisado pueden aplicarse en el campo de la criminología para identificar patrones delictivos, predecir crímenes y reconocer perfiles criminales. La problemática central que aborda es la falta de integración efectiva de herramientas computacionales en los sistemas de justicia penal, a pesar del crecimiento exponencial de datos disponibles. Se destaca que muchas agencias policiales disponen de grandes bases de datos subutilizadas que podrían fortalecer la prevención y resolución de crímenes.

El estudio está motivado por la necesidad de proporcionar a investigadores emergentes un marco de referencia sobre enfoques actuales en minería de datos aplicada al crimen. A lo largo del artículo, se detallan distintos métodos divididos en cinco grandes categorías: métodos basados en texto y NLP, en patrones y evidencia, en geolocalización, en comunicación, y en registros penitenciarios. La contribución clave del trabajo es compilar, comparar y evaluar cualitativa y cuantitativamente estas metodologías, destacando su utilidad, fortalezas y limitaciones en contextos reales. Concluye proponiendo el uso de sistemas predictivos que integren algoritmos optimizados para apoyar a las agencias del orden en su labor preventiva y reactiva.



**Figura 15:** Predicción y agrupación de delitos según el conjunto de datos de entrada

## B) Metodología:

El enfoque metodológico del artículo es **cuantitativo y comparativo**, basado en una **revisión teórica sistemática** de investigaciones previas aplicadas al análisis y



predicción del crimen. Se utilizan criterios específicos para agrupar y analizar los métodos existentes:

- **Recopilación de datos:** El estudio recoge ejemplos de diversas fuentes como redes sociales, mapas geográficos, registros de comunicación, bases de datos de antecedentes penales y textos electrónicos (como correos o publicaciones).
- **Técnicas de análisis:** Se clasifican en:
  1. **Aprendizaje supervisado:** algoritmos como Decision Tree, Naïve Bayes, SVM y regresión logística.
  2. **Aprendizaje no supervisado:** principalmente K-Means para clustering de patrones criminales.
  3. **Optimización:** se menciona el uso de Algoritmos Genéticos para mejorar parámetros de clasificación.
- **Fases del estudio:** El artículo organiza los métodos en cinco categorías metodológicas:
  1. **Texto/NLP:** Identificación de amenazas en correos mediante árboles de decisión.
  2. **Evidencia/patrones:** Uso de clasificación y clustering para identificar sospechosos a partir de evidencias del lugar del crimen.
  3. **Geolocalización:** Predicción de zonas peligrosas usando datos espaciales y de redes sociales.
  4. **Comunicación:** Análisis de redes criminales para identificar líderes usando registros de llamadas.
  5. **Prisiones:** Evaluación del riesgo de reincidencia en presos con enfermedades mentales.

- **Evaluación y comparación:** Se presentan tablas con métricas de rendimiento como Accuracy, Precision, Recall y F-value para evaluar los distintos enfoques. Destaca que la combinación de algoritmos de clasificación con técnicas de optimización (como Decision Tree + GA) mejora significativamente los resultados.

### Utilidad del artículo para el proyecto de tesis

Gracias a los aportes de este artículo, es de gran utilidad para nuestra tesis propuesta, **es por ello que implementaremos un módulo de predicción de hotspots criminales** que, basándose en la categoría de geolocalización mencionada, analizará datos espaciales y temporales de delitos en Lima Metropolitana. Este módulo utilizará algoritmos de aprendizaje supervisado como Decision Tree o SVM, adaptados a nuestro contexto urbano. Adicionalmente, se imple un **módulo de visualización y reportes interactivos** que generará **mapas de calor de riesgo criminal** por zona y tipo de delito, similar a la "Figura 1: Predicción y agrupación de delitos según el conjunto de datos de entrada", permitiendo a la Policía Nacional del Perú en Lima Metropolitana la asignación estratégica y optimizada de recursos policiales en las áreas de mayor incidencia delictiva, fortaleciendo así la prevención y respuesta efectiva al crimen..

En el artículo, se menciona que la minería de datos es fundamental para explorar grandes volúmenes de datos y descubrir patrones útiles en el ámbito criminal. En este sentido, se aplicarán técnicas de aprendizaje supervisado como árboles de decisión, Naïve Bayes y SVM para clasificar los tipos de delitos y los perfiles de los delincuentes, lo cual ayudará a predecir crímenes específicos. Además, se utilizarán técnicas de aprendizaje no supervisado, como K-Means, para identificar patrones delictivos basados en la agrupación de incidentes criminales según características comunes, como el lugar y la hora de ocurrencia. Esto permitirá la predicción de áreas con mayor probabilidad de delitos, facilitando la toma de decisiones informadas. El artículo también resalta cómo predecir crímenes mediante estos enfoques puede mejorar la asignación de recursos y aumentar la eficiencia operativa de las fuerzas del orden. Estas técnicas, integradas en el diseño del aplicativo, contribuirán significativamente a optimizar la respuesta policial en Lima Metropolitana.

### 3.1.7 Epistemologies of predictive policing: Mathematical social science, social physics and machine learning

#### Epistemologías de la vigilancia predictiva: ciencias sociales matemáticas, física social y aprendizaje automático

(Hälterlein, 2021)

(DOI: [10.1177/20539517211003118](https://doi.org/10.1177/20539517211003118))



Citado 44 veces en Scopus, Cuartil Q1

El artículo ofrece un análisis profundo sobre las diferentes epistemologías que sustentan las prácticas de la policía predictiva, un enfoque que utiliza modelos algorítmicos para predecir futuros crímenes. La principal premisa del artículo es que la predicción del crimen, al depender de modelos computacionales, está influenciada por varias formas de conocimiento, las cuales afectan directamente cómo se generan y se aplican las predicciones en el sistema de justicia penal.

El artículo distingue tres epistemologías clave: la ciencia social matemática (MSS), la física social (SP) y el aprendizaje automático (ML). Cada una de estas epistemologías tiene implicaciones significativas para la creación y la interpretación del conocimiento predictivo. La MSS se enfoca en la formalización de teorías sobre el comportamiento social a través de modelos matemáticos, que buscan predecir el crimen a partir de supuestos teóricos como la teoría de la elección racional y la teoría de las actividades rutinarias. Por otro lado, la física social toma prestados métodos de la física, en particular modelos matemáticos utilizados para predecir fenómenos naturales, aplicándolos a la dinámica social y criminal, como lo demuestran los modelos de secuencias de aftershocks adaptados para prever crímenes. Finalmente, el enfoque de aprendizaje automático, basado en algoritmos que aprenden de grandes volúmenes de datos, permite generar predicciones sin necesidad de basarse en teorías preexistentes, lo que plantea un cambio radical en la forma de concebir el crimen y su predicción.

A lo largo del artículo, Hälterlein examina las implicaciones de estas epistemologías en términos de génesis del conocimiento, alcance de la predicción, comprensión del crimen y

accesibilidad del sistema. La principal crítica es que las diferentes epistemologías generan conocimientos predictivos que no siempre son accesibles ni comprensibles para los profesionales del sistema de justicia, lo que puede resultar en una falta de rendición de cuentas y de explicabilidad en las decisiones tomadas basadas en predicciones algorítmicas.

Constitution of knowledge	Epistemology		
	Mathematical social science	Social physics	Machine learning
Genesis/predictive modelling	Theory-driven/top-down		Data-driven/bottom-up
Scope/limitations of prediction	Explanatory power of theories	Mobility of predictive models	No limits <sup>13</sup>
Intelligibility/explanation of crime	Human agency	Natural process	No explanation
Accessibility/algorithmic accountability	White box <sup>14</sup>	Black box	

*Tabla 2: Implicaciones de las epistemologías.*

### B) Metodología

El enfoque metodológico de este artículo es de tipo cualitativo y comparativo, ya que el autor distingue y analiza las tres epistemologías principales en el contexto de la predicción del crimen. No se trata de una investigación experimental, sino de una revisión profunda de las teorías existentes que sustentan los modelos predictivos en la policía. A través de una recopilación de datos provenientes de estudios previos y un análisis crítico de las diferentes aproximaciones teóricas, el autor detalla cómo los modelos matemáticos y los enfoques basados en datos se aplican a la predicción del crimen.

La metodología se enfoca en la distinción de tres enfoques clave para modelar la criminalidad:

1. **Mathematical Social Science (MSS):** Un enfoque top-down donde los modelos se construyen a partir de teorías criminológicas.
2. **Social Physics (SP):** Un enfoque que usa modelos matemáticos inspirados en la física para predecir el crimen, basándose en la idea de que los fenómenos sociales siguen patrones similares a los de la física.
3. **Machine Learning (ML):** Un enfoque bottom-up basado en la exploración de patrones en grandes volúmenes de datos sin necesidad de teorías preexistentes.

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

Gracias a los aportes de este artículo, que son de gran utilidad para nuestra tesis propuesta, es por ello que **implementaremos un módulo de predicción basado en la epistemología del aprendizaje automático (Machine Learning)**. Este módulo procesará datos históricos de crímenes en Lima Metropolitana utilizando algoritmos de regresión logística y Random Forest, que han demostrado ser eficaces para modelar la probabilidad de ocurrencia de delitos en lugares y momentos específicos. Además, se aplicará un **modelo ETAS (Epidemic-Type Aftershock Sequence)** adaptado del ámbito de la física social para identificar la propagación de delitos en zonas específicas y predecir patrones de criminalidad relacionados con "repeticiones cercanas" de crímenes, lo cual optimizará la asignación de recursos policiales. Este enfoque de modelado sin necesidad de una teoría criminológica explícita, apoyado en el análisis de patrones de datos, permitirá una mayor adaptabilidad y precisión en la predicción de nuevas áreas de riesgo. Finalmente, visualizaciones geográficas se generarán para mostrar las zonas con mayor probabilidad de ocurrir delitos, mejorando la toma de decisiones por parte de las autoridades.

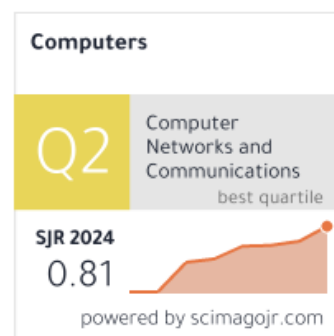
En el artículo, se menciona que la ciencia social matemática (MSS) se basa en teorías criminológicas y modelos matemáticos para predecir el crimen. En nuestro caso, aplicaremos este enfoque para construir modelos que integren teorías como la elección racional y las actividades rutinarias, lo cual nos permitirá identificar patrones en el comportamiento criminal y hacer predicciones fundamentadas sobre los lugares y tipos de delitos más comunes. Por otro lado, la física social (SP) utiliza modelos matemáticos inspirados en la física para predecir fenómenos sociales, en este caso, criminalidad, y será útil para analizar cómo los eventos criminales se difunden a través del espacio y el tiempo, algo que aplicaremos en nuestro sistema para entender las dinámicas de los delitos en áreas específicas de la ciudad. Finalmente, el aprendizaje automático (ML) nos permitirá usar grandes cantidades de datos para encontrar patrones y correlaciones sin necesidad de partir de teorías preexistentes, lo que facilitará la mejora continua del modelo predictivo conforme se integren más datos y se ajusten las predicciones basadas en las tendencias emergentes de criminalidad.

### 3.1.8 A Systematic Review of Using Machine Learning and Natural Language Processing in Smart Policing

Una revisión sistemática del uso del aprendizaje automático y el procesamiento del lenguaje natural en la vigilancia policial inteligente

(Sarzaeim et al., 2023)

(DOI:10.3390/computadoras12120255)



Citado 9 veces en Scopus, Cuartil Q2

Este artículo presenta una revisión sistemática de diversas aplicaciones del aprendizaje automático (ML) y el procesamiento del lenguaje natural (NLP) en la vigilancia policial inteligente, enfocándose en su papel en la mejora de la eficiencia y efectividad de las agencias de policía. El uso de tecnologías avanzadas como IA, ML y NLP en las prácticas policiales permite la gestión eficiente de grandes volúmenes de datos, lo que facilita tareas como la predicción del crimen, la identificación de patrones delictivos y la optimización de recursos en las intervenciones policiales.

El artículo destaca ejemplos específicos de aplicaciones de estas tecnologías, como la detección de huellas dactilares, el emparejamiento de ADN, la vigilancia mediante cámaras de CCTV y la predicción de delitos, donde los algoritmos de IA y ML han demostrado ser útiles para reducir los errores humanos y los sesgos inherentes en la toma de decisiones. Sin embargo, los autores subrayan la importancia de tener en cuenta que los algoritmos de IA reflejan los sesgos presentes en los datos con los que son entrenados, los cuales provienen de inputs humanos y pueden influir en los resultados.

A lo largo del estudio, se exploran diferentes enfoques metodológicos utilizados en la vigilancia policial inteligente, y se discuten tanto los beneficios como las limitaciones de la implementación de estos métodos. En particular, se aborda el uso de técnicas de ML, como algoritmos de clasificación y regresión, y su aplicación en la predicción de patrones delictivos. Asimismo, se menciona el uso de NLP para el análisis de informes narrativos de crímenes, lo cual facilita la extracción automática de información clave y la clasificación de datos en grandes volúmenes de documentos de texto no estructurado.

Los desafíos éticos asociados con el uso de IA en el ámbito policial también son una parte crucial de la discusión, ya que la implementación de estos algoritmos debe ser cuidadosa y responsable para evitar la discriminación o el uso indebido de los datos.

## **B) Metodología**

Mediante la presente revisión sistemática, se emplearon estrategias de búsqueda en bases de datos como IEEE y Google Scholar, utilizando palabras clave como Machine Learning (ML) y Natural Language Processing (NLP) en el contexto de la vigilancia policial inteligente. Se definieron dos preguntas clave: ¿Qué métodos de ML y NLP se han propuesto para procesar datos del crimen y predecir la actividad delictiva? y ¿Cuáles son las fortalezas y limitaciones de los métodos actuales y cómo pueden abordarse? Los estudios fueron seleccionados bajo estrictos criterios de inclusión, evaluando investigaciones en inglés relacionadas con la predicción del crimen. Se aplicó la técnica de snowballing para asegurar la inclusión de estudios clave, y en total se revisaron 103 artículos, de los cuales 45 fueron seleccionados como estudios primarios. Los estudios fueron clasificados en técnicas de ML, NLP y estadísticas, destacando el uso de algoritmos como SVM, Random Forest, y KDE, así como el reconocimiento de entidades nombradas (NER) para el análisis de informes del crimen. También se consideraron las implicaciones éticas y la aplicabilidad práctica de estos enfoques en la mejora de la eficiencia policial.

## **C) Etapas**

Etapa 1: Búsqueda y selección de estudios.

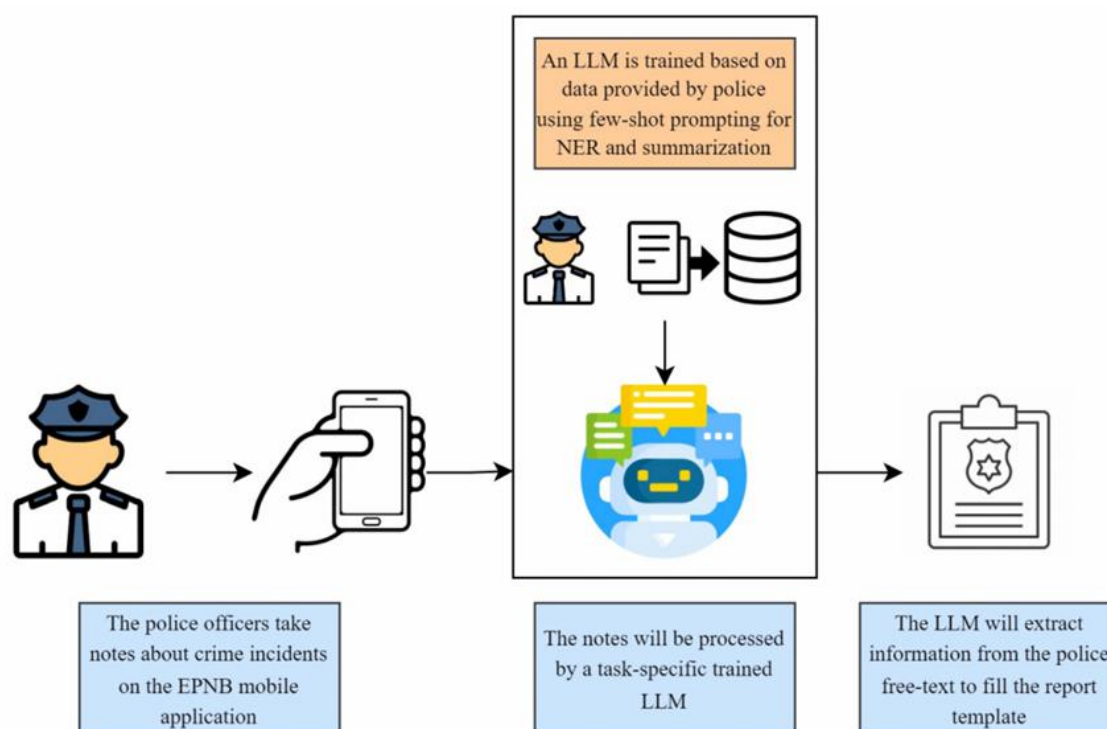
Etapa 2: Criterios de inclusión y snowballing para ampliar la búsqueda.

Etapa 3: Clasificación de los estudios según las técnicas utilizadas: ML, NLP y estadísticas.

Etapa 4: Análisis de la aplicabilidad práctica, precisión y evaluación ética de los métodos.

**Figura 16**

*Un esquema simplificado de la aplicación diseñada para Mobile Innovations. La aplicación EPNB está integrada en la nube de Microsoft Azure, que también ofrece servicios para usar tecnologías avanzadas como OpenAI y el Servicio de Lenguaje de Azure a través de llamadas API.*



**Nota. Fuente:** Sarzaeim, P., Mahmoud, Q. H., Azim, A., Bauer, G., & Bowles, I., (2023).

### Utilidad del artículo para el proyecto de tesis

Gracias a los aportes de este artículo, es de gran utilidad para nuestra tesis propuesta, es por ello que implementaremos un **módulo de procesamiento de lenguaje natural (NLP) para análisis de texto criminal**, siguiendo la revisión de las técnicas de NLP en el "Smart Policing" detalladas en el estudio. Este módulo procesa las descripciones textuales de los registros policiales y denuncias de Lima Metropolitana para extraer entidades, eventos y relaciones relevantes que ayuden a identificar patrones y contextos delictivos, lo cual es fundamental para una comprensión profunda del crimen. Adicionalmente, se implementará un **módulo de modelado predictivo** híbrido que combinará los algoritmos de Machine Learning supervisado con los datos enriquecidos por el NLP para predecir no solo la ubicación, sino también el tipo y las características contextuales del crimen en Lima Metropolitana, generando **reportes detallados y visualizaciones interactivas** que incluirán la semántica extraída de los textos, permitiendo a la Policía Nacional del Perú una asignación de recursos más precisa



y una planificación operativa que considere los elementos implícitos y explícitos de la criminalidad.

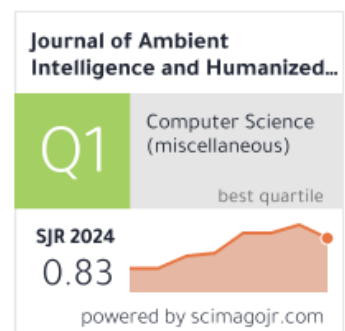
En el artículo destaca ejemplos específicos de aplicaciones de estas tecnologías, como la detección de huellas dactilares, el emparejamiento de ADN, la vigilancia mediante cámaras de CCTV y la predicción de delitos, donde los algoritmos de IA y ML han demostrado ser útiles para reducir los errores humanos y los sesgos inherentes en la toma de decisiones. Sin embargo, los autores subrayan la importancia de tener en cuenta que los algoritmos de IA reflejan los sesgos presentes en los datos con los que son entrenados, los cuales provienen de inputs humanos y pueden influir en los resultados.

### 3.1.9 Multi-UAV Allocation Framework for Predictive Crime Deterrence and Data Acquisition

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

(Miyano et al., 2020)

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



Citado 16 veces en Scopus, Cuartil Q1

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.

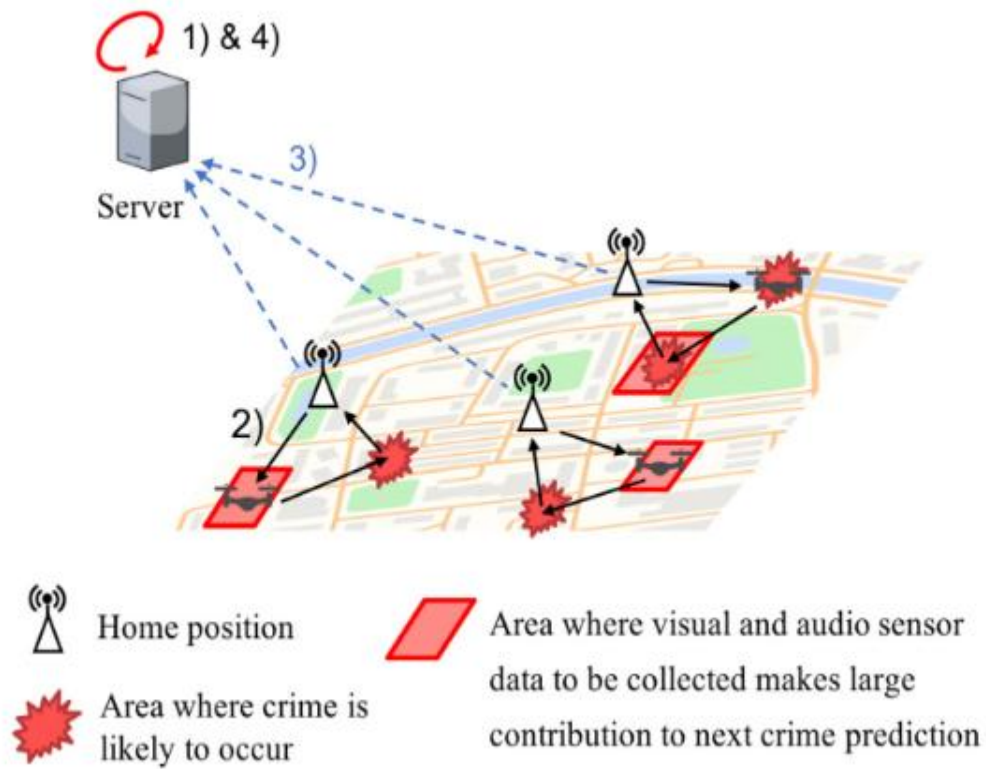
El marco se basa en un modelo predictivo que guía el despliegue de los UAVs no solo para disuadir incidentes delictivos, sino también para recolectar datos históricos relevantes que retroalimentan y mejoran la eficacia futura del sistema. Esta estrategia permite una optimización continua en la asignación 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.

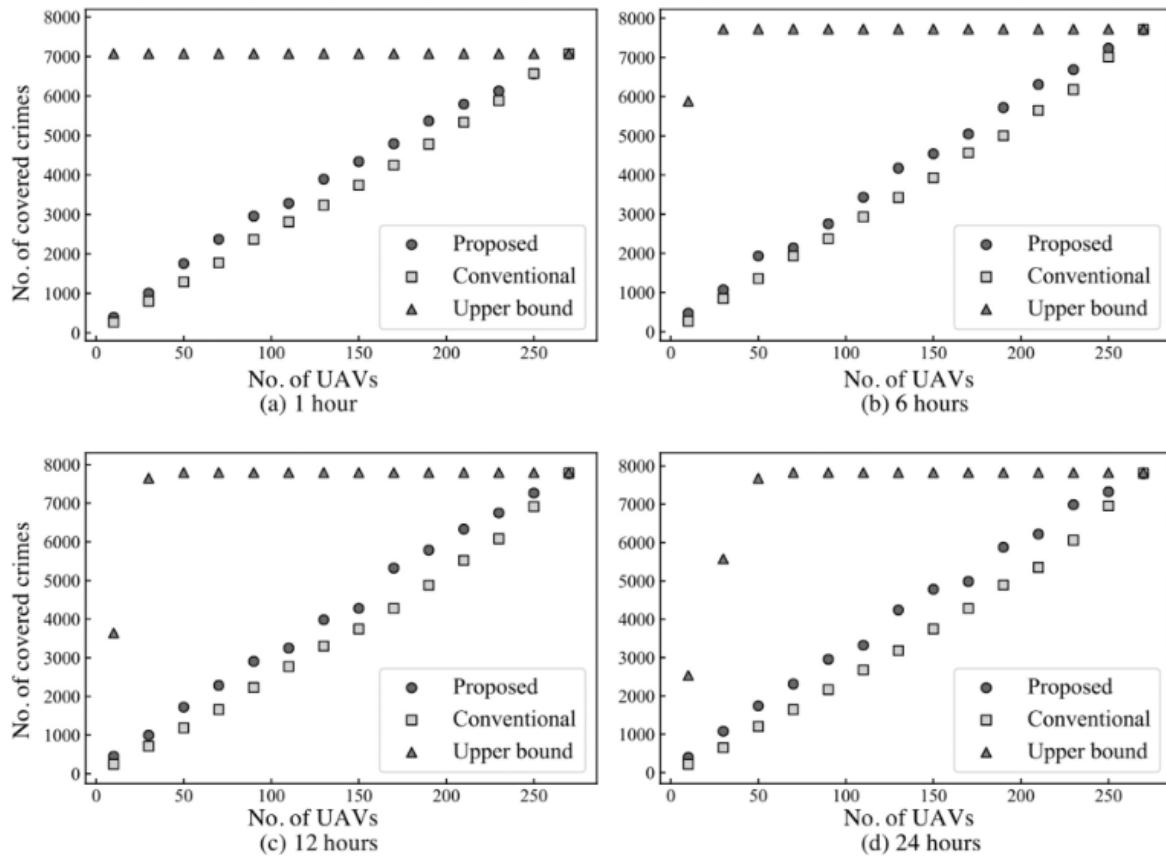
**Figura 17:** Arquitectura del sistema asumida en este estudio.



### 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 18:** 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

Gracias a los aportes de este artículo, es de gran utilidad para nuestra tesis propuesta, es por ello que implementaremos un **módulo de predicción de riesgo espacial y temporal** que utilizará modelos de Machine Learning como redes neuronales para analizar datos históricos de delitos en Lima Metropolitana. Este módulo se encargará de predecir las "times and places that are high-risk" para la ocurrencia de crímenes, siguiendo el enfoque del artículo. Adicionalmente, se integrará un **módulo de visualización estratégica de recursos** que generará **mapas de calor dinámicos y reportes de áreas de alta probabilidad de delitos**, permitiendo a la Policía Nacional del Perú en Lima Metropolitana optimizar la asignación de sus recursos y efectivos, simulando el despliegue de "unmanned aerial vehicles (UAVs)" como se discute en el artículo para el monitoreo y disuasión, aunque adaptado a recursos policiales

tangibles en nuestro contexto, con el fin de mejorar la disuasión del crimen y la capacidad de respuesta en zonas críticas.

En el artículo, se menciona que 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.

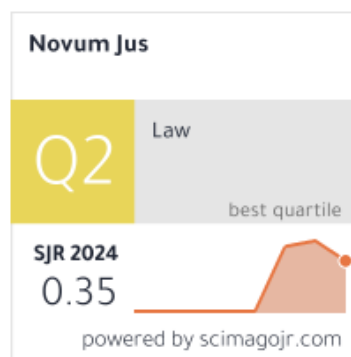
### 3.1.10 Predictive Criminology: A Near Future or a Distant Fiction?

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

(Salazar, 2024)

(DOI: [10.14718/NovumJus.2024.18.3.13](https://doi.org/10.14718/NovumJus.2024.18.3.13))

Citado 16 veces en Scopus, Cuartil Q2



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.

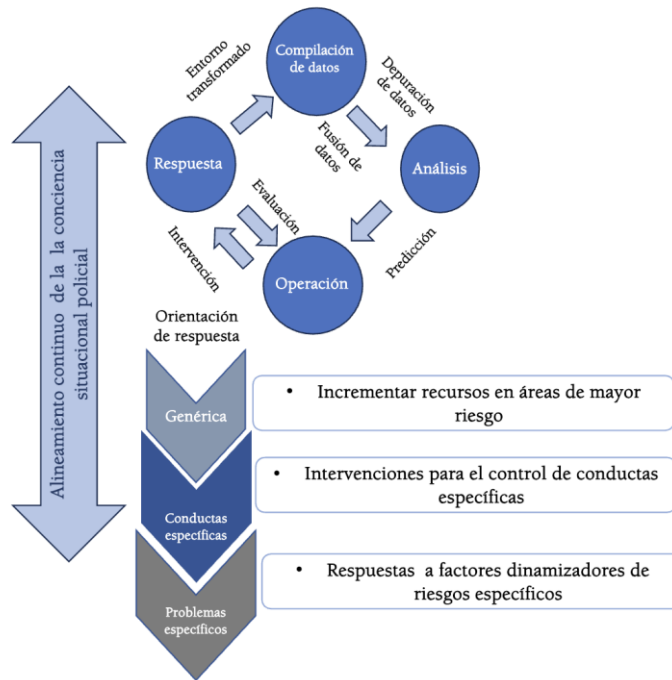
Pérez Salazar señala que si bien estos sistemas ofrecen beneficios operativos y contribuyen a la prevención del delito, presentan riesgos en cuanto a sesgos algorítmicos y falta de interpretación contextual. Aun así, se plantea que el uso de estos modelos puede fortalecer la capacidad del Estado para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

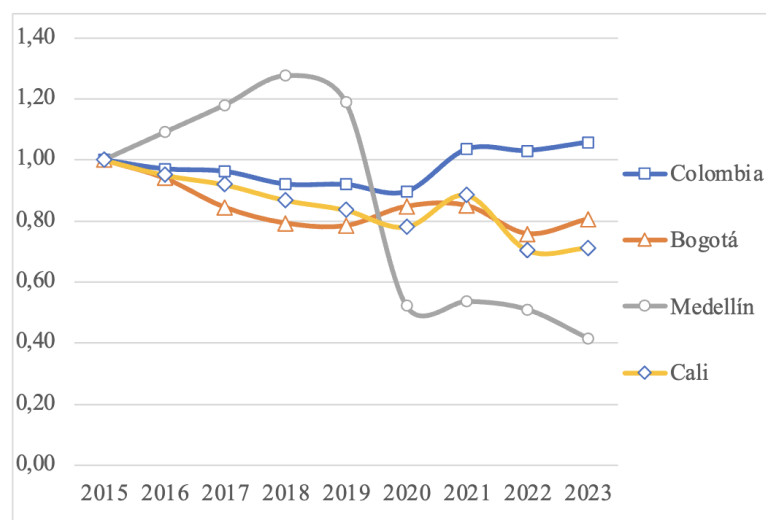
**Figura 19:** Ciclo de gestión del conocimiento en el enfoque de policía predictiva



### 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**

Gracias a los aportes de este artículo, es de gran utilidad para nuestra tesis propuesta, es por ello que implementaremos un **módulo de modelado de policía predictiva** que integrará los modelos estadísticos predictivos para analizar datos históricos de criminalidad en Lima Metropolitana. Este módulo se encargará de identificar patrones delictivos en grandes centros urbanos, siguiendo la problemática central que aborda el artículo sobre la falta de integración efectiva de herramientas computacionales. Adicionalmente, se implementará un **módulo de generación de informes y visualizaciones estratégicas** que producirá **reportes predictivos y mapas de riesgo interactivos** que detallarán las zonas, horarios y tipos de delitos con mayor probabilidad de ocurrencia. Esto permitirá a la Policía Nacional del Perú en Lima Metropolitana optimizar la asignación de recursos policiales de manera proactiva y basada en evidencia, mejorando la labor preventiva y reactiva frente a la criminalidad.

En el artículo, se menciona 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.

### **3.1.11 Institutional factors driving citizen perceptions of AI in government: Evidence from a survey experiment on policing**



## Factores institucionales que impulsan las percepciones ciudadanas de la IA en el gobierno: evidencia de un experimento de encuesta sobre la policía

(Schiff et al., 2023)

(DOI: [10.1111/puar.13754](https://doi.org/10.1111/puar.13754))

Citado 11 veces en Scopus, Cuartil Q1

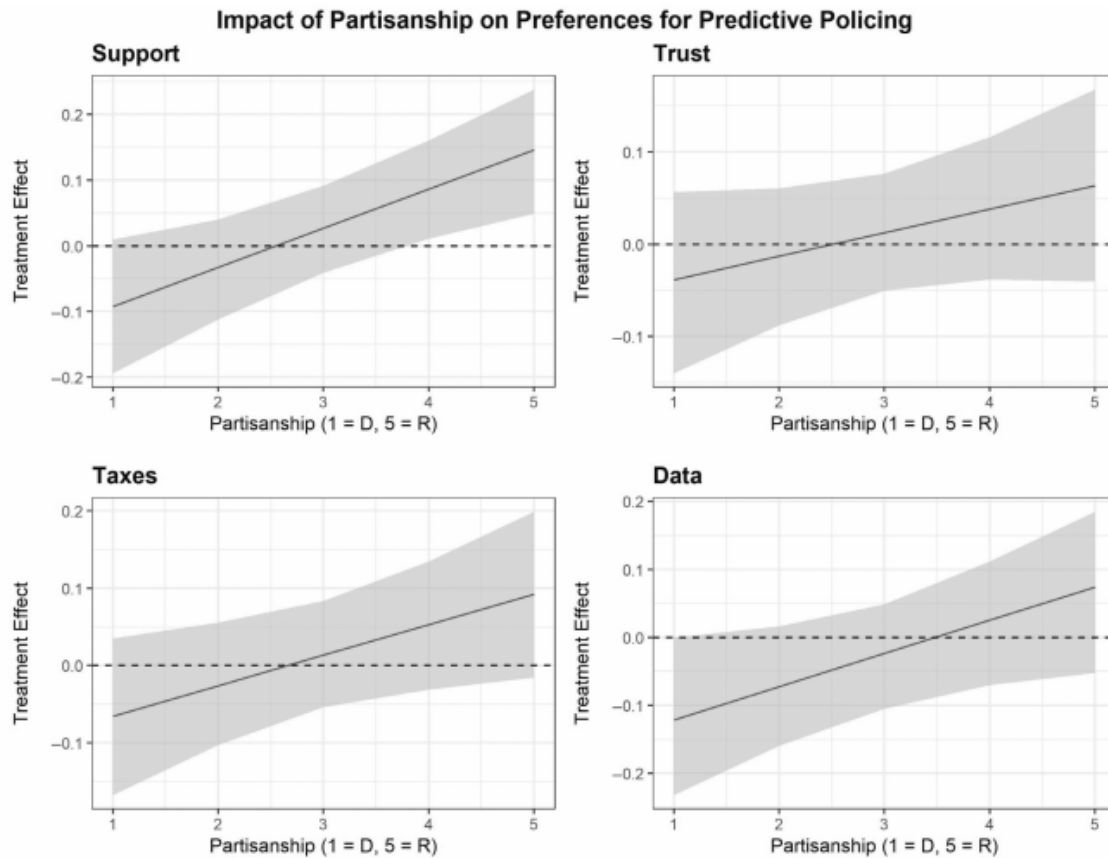


El artículo "Institutional factors driving citizen perceptions of AI in government" explora cómo los factores institucionales afectan la percepción pública sobre el uso de la inteligencia artificial (IA) en la administración pública, específicamente en el ámbito de la policía. El contexto de estudio se centra en cómo la confianza pública y el apoyo hacia la IA en el gobierno pueden variar dependiendo del contexto institucional. En particular, se investigan tres factores clave: la proximidad burocrática, los objetivos algorítmicos y la capacidad de las agencias para implementar herramientas de IA.

El objetivo principal del estudio es comprender cómo estos factores institucionales impactan la forma en que el público evalúa la implementación de IA en la policía. A través de un experimento de encuesta, el estudio analiza las percepciones de 4200 ciudadanos estadounidenses sobre el uso de IA en casos de policía predictiva y la detección de conductas indebidas de los oficiales a través de la revisión automatizada de casos. Los resultados clave revelan que los ciudadanos tienen una preferencia notable por el uso local de IA en la policía, particularmente cuando los actores responsables son alguaciles locales, en comparación con agencias nacionales como el FBI. Esta preferencia se debe a la mayor confianza en los alguaciles locales, quienes son percibidos como más cercanos y representativos de la comunidad. Además, aunque las percepciones sobre los objetivos algorítmicos (policía predictiva vs. revisión interna de los oficiales) son menos pronunciadas, se observan diferencias políticas y raciales en las respuestas.

En términos de la capacidad de las agencias, el estudio muestra que, aunque se reconocen las limitaciones de las agencias locales para implementar IA de manera efectiva, la capacidad institucional no influye significativamente en las actitudes del público. Esto sugiere que la confianza en los actores gubernamentales y la percepción de la proximidad burocrática son más influyentes que la capacidad técnica percibida de las agencias. El artículo concluye que para una implementación exitosa de IA en el gobierno, no solo se debe atender a las características técnicas de las herramientas, sino también a los factores organizacionales e institucionales, que juegan un papel crucial en la percepción pública de la IA.

**Figura 20 : Impact Of Partisanship**



### Metodología:

La metodología utilizada en este estudio es un diseño experimental pre-registrado que se centra en un experimento de encuesta de tipo factorial 2x2x2, donde se manipulan tres factores clave para evaluar cómo influyen en las percepciones del público sobre el uso de IA en la policía. Estos tres factores son:

1. **Proximidad burocrática:** se analiza la diferencia en la percepción pública sobre la IA cuando la herramienta es implementada por actores locales (alguaciles) versus actores nacionales (FBI). La hipótesis es que los ciudadanos prefieren la implementación local de IA debido a una mayor confianza en los actores más cercanos a la comunidad.
2. **Objetivos algorítmicos:** se comparan las reacciones del público hacia la IA utilizada para la policía predictiva (que impacta a la comunidad) frente a la utilizada para la revisión interna de conductas indebidas de los oficiales (que afecta a los empleados dentro de la institución). Se espera que los ciudadanos prefieran la IA dirigida a los

oficiales de policía debido a la creciente preocupación sobre la conducta policial.

3. **Capacidad de la agencia:** se investiga cómo la capacidad de una agencia (en términos de recursos y experiencia) influye en la confianza del público hacia la implementación de IA. La hipótesis es que las agencias con mayor capacidad deberían generar más confianza, pero el estudio también examina si esto es relevante para el público.

Para llevar a cabo el estudio, se reclutaron 4200 adultos estadounidenses a través de la organización de encuestas Data for Progress, que administró una encuesta en línea entre noviembre de 2021 y enero de 2022. Los participantes fueron asignados aleatoriamente a una de las ocho posibles combinaciones de las variables experimentales. Las respuestas se recopilaron en función de las actitudes hacia la IA, incluyendo la confianza en la agencia que implementa la IA, el apoyo a la implementación de la herramienta, la disposición a pagar más impuestos para financiar la IA y la disposición a compartir datos personales para mejorar la precisión de las herramientas de IA.

El análisis se realizó utilizando modelos de regresión OLS con errores estándar robustos, ajustando los resultados por las variables demográficas de los encuestados, como edad, género, raza, nivel educativo, ingreso, y estado de empleo. Además, se incluyeron medidas de control de atención para garantizar la calidad de las respuestas. El análisis también se centró en las diferencias de actitudes según las características partidarias y raciales de los participantes.

El diseño experimental permitió probar las hipótesis sobre la proximidad burocrática, los objetivos algorítmicos y la capacidad de las agencias de manera controlada, proporcionando resultados más robustos sobre las actitudes del público frente a la IA en la policía. Los resultados fueron consistentes y robustos a través de varias especificaciones de modelos y análisis de efectos heterogéneos según las características demográficas, como la afiliación política y la raza.

Los hallazgos clave incluyen que los ciudadanos prefieren la implementación local de IA, pero no muestran una clara preferencia entre los objetivos algorítmicos. Además, la capacidad de la agencia no tiene un impacto significativo en las actitudes del público, lo que sugiere que otros factores, como la proximidad institucional, son más influyentes. Esto resalta la importancia de considerar los factores institucionales y organizacionales al diseñar políticas de implementación de IA en el gobierno.

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

Gracias a los aportes de este artículo, que son de gran utilidad para nuestra tesis propuesta, es por ello que implementaremos un **módulo de análisis de aspectos implícitos de criminalidad** que procesará datos históricos de delitos en Lima Metropolitana. Este módulo aplicará el concepto de proximidad burocrática descrito en el artículo, priorizando el uso de inteligencia artificial (IA) en la policía local en lugar de agencias nacionales como el FBI, ya que se ha encontrado que el público prefiere el uso de IA por parte de las autoridades cercanas a su comunidad. Se generarán reportes predictivos que detallarán los tipos de crímenes y las zonas geográficas de mayor incidencia, lo cual permitirá a las autoridades policiales tomar decisiones informadas sobre la asignación de recursos. Esto se complementará con la implementación de un **módulo de preprocesamiento de datos**, que se encargará de extraer características relevantes de los registros policiales, utilizando técnicas como el análisis de relaciones semánticas y el filtrado de datos, lo cual ayudará a mejorar la precisión del sistema predictivo, alineándose con las mejores prácticas sugeridas en el artículo.

En el artículo, se menciona que los factores institucionales son fundamentales para determinar cómo el público percibe el uso de inteligencia artificial (IA) en la administración pública. La confianza pública y el apoyo hacia la IA dependen en gran medida de la proximidad burocrática, donde la cercanía de las autoridades locales genera mayor confianza en la implementación de herramientas tecnológicas, como el caso de los alguaciles locales frente a agencias nacionales como el FBI. Además, el estudio destaca la importancia de los objetivos algorítmicos, como el uso de la IA para la revisión automatizada de casos y la policía predictiva. La revisión automatizada de casos y la implementación de IA en la policía predictiva son aspectos clave en nuestra tesis, ya que estos enfoques ayudarán a optimizar la asignación de recursos policiales en tiempo real, basado en patrones de criminalidad previamente identificados por el algoritmo. Por último, el artículo resalta que, aunque la capacidad de la agencia para implementar IA es un factor relevante, lo más importante es cómo la proximidad burocrática y la relación de la comunidad con las autoridades influyen en la efectividad y la aceptación de estas tecnologías. Esto refuerza la importancia de considerar el contexto institucional en el diseño de nuestro sistema predictivo.

### 3.1.12 We have to talk about emotional AI and crime

#### Tenemos que hablar de la IA emocional y el crimen

(Podoletz, 2023)

(DOI: [10.1007/s00146-022-01435-w](https://doi.org/10.1007/s00146-022-01435-w))

Citado 29 veces en Scopus, Cuartil Q1



Este artículo aborda el uso emergente de la inteligencia artificial emocional (emotional AI) en contextos de policía, vigilancia y prevención del crimen, específicamente en espacios urbanos públicos. La autora examina críticamente cómo estas tecnologías, que intentan inferir estados emocionales o intenciones de las personas a partir de datos como expresiones faciales, tono de voz o movimientos corporales, están siendo consideradas para tareas de seguridad pública y policiacas, a pesar de la débil base científica que respalda su precisión. El objetivo central del artículo es evaluar las implicaciones éticas, jurídicas y sociales del uso de emotional AI en democracias liberales, argumentando que su implementación en el ámbito público debe evitarse.

La autora sostiene que, si bien estas tecnologías prometen contribuir a la prevención del delito mediante la inferencia de estados emocionales o intenciones a partir de señales como microexpresiones faciales, tono de voz o gestos corporales, la evidencia empírica que respalda su validez es sumamente limitada y cuestionable. En este sentido, se advierte que la implementación de dichas herramientas en contextos urbanos puede representar no solo una grave amenaza a la privacidad individual, sino también una preocupante expansión del poder policial bajo el paradigma del “pre-crimen”.

El artículo propone que la capacidad de estas tecnologías para generar predicciones probabilísticas sobre eventos futuros no justifica su uso en sociedades democráticas, donde los derechos fundamentales deben ser protegidos. Además, la autora argumenta que el despliegue de emotional AI en políticas de seguridad pública puede promover una forma de control social algorítmico, impactando negativamente en el derecho a la ciudad, la autonomía personal y la libre asociación.

#### **B: Metodología:**

El artículo adopta un enfoque crítico interdisciplinario dentro del marco de las ciencias sociales, específicamente desde la criminología, vigilancia tecnológica, derecho y estudios

urbanos. No se trata de un estudio experimental, sino de un análisis teórico-normativo con revisión documental y argumentación basada en estudios previos.

### **1. Contextualización histórica y tecnológica:**

Se traza la evolución del uso de biometría en la policía, diferenciando entre su aplicación para identificación y su uso especulativo para inferir rasgos delictivos o emocionales.

### **2. Evaluación de aplicaciones actuales:**

Se analizan las tecnologías de reconocimiento emocional aplicadas a la vigilancia urbana, destacando su potencial para integrarse con sistemas de videovigilancia, reconocimiento facial para realizar tareas de detección de anomalías o prevención.

### **3. Crítica ética y jurídica:**

Se discuten las implicancias legales de estas tecnologías, en términos de rendición de cuentas, transparencia algorítmica, impacto en derechos humanos y sesgo sistémico.

### **4. Reflexión estratégica:**

Se plantea que, aunque estas tecnologías prometen optimizar las tareas de seguridad ciudadana, aún no existen garantías suficientes respecto a su precisión, lo que podría derivar en discriminación algorítmica y vigilancia desproporcionada, especialmente en comunidades vulnerables.

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

Gracias a los aportes de este artículo, que son de gran utilidad para nuestra tesis propuesta, es por ello que **implementaremos un módulo de predicción de criminalidad basado en señales emocionales** que utilizará datos de expresiones faciales, movimientos corporales, tono de voz y comportamiento verbal para identificar patrones que preceden posibles delitos en áreas urbanas de Lima Metropolitana. Siguiendo la línea de análisis de AI emocional planteada en el artículo, aplicaremos un modelo de machine learning que entrenará con datos históricos de incidentes delictivos y las señales emocionales de los involucrados, para inferir comportamientos agresivos o intenciones delictivas en eventos públicos. Este modelo será integrado al **módulo de vigilancia en tiempo real**, capaz de identificar señales emocionales asociadas con riesgos de violencia en lugares públicos, utilizando cámaras de CCTV y otras tecnologías de reconocimiento de emociones que pueden predecir la probabilidad de comportamientos violentos. A partir de los análisis realizados, se generarán reportes predictivos que permitirán visualizar los riesgos emocionales asociados a lugares y momentos

específicos, y así proporcionar información valiosa a las autoridades policiales para la asignación de recursos de manera más eficiente y preventiva. Además, se incluirá un **módulo de análisis contextual** que tomará en cuenta las características emocionales y contextuales del comportamiento humano en situaciones de alto riesgo, mejorando la precisión del sistema predictivo en la detección de posibles eventos delictivos.

En el artículo, se menciona que la inteligencia artificial emocional (emotional AI) está siendo aplicada en sistemas de vigilancia y prevención del crimen, a pesar de que su base científica es débil y su uso representa riesgos significativos para los derechos humanos. Aunque estas tecnologías se presentan como una solución para la prevención del delito, su implementación puede derivar en mecanismos de control excesivo y discriminación. También se discute que este tipo de IA busca generar predicciones probabilísticas sobre eventos futuros a partir de gestos, expresiones y comportamientos, sin una validación empírica sólida. Esta reflexión crítica refuerza la importancia de que nuestro proyecto utilice técnicas de aprendizaje automático éticamente fundamentadas, basadas en datos reales y objetivos, evitando así caer en prácticas de vigilancia desproporcionada o decisiones policiales automatizadas sin justificación confiable.

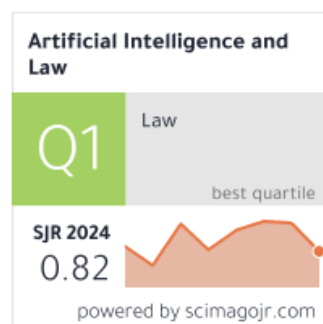
### 3.1.13 Smart criminal justice: exploring the use of algorithms in the Swiss criminal justice system

#### Justicia penal inteligente: exploración del uso de algoritmos en el sistema de justicia penal suizo

(Simmler et al., 2023)

(DOI: [10.1007/s10506-022-09310-1](https://doi.org/10.1007/s10506-022-09310-1))

Citado 12 veces en Scopus, Cuartil Q1



El artículo examina el estado actual y las implicancias del uso de algoritmos en el sistema de justicia penal suizo, en un contexto donde la transformación digital y el desarrollo de tecnologías inteligentes están remodelando la gestión pública y, en particular, la estrategia policial y la seguridad ciudadana.

El estudio parte del reconocimiento de que el uso de [algoritmos de decisión automatizada](#), algunos de ellos dotados de [aprendizaje automático](#), está ganando presencia en procesos clave de prevención del delito, análisis criminal y administración penal. En particular, se identifican [herramientas aplicadas en policía predictiva, análisis de patrones delictivos, evaluación del riesgo de reincidencia y gestión de casos en el sistema penitenciario](#). Estas soluciones tecnológicas se sustentan en el uso de datos para generar predicciones que permiten optimizar la [asignación de recursos policiales y penitenciarios](#), con el fin último de mejorar la eficiencia institucional.

El trabajo se estructura en torno a tres ejes: la identificación de las aplicaciones algorítmicas actualmente implementadas; los motivos institucionales y políticos que impulsaron su adopción; y la evaluación de su impacto percibido en la práctica cotidiana de operadores policiales y penitenciarios. Entre los resultados más relevantes, se destaca que si bien los algoritmos ya están integrados en varias áreas de la justicia penal suiza, su nivel de complejidad y autonomía es bajo, predominando modelos lineales o estructuras tipo checklist, más cercanas a la digitalización administrativa que al uso pleno de inteligencia artificial.

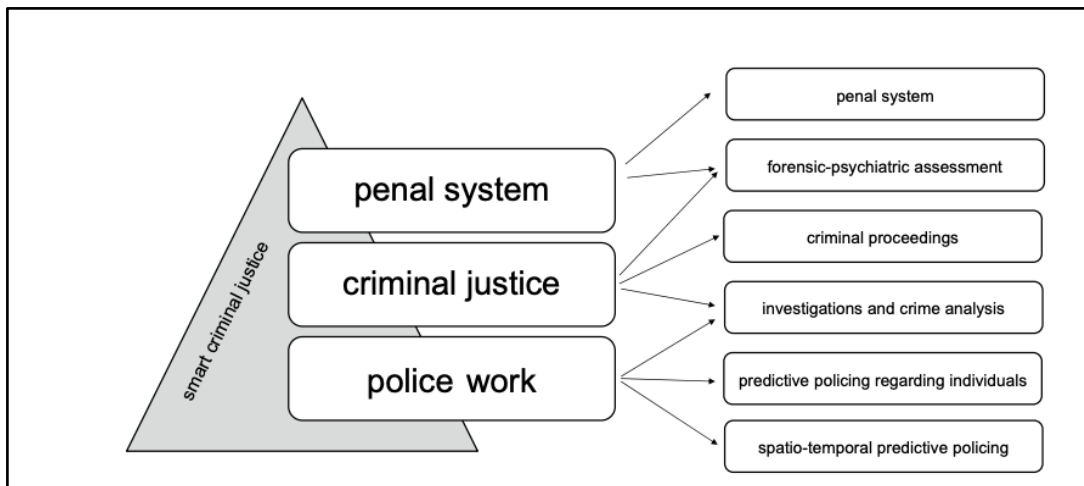
Este trabajo representa una importante [contribución al estudio de la justicia penal inteligente, proporcionando evidencia empírica relevante](#) y estableciendo las bases para una futura agenda de investigación centrada en el diseño, [implementación y evaluación de tecnologías predictivas aplicadas al sistema penal](#), siempre en coherencia con los principios del Estado de derecho y la garantía de derechos fundamentales.

## **B: Metodología:**

El estudio adopta un enfoque cualitativo y exploratorio, con un diseño empírico-descriptivo centrado en entrevistas semiestructuradas. Su objetivo es captar percepciones y experiencias directas de los usuarios de herramientas algorítmicas en el ámbito penal suizo. Se realizaron 25 entrevistas personales con 32 expertos provenientes de 14 cantones. Los participantes incluían policías, juristas, psicólogos forenses, informáticos y trabajadores sociales y se utilizó una guía temática uniforme que incluía preguntas sobre herramientas utilizadas, motivos de adquisición, funcionamiento, efectos, bases legales y desafíos.

***Figura 21: Categorías de justicia penal inteligente***





La información recogida fue procesada mediante análisis temático codificado en dos fases. Primero, se identificaron patrones recurrentes en las respuestas; luego, estas se organizaron según un sistema de codificación numérica, lo que permitió analizar transversalmente las entrevistas. Se distinguieron además diferencias entre grupos (policía vs sistema penitenciario, expertos legales y técnicos). La integración de los datos provenientes del sondeo por correo electrónico se realizó bajo el mismo esquema analítico.

#### **Fases del estudio:**

1. Contacto y validación de expertos institucionales.
2. Recolección de datos mediante entrevistas presenciales y encuestas.
3. Transcripción literal de los testimonios.
4. Codificación temática y categorización comparativa.
5. Análisis conjunto de tendencias y elaboración de conclusiones.

Los resultados permiten concluir que, si bien las tecnologías actuales no desplazan la toma de decisiones humanas, sí actúan como instrumentos de estructuración, estandarización y apoyo, generando valor al facilitar la prevención y el uso más eficiente de los recursos en contextos de seguridad ciudadana. No obstante, los autores alertan sobre el riesgo de una pseudo-legitimación algorítmica, donde las decisiones humanas son simplemente respaldadas a posteriori por los resultados del sistema, sin un verdadero contraste crítico.

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

Gracias a los aportes de este artículo, es de gran utilidad para nuestra tesis propuesta, es por ello que **desarrollaremos un módulo de predicción espacial-temporal de criminalidad** que procesa datos históricos de delitos en Lima Metropolitana. Este módulo se basará en el uso del enfoque de predicción de crimen de tipo "near repeat", que permite identificar zonas con mayor riesgo de ser blanco de delitos, tal como se describe en el artículo. Otra aportación de gran utilidad es que **diseñaremos un módulo de geoespaciales y técnicas de minería de datos** para generar perfiles de riesgo de áreas específicas, enfocándonos en la predicción de crímenes recurrentes a partir de datos históricos de delitos, como robos y hurtos.. Además, **desarrollaremos reportes predictivos que visualizarán las áreas y momentos de mayor probabilidad de ocurrencia de crímenes**, lo que optimizará la asignación de recursos policiales y la toma de decisiones informadas.

En este sentido, el artículo proporciona información valiosa sobre los avances tecnológicos están remodelando la **estrategia policial y la seguridad ciudadana**, mediante el uso de **algoritmos de decisión automatizada**, algunos de ellos alimentados por **aprendizaje automático**. Estas herramientas permiten optimizar la **asignación de recursos policiales y penitenciarios**, generando beneficios operativos concretos en la gestión institucional. Asimismo, el texto destaca la necesidad de una correcta **implementación y evaluación de tecnologías predictivas aplicadas al sistema penal**, reconociendo tanto sus ventajas como los riesgos de una automatización sin supervisión crítica. Esta perspectiva respalda la orientación de nuestro proyecto hacia un desarrollo tecnológico responsable, que complemente el trabajo policial sin desplazar la evaluación humana ni comprometer derechos fundamentales.

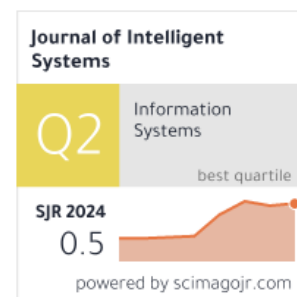
#### **3.1.14 A study on predicting crime rates through machine learning and data mining using text**

**Un estudio sobre la predicción de las tasas de delincuencia mediante el aprendizaje automático y la minería de datos utilizando texto**

(Saeed & Abdulmohsin, 2023)

(DOI: [10.1515/jisys-2022-0223](https://doi.org/10.1515/jisys-2022-0223))

**Citado 15 veces en Scopus, Cuartil Q1**



Este artículo examina el uso de técnicas de aprendizaje automático y minería de datos para la predicción de tasas de criminalidad, destacando la importancia de estas herramientas para mejorar la seguridad pública. La principal problemática que se aborda es la limitación de los métodos tradicionales, como el análisis manual y los enfoques estadísticos convencionales, que no son lo suficientemente eficaces para predecir de manera precisa la ubicación y el momento exacto de los crímenes. Esto genera dificultades en la asignación de recursos policiales y en la planificación de estrategias de prevención.

Con el avance de la tecnología, el uso de técnicas de minería de datos y [aprendizaje automático](#) ha permitido mejorar considerablemente la precisión en la predicción de crímenes. En particular, el artículo explora cómo los métodos supervisados, como la [regresión logística](#), [las máquinas de soporte vectorial \(SVM\)](#) y otros algoritmos de clasificación, han demostrado ser más eficaces que los métodos tradicionales. Además, se abordan los métodos no supervisados, como el [clustering](#), que también se han utilizado para identificar patrones de comportamiento criminal. A través de la aplicación de estos algoritmos, los investigadores han podido identificar áreas de alta criminalidad y predecir las tendencias delictivas futuras basándose en datos históricos de crímenes.

El artículo también hace un análisis comparativo de diferentes enfoques utilizados en investigaciones previas, mostrando cómo las técnicas de aprendizaje automático pueden proporcionar predicciones más precisas al trabajar con grandes volúmenes de datos espaciales y temporales. A pesar de los avances en este campo, se reconoce que la calidad de los datos y la heterogeneidad de las fuentes de información representan desafíos importantes. Por ejemplo, la variabilidad en los tipos de crímenes y la complejidad de las relaciones entre las variables espaciales y temporales dificultan la creación de un modelo único que funcione de manera eficiente en todos los contextos.

**Tabla 3:** Estudio de la literatura sobre trabajos de investigación de predicción del delito con minería de datos

Ref.	Year	Method	Dataset	Classification technique	Acc. %
[62]	2013	Data mining	Different states of USA	DT	83.9
[61]	2014	Data mining	India crime data	WEKA on two K-mean clusters	93.62 for C1 93.99 for C2
[63]	2015	Data mining	Denver crime data	NB	51
				DT	42
			Los Angeles crime data	NB	54
				DT	43
[64]	2017	Data mining	Chicago crime data	DT	75.9

## B) Metodología

El enfoque metodológico del artículo es una revisión sistemática de la literatura, con un enfoque en el análisis comparativo de diferentes algoritmos de aprendizaje automático y minería de datos aplicados a la predicción de crímenes. La investigación se divide en dos etapas principales: primero, la recopilación de estudios relevantes sobre la predicción de crímenes utilizando técnicas de aprendizaje automático; y segundo, la clasificación y análisis de estos estudios, evaluando la precisión de los métodos utilizados.

Los pasos principales de la metodología incluyen:

1. **Recopilación de estudios relevantes:** Se seleccionaron artículos publicados entre 2001 y 2022 que aborden la predicción de crímenes a través de técnicas de aprendizaje automático y minería de datos. Se prestó especial atención a la precisión reportada por cada estudio.
2. **Análisis de algoritmos utilizados:** Se revisaron y analizaron los diversos algoritmos aplicados en los estudios, incluyendo la regresión logística, SVM, Naïve Bayes, y K-Nearest Neighbors (KNN). También se evaluaron los enfoques de aprendizaje supervisado y no supervisado para determinar cuáles ofrecían mejores resultados.
3. **Evaluación de la precisión y desempeño:** Se presentó una tabla comparativa con los resultados obtenidos de diferentes algoritmos aplicados a conjuntos de datos específicos de crímenes, como los de Chicago y Los Ángeles. Se analizó cómo la elección del algoritmo y el tipo de datos afectaron los resultados de las predicciones.
4. **Desafíos en la predicción de crímenes:** Se discutieron los principales desafíos en el campo, como la falta de datos completos, la variabilidad en la calidad de los datos, y la dificultad de aplicar los mismos algoritmos a diferentes conjuntos de datos.

### Utilidad del artículo para el proyecto de tesis

Gracias a los aportes de este artículo, que son de gran utilidad para nuestra tesis propuesta, es por ello que **desarrollaremos un módulo de predicción de crímenes que utilizará técnicas de aprendizaje automático basadas en los algoritmos descritos en el artículo. Este módulo procesa datos históricos de delitos y registros policiales de Lima Metropolitana, con el objetivo de predecir zonas de alta criminalidad y tendencias futuras.** Otra aportación de

gran utilidad es que **diseñaremos un módulo de visualización interactiva que generará reportes predictivos** detallados, mostrando las zonas de riesgo y la evolución temporales de crímenes, identificando patrones de criminalidad a través de la minería de datos y las técnicas de clustering también será utilizada para segmentar áreas con características similares, lo que facilitará la asignación eficiente de recursos policiales, tal como lo proponen los modelos de predicción discutidos en el artículo.

En ese sentido, el artículo proporciona información valiosa sobre cómo el **aprendizaje automático** es esencial para la predicción de crímenes, modelando patrones complejos en grandes volúmenes de datos. También se destaca la efectividad de la **regresión logística**, que utilizaremos para predecir la probabilidad de crímenes en áreas específicas. Además, el artículo resalta el uso de **máquinas de soporte vectorial (SVM)**, que ayudarán a clasificar las zonas de alto riesgo. Finalmente, se menciona el uso de **clustering**, un método no supervisado útil para identificar patrones criminales y ajustar estrategias preventivas en zonas de alta incidencia delictiva.

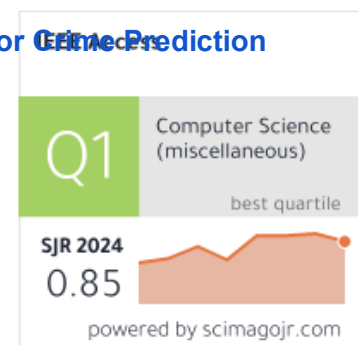
### 3.1.15 An Empirical Analysis of Machine Learning Algorithms for Crime Prediction Using Stacked Generalization: An Ensemble Approach

**Un análisis empírico de algoritmos de aprendizaje automático para la predicción de delitos mediante generalización apilada: un enfoque de conjunto**

(Kshatri et al., 2021)

(DOI: [10.1109/ACCESS.2021.3075140](https://doi.org/10.1109/ACCESS.2021.3075140))

**Citado 124 veces en Scopus, Cuartil Q1**

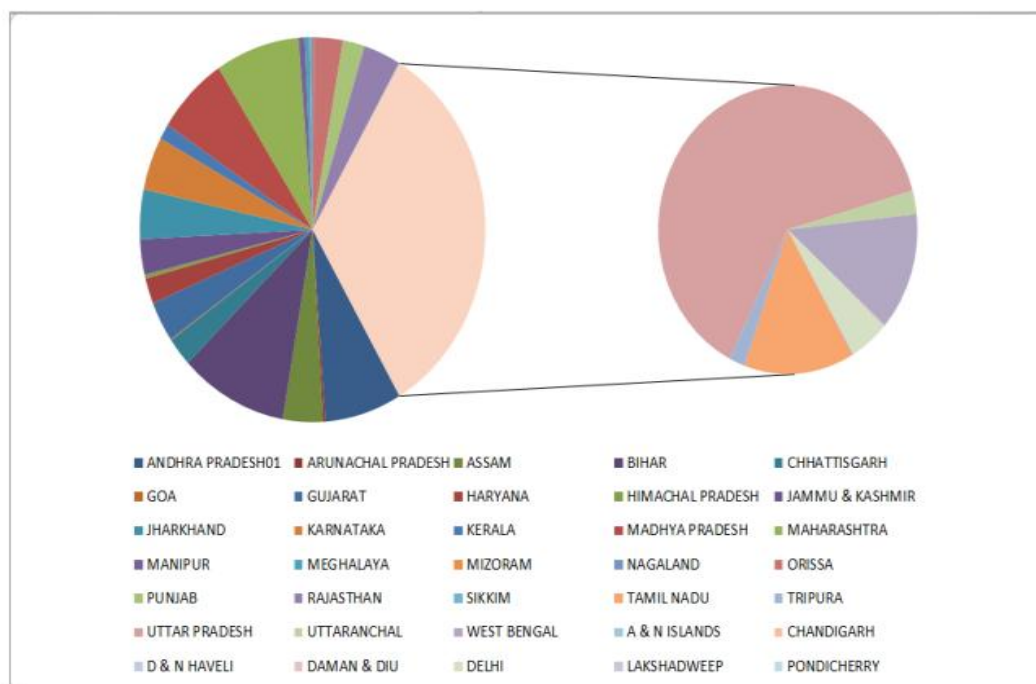


Este artículo tiene como objetivo evaluar el desempeño de distintos algoritmos de aprendizaje automático para predecir delitos, específicamente crímenes violentos. Utilizando técnicas avanzadas de **generalización apilada (stacked generalization)** y **aprendizaje en conjunto (ensemble learning)**, el estudio compara varios modelos de clasificación, como SVM, J48, SMO, Naïve Bayes, y Random Forest, con el fin de identificar el modelo más efectivo para predecir crímenes en diferentes áreas. Los resultados clave indican que el modelo propuesto, basado en **SVM stacking**, alcanzó una precisión de clasificación del 99.5%, superando a otros modelos tradicionales que solo lograron entre 95.55% y 97.21% de precisión.

La problemática central que aborda el artículo es la alta tasa de crímenes violentos y la necesidad de sistemas predictivos más efectivos que ayuden a las autoridades a asignar

recursos de manera eficiente para prevenir delitos. A pesar de los avances en el uso de **machine learning** para la predicción de crímenes, los modelos existentes tienen limitaciones, como la falta de precisión en algunas áreas y la dificultad de interpretar los resultados. El artículo se relaciona con el tema de estudio al aplicar tecnologías avanzadas de **inteligencia artificial (IA)** para mejorar la capacidad predictiva de los sistemas de prevención de crímenes, específicamente en India, donde la tasa de crímenes violentos es elevada.

La metodología del artículo está orientada a la investigación empírica con un enfoque comparativo entre diferentes algoritmos de machine learning. El artículo utiliza un conjunto de datos sobre crímenes violentos en India, y evalúa el desempeño de los algoritmos mediante el uso de técnicas de **validación cruzada**. Los resultados obtenidos demuestran que el enfoque de ensemble stacking tiene el mejor desempeño, lo que sugiere que combinar varios modelos puede ser más efectivo que usar un solo modelo de clasificación.



**Figura 22:** Distribución de la delincuencia en las regiones seleccionadas.

## B) Metodología

El enfoque metodológico adoptado es **investigación empírica comparativa**, en la que se realiza un análisis detallado de varios algoritmos de machine learning aplicados a la predicción de crímenes. Se recopiló un conjunto de datos de crímenes violentos en India,

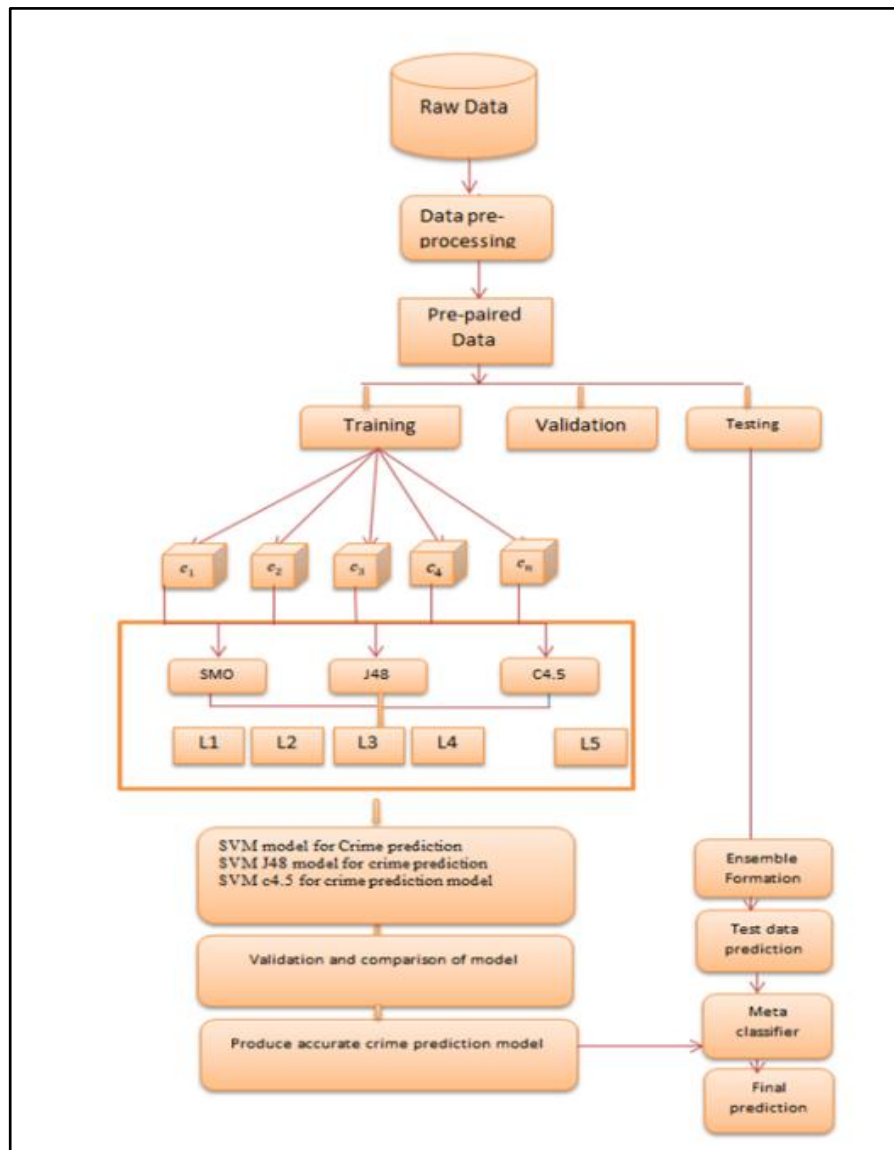
utilizando fuentes como el **National Crime Record Bureau (NCRB)**. Los datos fueron preprocesados para eliminar valores faltantes y ruidosos, y luego se aplicaron técnicas de **5-fold cross-validation** para evaluar la precisión de los modelos propuestos.

Los principales pasos de la metodología incluyen:

4. **Recopilación de datos:** Se reunieron datos de crímenes violentos en India, abarcando tipos como homicidios, violaciones, robos y secuestros, entre otros, con información sobre la naturaleza del crimen, el tiempo y la ubicación.
5. **Preprocesamiento de datos:** El conjunto de datos fue limpiado y transformado para asegurar que no hubiera valores faltantes ni inconsistencias. Se redujo el tamaño de los datos eliminando registros incompletos.
6. **Evaluación de modelos:** Se probaron seis algoritmos de machine learning, incluyendo **J48**, **SMO**, **Naïve Bayes**, **Random Forest**, y el **modelo de apilamiento (stacking)** basado en **SVM**. Los resultados se compararon en términos de **precisión**, **sensibilidad**, **especificidad**, **error cuadrático medio (RMSE)** y **error absoluto medio (MAE)**.
7. **Modelo de apilamiento:** Se aplicó un enfoque de **ensemble stacking** para combinar las predicciones de varios clasificadores y mejorar la precisión de las predicciones.

Las fases del estudio incluyen:

1. **Entrenamiento de modelos:** Se entrenaron los modelos de clasificación utilizando un conjunto de datos de entrenamiento (80% de los datos).
2. **Validación y prueba de modelos:** El conjunto de datos restante (20%) se utilizó para validar y probar los modelos, midiendo su rendimiento mediante métricas estándar de precisión y error.
3. **Análisis comparativo:** Se compararon los resultados obtenidos de los diferentes algoritmos de machine learning y se evaluaron las ventajas del modelo de apilamiento.



**Figura 23:** Clasificador propuesto para el modelo de predicción de delitos basado en conjuntos.

### Utilidad del artículo para el proyecto de tesis

Gracias a los aportes de este artículo, es de gran utilidad para nuestra tesis propuesta, es por ello que implementaremos un **modelo de predicción basado en ensemble stacking** para identificar patrones criminales, tal como se sugiere en el estudio. Este enfoque mejorará la precisión en la predicción de tipos de crímenes y sus ubicaciones en Lima Metropolitana. Utilizaremos un **clasificador SVM apilado**, que combina diferentes modelos base como J48,



C4.5, y SVM, para crear una estructura robusta capaz de manejar la variabilidad y la dinámica de los datos criminales. Además, se implementará un **módulo de preprocesamiento de datos** que limpiará, transformará y preparará los datos históricos de crímenes, integrando técnicas de normalización y validación cruzada, con el fin de mejorar la calidad de las predicciones. Los reportes generados por el sistema mostrarán la probabilidad de ocurrencia de delitos violentos como homicidios y robos, permitiendo a las autoridades **optimizar la asignación de recursos**. Este enfoque será clave para maximizar la eficiencia del sistema y facilitar decisiones informadas basadas en datos históricos.

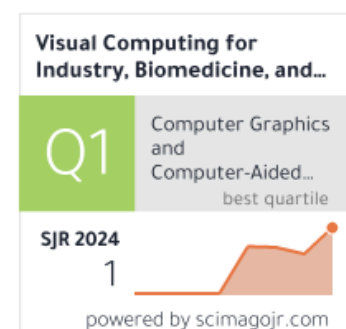
En el artículo, se menciona que los modelos de generalización apilada (stacked generalization) combinan varios clasificadores para mejorar la precisión de las predicciones. Este enfoque es esencial para nuestro proyecto, ya que permite optimizar el rendimiento predictivo al combinar lo mejor de diferentes modelos, lo cual será útil para obtener predicciones más robustas sobre la criminalidad. Además, se destaca la importancia del aprendizaje en conjunto (ensemble learning), que permite mejorar los modelos individuales mediante la colaboración de varios algoritmos, lo que aumenta la fiabilidad de nuestras predicciones. El uso de SVM stacking en particular, basado en máquinas de soporte vectorial (SVM), se aplicará en nuestro proyecto para optimizar la clasificación de los crímenes en categorías específicas, como homicidios y robos. El artículo también hace énfasis en la utilidad de las técnicas de machine learning y inteligencia artificial (IA) en la predicción de patrones delictivos, lo cual se alinea con el enfoque de nuestro proyecto para desarrollar un sistema predictivo inteligente. Además, se resalta el uso de validación cruzada, una técnica que aplicaremos para asegurar que nuestro modelo sea robusto y preciso antes de ser implementado en el entorno real de Lima Metropolitana.

### 13.1.16 Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention

**Pronóstico del delito: un enfoque de aprendizaje automático y visión artificial para la predicción y prevención del delito**

(Shah et al., 2021)

(DOI: [10.1186/s42492-021-00075-z](https://doi.org/10.1186/s42492-021-00075-z))



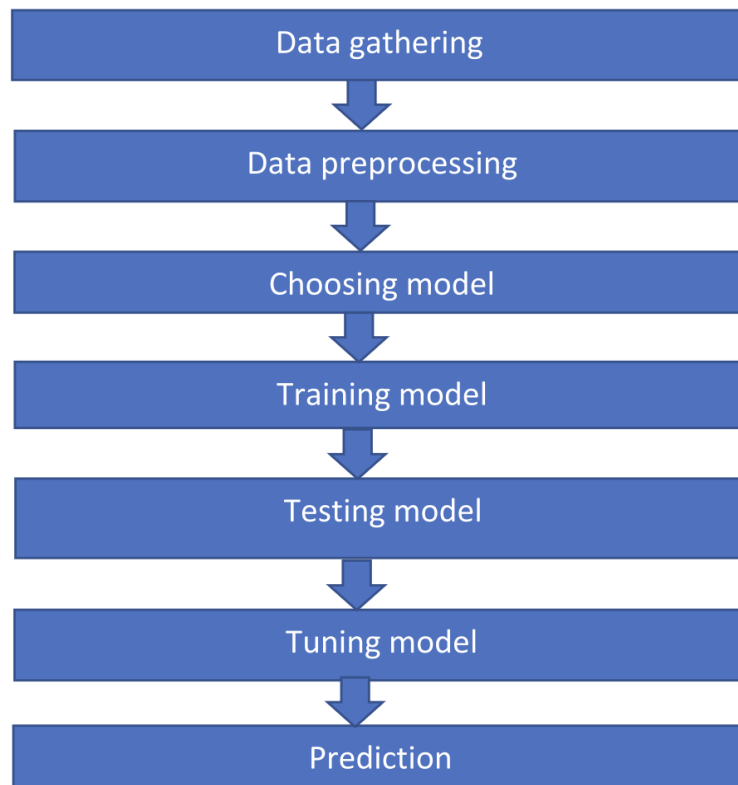
Citado 167 veces en Scopus, Cuartil Q1

Este artículo analiza un enfoque integral para la predicción y prevención del crimen basado en aprendizaje automático (Machine Learning) y visión por computadora, tecnologías emergentes que buscan complementar y superar las limitaciones de los métodos tradicionales de vigilancia policial. Se incluye una exhaustiva comparación de modelos predictivos y la propuesta de un sistema con 17 características técnicas, incluyendo redes neuronales, reconocimiento de voz, análisis de marcha, minería de patrones y simulación de escenarios delictivos. Además, se recopilan resultados empíricos de múltiples estudios y bases de datos de crímenes urbanos. Ante el crecimiento exponencial de la criminalidad y la ineficacia de las técnicas convencionales para resolver crímenes, los autores proponen el diseño de un sistema inteligente que anticipe delitos antes de que ocurran, apoyando a las fuerzas del orden mediante análisis predictivo.

La problemática central abordada es la ineficiencia de las técnicas actuales en detectar patrones delictivos complejos en tiempo real. Para resolverlo, se plantea el desarrollo de un sistema que, utilizando datos históricos, imágenes, grabaciones de audio, análisis de patrones y reconocimiento facial, permite anticipar crímenes con alta precisión, identificar sospechosos y apoyar en decisiones estratégicas. Se destacan ejemplos concretos de implementación de ML y visión por computadora en contextos reales como EE.UU., Canadá e India.

Los autores proponen la construcción de un "oficial policial universal digital", capaz de operar de forma autónoma mediante el aprendizaje profundo, con potencial para revolucionar la vigilancia, reducir la criminalidad y optimizar la asignación de recursos policiales en entornos urbanos. El sistema sería inicialmente probado en áreas metropolitanas para posteriormente escalar.

**Figura 24: Diagrama Dataflow**



## B) Metodología

El artículo se basa en una investigación aplicada y exploratoria, con orientación tecnológica. Emplea el análisis comparativo de técnicas de ML y visión artificial implementadas en contextos reales para diseñar un sistema integral de predicción del delito. También integra elementos de investigación proyectiva, dado que culmina en una propuesta tecnológica concreta.

Los estudios analizados evidencian que la combinación de ML con visión por computadora supera la eficacia de métodos tradicionales. Algunos modelos predictivos alcanzan más del 80% de precisión, especialmente cuando integran fuentes de datos complementarias (clima, transporte, patrones históricos). El modelo DNN propuesto por los autores alcanza un 84.25% de precisión, mientras que modelos más optimizados llegan hasta el 99.16%. Estas cifras demuestran el potencial real de estas tecnologías para la optimización, prevención y asignación estratégica de recursos policiales y gubernamentales.

Los principales pasos de la metodología incluyen:

1. **Diagnóstico del estado actual** de los métodos de vigilancia y predicción del crimen.
2. **Análisis experimental** de modelos predictivos aplicados a distintas ciudades (Chicago, Vancouver, Filadelfia, entre otras).

3. **Diseño conceptual del sistema propuesto**, incluyendo arquitectura modular y funcionamiento autónomo.
4. **Propuesta de prueba piloto en entornos urbanos controlados**, con simulación de escenarios criminales y evaluación de amenazas.

**Figura 25: Functionality of proposed approach**



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

Gracias a los aportes de este artículo, es de gran utilidad para nuestra tesis propuesta, es por ello que implementaremos un módulo de predicción de criminalidad utilizando algoritmos de Machine Learning y visión por computadora que procesará datos históricos de delitos en Lima Metropolitana. Este módulo aplicará algoritmos de aprendizaje automático, como Redes Neuronales Convolucionales (CNN) y Redes Neuronales Recurrentes (RNN), para identificar patrones espaciales y temporales en los delitos, tal como se sugiere en el artículo. Además, desarrollaremos un algoritmo de predicción basado en un modelo de regresión logística y redes neuronales profundas (DNN) para predecir la ocurrencia de crímenes en áreas específicas de Lima, mejorando la precisión del sistema predictivo. Los reportes generados por el sistema incluirán mapas de calor y visualizaciones interactivas de zonas de alto riesgo, lo que permitirá a las autoridades visualizar las áreas con mayor probabilidad de ocurrencia de crímenes y asignar recursos policiales de manera más eficiente. Este enfoque optimizará la prevención del crimen en Lima Metropolitana y fortalecerá la toma de decisiones estratégicas de las fuerzas de seguridad.

En el artículo, se menciona que la predicción y prevención del crimen se logra a través de un enfoque que integra aprendizaje automático (Machine Learning) y visión artificial para anticipar comportamientos delictivos. Además, se incluye una comparación de modelos predictivos, lo cual permite identificar cuáles ofrecen mayor precisión en la identificación de patrones criminales. Este análisis comparativo refuerza la necesidad de evaluar diferentes algoritmos en nuestro proyecto con el fin de seleccionar aquellos que ofrezcan mejores resultados en contextos urbanos como Lima. Asimismo, el artículo destaca que es posible anticipar crímenes con alta precisión mediante el análisis de datos históricos, imágenes y otras variables contextuales, una capacidad que resulta esencial para nuestra tesis, ya que el aplicativo que se propone busca proporcionar alertas predictivas confiables para una mejor

toma de decisiones estratégicas por parte de las autoridades policiales. Esta evidencia respalda técnicamente el valor de nuestra propuesta y demuestra el potencial de las tecnologías emergentes para transformar los sistemas de seguridad pública en contextos latinoamericanos.

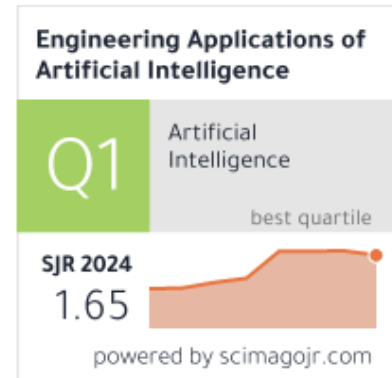
### 3.1.17 Prediction of crime rate in urban neighborhoods based on machine learning

#### Predicción de la tasa de criminalidad en barrios urbanos basada en aprendizaje automático

(He & Zheng, 2021)

(DOI: [10.1016/j.engappai.2021.104460](https://doi.org/10.1016/j.engappai.2021.104460))

Citado 60 veces en Scopus, Cuartil Q1

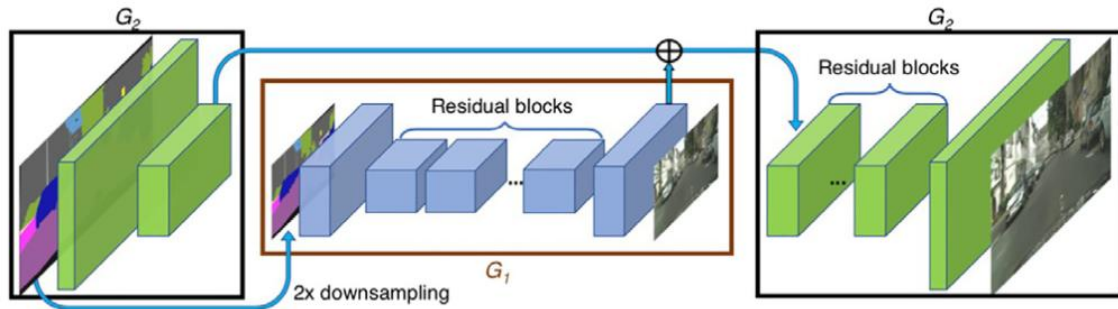


Este artículo propone un enfoque innovador para predecir la tasa de criminalidad en vecindarios urbanos utilizando redes neuronales generativas adversarias (GAN). El objetivo principal es visualizar y anticipar con precisión la distribución del crimen en diversas áreas urbanas, a fin de mejorar la planificación urbana y la seguridad ciudadana mediante herramientas de inteligencia artificial. La principal contribución del artículo radica en presentar una metodología eficaz y visual para la prevención del delito, facilitando la toma de decisiones en el diseño urbano mediante la predicción automatizada de zonas críticas

La problemática central que aborda el estudio es la limitación de los modelos tradicionales de predicción criminal, los cuales suelen apoyarse exclusivamente en datos históricos de delitos sin considerar los factores espaciales ni las características ambientales del entorno urbano. Esta falta de integración de datos contextuales reduce significativamente la precisión de las predicciones. Asimismo, los modelos como regresiones múltiples, SVM o redes neuronales tradicionales no son óptimos para representar de forma visual el impacto espacial de estos fenómenos.

Los resultados muestran que el modelo es capaz de predecir de forma precisa las zonas de alta criminalidad, identificando patrones recurrentes relacionados con factores como la densidad de edificaciones, disposición de calles, parques y edificios municipales. Además, se realizaron pruebas con mapas de ciudades como Princeton, Seattle y Nueva York, verificando su aplicabilidad y utilidad para mejorar la seguridad urbana en diferentes contextos.

**Figura 26: The basic principles of GAN**



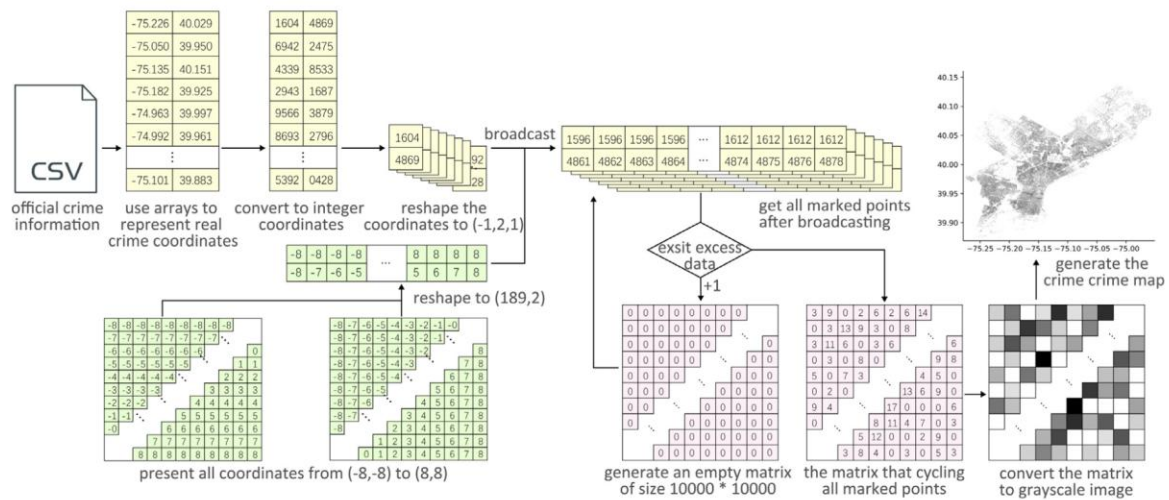
## B) Metodología

El estudio se basa en un enfoque cuantitativo de predicción mediante aprendizaje profundo, específicamente utilizando Generative Adversarial Networks (GAN), una técnica avanzada de machine learning para el aprendizaje de patrones complejos entre mapas urbanos (input) y mapas de calor del crimen (output).

Los principales pasos de la metodología incluyen:

1. **Recopilación de datos:** Se generaron mapas de calor a partir de coordenadas de delitos, usando Python y librerías como NumPy, Pandas y Matplotlib.
2. **Procesamiento de mapas urbanos:** Se estandarizó la escala visual y la codificación de colores para mejorar la capacidad de aprendizaje del modelo.
3. **Entrenamiento del modelo GAN:** Se usó el modelo Pix2pixHD, que permite transformar mapas urbanos en mapas de predicción del crimen.
4. **Pruebas de validación:** Se evaluó la capacidad del modelo para predecir patrones de criminalidad en áreas no incluidas en el entrenamiento, destacando su generalización.

**Figura 2: The principle of generating the crime heat map.**



## Utilidad del artículo para el proyecto de tesis

Gracias a los aportes de este artículo, es de gran utilidad para nuestra tesis propuesta, ya que desarrollaremos un módulo de predicción de tasas de criminalidad a nivel de vecindario que aplicará los modelos de Machine Learning, específicamente redes neuronales como las GAN, para analizar datos históricos de delitos en los distintos distritos de Lima Metropolitana. Como otra aportación de gran utilidad diseñaremos un módulo de visualización cartográfica interactiva que generará mapas de calor de distribución de criminalidad y planos urbanos con superposiciones de riesgo , para Lima Metropolitana. Esto proporcionará a la Policía Nacional del Perú una herramienta visual y predictiva robusta para la planificación de la seguridad y la asignación estratégica de recursos en zonas específicas, mejorando la prevención y la intervención policial.

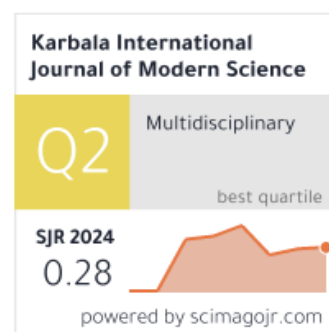
En el artículo, se menciona que las redes neuronales generativas adversarias (GAN) fueron utilizadas para establecer una relación directa entre los mapas urbanos y los patrones de criminalidad, lo cual permitió una visualización detallada del riesgo delictivo en distintas zonas. Esta técnica resulta esencial para nuestro proyecto, ya que su implementación posibilitará la prevención del delito mediante la predicción automatizada de zonas críticas, incluso en lugares donde no se cuenta con datos históricos suficientes. Además, el estudio demuestra que es posible predecir de forma precisa las zonas de alta criminalidad, lo cual respalda nuestra propuesta de desarrollar un sistema predictivo capaz de identificar patrones delictivos y anticipar focos de inseguridad. La integración de esta tecnología permitirá transformar datos urbanos en mapas de riesgo que serán de gran valor operativo para las unidades policiales en Lima Metropolitana.

### 3.1.18 Machine Learning based Soft Computing Regression Analysis Approach for Crime Data Prediction

**Enfoque de análisis de regresión de computación blanda basado en aprendizaje automático para la predicción de datos sobre delincuencia**

(Aziz et al., 2022)

(DOI: [10.33640/2405-609X.3197](https://doi.org/10.33640/2405-609X.3197))



**Citado 47 veces en Scopus , Cuartil Q1**

El artículo presenta un enfoque basado en técnicas de [aprendizaje automático \(machine learning\)](#) para predecir datos delictivos en India mediante análisis de regresión de soft computing. Dada la creciente tasa de criminalidad en India y el volumen acumulado de datos, se hace imperativo utilizar métodos predictivos que ayuden a las autoridades a anticipar incidentes y [asignar recursos de manera más eficiente](#).

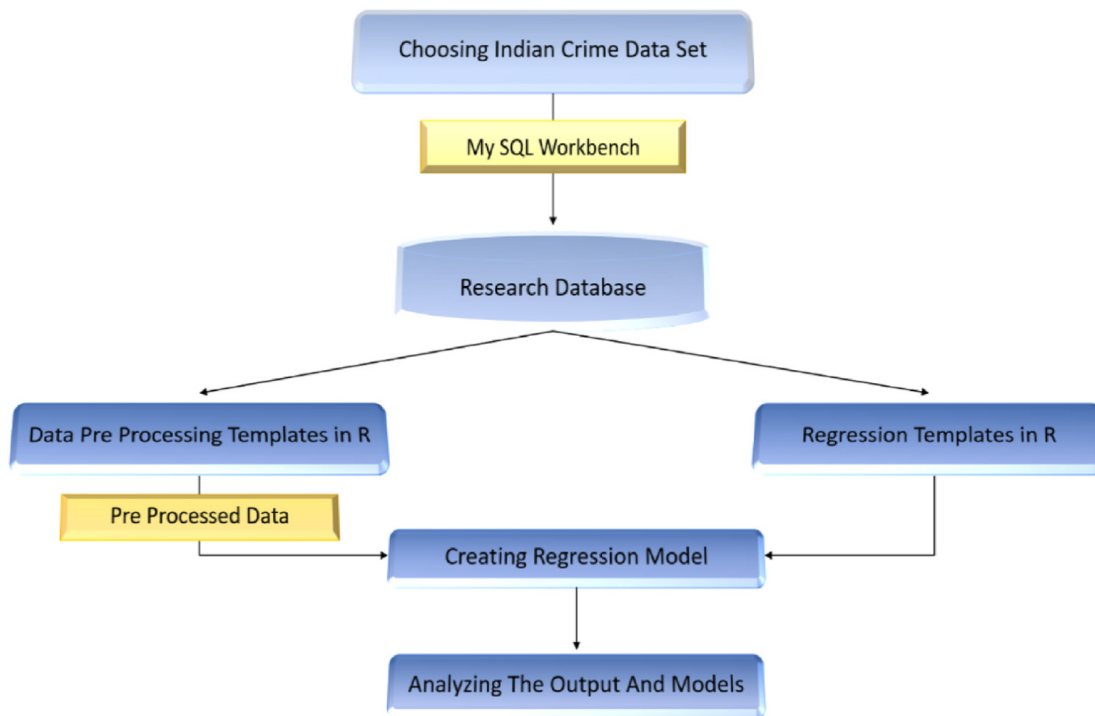
La problemática central es la dificultad para predecir crímenes debido a la complejidad de los factores sociales y económicos involucrados, así como la necesidad de aprovechar grandes volúmenes de datos delictivos. El objetivo del estudio es [desarrollar modelos de predicción que puedan estimar, por año y región, los conteos de crímenes registrados bajo el Código Penal Indio \(IPC\), tanto en total como desglosados por tipo de delito \(asesinato, violación, secuestro, disturbios, etc.\)](#).

Para ello, se propone un [enfoque computacional](#) que emplea cinco algoritmos de regresión distintos: Regresión Lineal Simple (SLR), Regresión Lineal Múltiple (MLR), Regresión de Árbol de Decisión (DTR), Regresión de Soporte Vectorial (SVR) y Regresión de Bosque Aleatorio (RFR). Estos modelos fueron entrenados y validados usando datos espaciales-temporales de distritos indios, entre 2001 y 2012, extraídos del portal del NCRB (Oficina Nacional de Registro de Crímenes).

Este trabajo ofrece una base estructurada para desarrollar [modelos predictivos en análisis de crimen](#) en India y puede ser aprovechado por otros investigadores y fuerzas de seguridad para mejorar sus procesos de prevención y planificación.



**Figura 27: The flow chart of proposed ICDA Approach.**



## **B) Metodología**

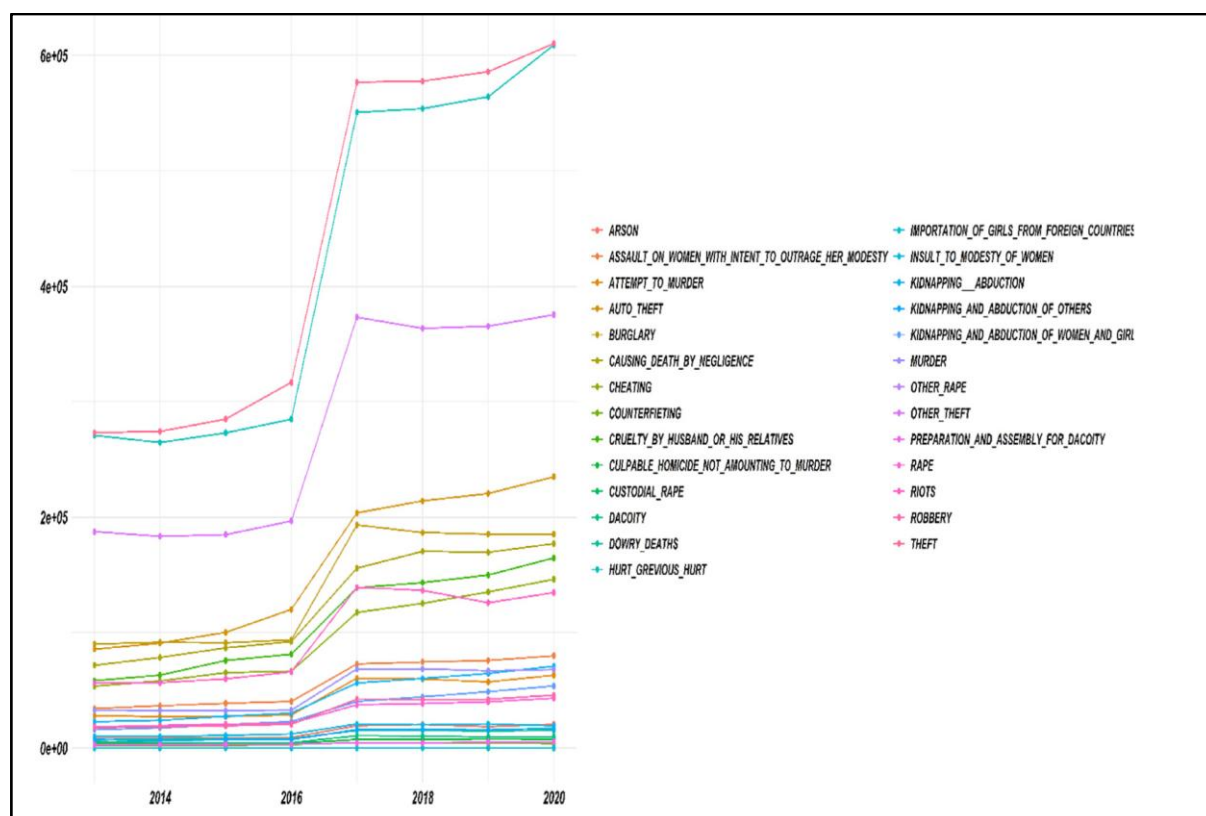
El estudio se basa en una investigación aplicada, utilizando un enfoque cuantitativo de aprendizaje automático con componentes de soft computing. Se emplea un método experimental, comparando múltiples modelos de regresión para encontrar el mejor predictor del comportamiento criminal en diferentes regiones y años.

Los principales pasos de la metodología incluyen:

1. **Recopilación de datos:** Incluyen 28 tipos distintos de crímenes IPC, otros crímenes IPC y el total anual por región.
2. **Preprocesamiento:** Transformación de datos usando R y MySQL para crear subconjuntos derivados y procesados.
3. **Entrenamiento del modelo GAN:** Se usaron cinco modelos de regresión: SLR, MLR, DTR, SVR y RFR.

El estudio no se divide explícitamente en ciclos de intervención, pero sí se estructura en etapas secuenciales: preparación del conjunto de datos, creación de modelos de predicción, evaluación del rendimiento de los modelos, y visualización de los resultados en mapas (Leaflet en R).

**Figura 28: Crimes evolution per type of crime 2014 - 2020.**



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

Gracias a los aportes de este artículo, es de gran utilidad para nuestra tesis propuesta, es por ello que **desarrollaremos un módulo de predicción de criminalidad** utilizando técnicas de regresión que procesa datos históricos de delitos en Lima Metropolitana. Este módulo empleará algoritmos de regresión como Random Forest Regression (RFR), Support Vector Regression (SVR), Decision Tree Regression (DTR), y Multiple Linear Regression (MLR), tal como se describe en el artículo, para predecir el número de crímenes en diferentes áreas geográficas. Otra aportación de gran utilidad es que **diseñaremos un módulo de predicción espacial-temporal** que utilizará los datos de crímenes de Lima para predecir las zonas y momentos de mayor probabilidad de ocurrencia de delitos, basándose en patrones históricos y factores espaciales y temporales. Además, **desarrollaremos un módulo de preprocesamiento de datos**, que limpiará y transformará los datos históricos para mejorar la precisión de los modelos predictivos, asegurando que el sistema sea lo más preciso posible para la prevención del crimen en Lima Metropolitana.

En el artículo, se menciona que el uso del aprendizaje automático (machine learning) permite extraer conocimiento útil de grandes volúmenes de datos criminales, lo cual es clave para el desarrollo de sistemas predictivos. Asimismo, se destaca la importancia de asignar recursos de manera más eficiente, objetivo que compartimos en nuestra tesis al buscar optimizar el despliegue policial según el nivel de riesgo por distrito. Además, se propone un enfoque computacional centrado en el uso de algoritmos de regresión que se entrenan con datos espaciales-temporales, lo que orienta nuestro proceso técnico en el desarrollo del aplicativo. Finalmente, el artículo establece las bases para la construcción de modelos predictivos en análisis de crimen, los cuales adaptamos al contexto de Lima Metropolitana para generar una herramienta tecnológica útil para la gestión y prevención del delito.

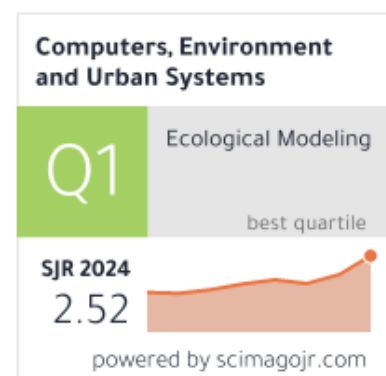
**De acuerdo a la figura 28**, que ilustra la evolución de los crímenes por tipo de delito, se indica **cómo los patrones históricos de criminalidad pueden ser utilizados para predecir la ocurrencia futura de delitos en distintas áreas geográficas y por tipo de crimen**. En este sentido, se tomará el enfoque de regresión con modelos como aprendizaje automático y modelos predictivos en análisis de crimen, tal como se describe en el artículo, para predecir el número de crímenes por tipo y zona. Este enfoque permitirá generar visualizaciones interactivas, como mapas de calor y gráficos de tendencia, que ayudarán a las autoridades a identificar áreas de alto riesgo y asignar recursos de manera eficiente en Lima Metropolitana.

### 3.1.19 A data driven agent based simulation to predict crime patterns in an urban environment

Una simulación basada en agentes de datos para predecir patrones delictivos en un entorno urbano

(Rosés et al., 2021)

DOI: [10.1016/j.compenvurbsys.2021.101660](https://doi.org/10.1016/j.compenvurbsys.2021.101660)



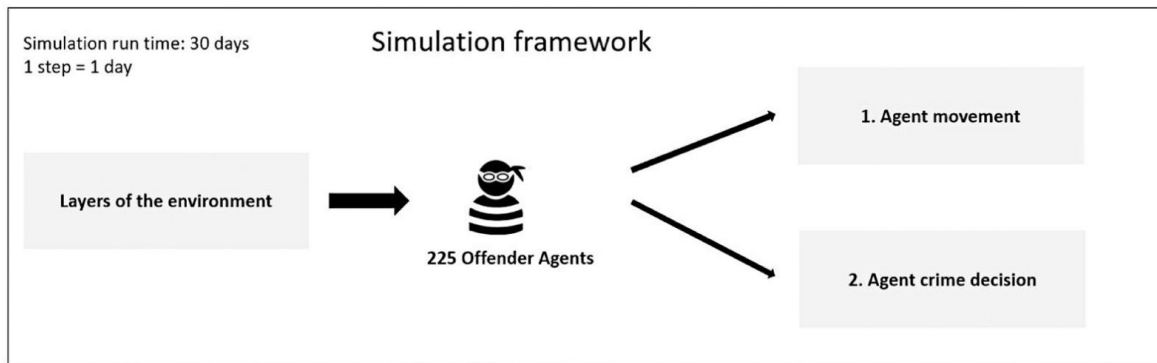
Citado 42 veces en Scopus , Cuartil Q1

La presente investigación se enmarca en el campo de la criminología computacional y la modelación urbana, con un enfoque centrado en la predicción de patrones delictivos en entornos urbanos. La problemática central que aborda el artículo es la dificultad de representar y predecir el crimen urbano de forma realista mediante modelos teóricos o puramente estadísticos, lo cual limita su aplicabilidad práctica en políticas de prevención. Frente a esto, el estudio propone una simulación basada en agentes (ABM) impulsada por datos abiertos y técnicas de aprendizaje automático (Machine Learning), con el objetivo de predecir patrones de robo a nivel de segmento de calle en la ciudad de Nueva York. El modelo simula el comportamiento de agentes virtuales que representan a potenciales delincuentes navegando por una red urbana construida a partir de datos abiertos.

El artículo tiene como propósito principal construir un modelo predictivo de patrones criminales que sea al mismo tiempo teóricamente informado y empíricamente fundamentado. Busca responder si un modelo de simulación impulsado por datos reales puede generar patrones delictivos similares a los observados en la realidad, y qué capas de información (espacial, temporal, interactiva) contribuyen de manera más significativa a su capacidad predictiva.

El modelo demuestra una alta capacidad predictiva a nivel de micro-espacios urbanos, especialmente al combinar múltiples fuentes de datos. Los resultados indican que los datos espaciales son el componente más influyente en la predicción, aunque las capas temporales y de interacción también aportan mejoras significativas. El trabajo establece una base sólida para el uso práctico de simulaciones impulsadas por datos en estrategias de prevención del delito, y representa un avance relevante en la integración de inteligencia artificial, datos abiertos y teoría criminológica en simulaciones urbanas.

**Figura 29: Model framework**



## B) Metodología

La investigación adopta un enfoque cuantitativo, computacional y aplicado, centrado en la simulación basada en agentes (ABM). La metodología combina elementos de la modelación urbana, análisis geoespacial y aprendizaje automático, integrando tanto fundamentos teóricos de la criminología ambiental como datos empíricos en múltiples dimensiones.

El entorno urbano virtual fue construido a partir de datos abiertos de la ciudad de Nueva York. Entre las fuentes de información se incluyen:

- Registros de robos del NYPD (junio 2014 - junio 2015),
- Llamadas al 311 como proxy de desorden físico y social,
- Datos de movilidad urbana (trips de taxi),
- Check-ins de Foursquare como indicador de actividad humana,
- Condiciones meteorológicas diarias,
- Datos censales y geográficos: red vial, árboles urbanos, escuelas, transporte público, uso del suelo, densidad poblacional, etc.

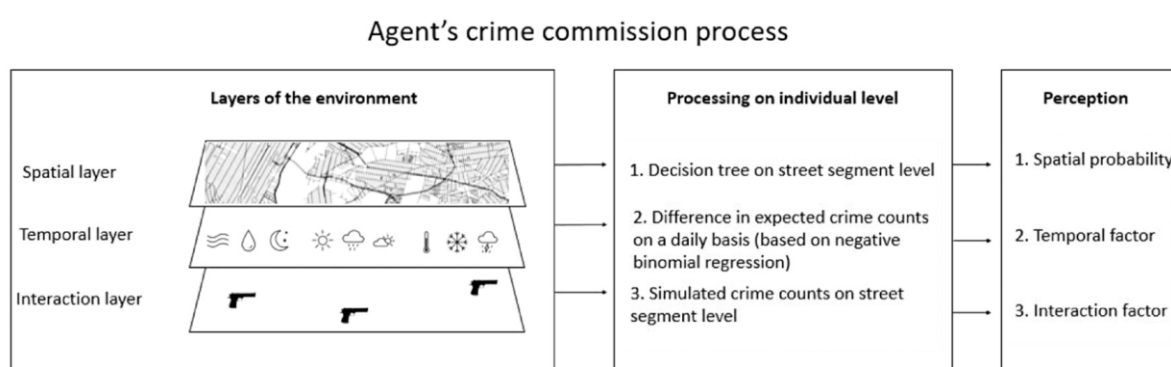
Los agentes simulan comportamientos cotidianos (movimientos de hogar a nodos de actividad), y toman decisiones sobre si cometer un delito evaluando tres capas del entorno:

1. **Capa espacial (S):** Evaluada mediante un árbol de decisión, entrenado con datos históricos a nivel de segmento de calle.
2. **Capa temporal (T):** Ajusta la probabilidad de cometer un delito en función de las condiciones climáticas y el día de la semana, usando regresión binomial negativa.

3. **Capa de interacción (I):** Incorpora el efecto de repetición criminal a corto plazo en segmentos donde ya ocurrió un delito (near-repeat).

Cada escenario fue calibrado usando datos de mayo 2015 y probado sobre los datos de junio 2015. La evaluación del modelo se realizó a través de métricas estándar en criminología computacional: índice de precisión predictiva (PAI), error cuadrático medio (RMSE) y curva ROC (AUC-ROC).

Figura 2: Process of agent deciding whether to commit crimes in three layers containing the layers of the environment, how individual agents processed these layers. and what individual agents perceive.



### Utilidad del artículo para el proyecto de tesis

Gracias a los aportes de este artículo, es de gran utilidad para nuestra tesis propuesta, es por ello que desarrollaremos un módulo de predicción de criminalidad basado en un modelo de simulación de agentes que procesa datos históricos de delitos en Lima Metropolitana. Este módulo aplicará técnicas de simulación basadas en agentes (ABM), en las que los agentes virtuales tomarán decisiones sobre si cometer o no delitos, basándose en las características de su entorno, tal como se propone en el artículo. Otra aportación de gran utilidad es que diseñaremos un módulo de reportes que combinará datos geoespaciales con características temporales, como el clima y la actividad humana, para mejorar la precisión del modelo predictivo y optimizar la asignación de recursos policiales en Lima Metropolitana.

En el artículo, se menciona que el uso del aprendizaje automático (machine learning) permite a los agentes virtuales procesar información del entorno para decidir si cometieron o no un delito, lo cual es crucial para la generación de predicciones realistas.

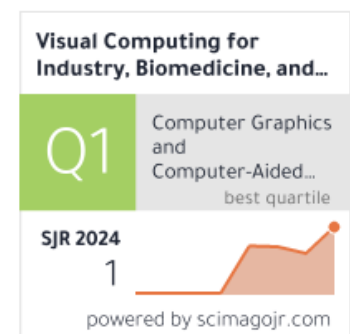
Además, se destaca el valor de la simulación basada en agentes (ABM) como una herramienta eficaz para representar el comportamiento dinámico de los delincuentes en escenarios urbanos complejos. Esta metodología permite modelar interacciones entre individuos y su entorno, aportando un nivel de realismo que mejora la comprensión de los fenómenos delictivos. Finalmente, el estudio enfatiza la importancia de la predicción de patrones delictivos en entornos urbanos, ya que permite anticipar zonas de riesgo y diseñar estrategias preventivas más efectivas, lo cual se alinea directamente con los objetivos de nuestro proyecto en Lima Metropolitana.

### 3.1.20 Predicting offenses among individuals with psychiatric disorders - A machine learning approach

#### Predicción de delitos en personas con trastornos psiquiátricos: un enfoque de aprendizaje automático

(Watts et al., 2021)

(DOI: [10.1016/j.jpsychires.2021.03.026](https://doi.org/10.1016/j.jpsychires.2021.03.026))



Citado 31 veces en Scopus, Cuartil Q1

El artículo presenta una propuesta innovadora para mejorar la predicción del comportamiento delictivo en personas con trastornos psiquiátricos. El enfoque tradicional en psiquiatría forense se ha basado en herramientas actuariales que, si bien son útiles para estimar el riesgo general de reincidencia dentro de poblaciones, resultan insuficientes cuando se busca realizar predicciones específicas a nivel individual. Estas herramientas priorizan factores estáticos como antecedentes penales o experiencias infantiles adversas, lo cual puede conducir a sesgos, estigmatización y una limitada capacidad de intervención terapéutica eficaz.

Con el propósito de superar estas limitaciones, el estudio emplea modelos de aprendizaje automático (machine learning) para predecir el tipo de delito (violento, no violento

o sexual) que una persona con enfermedad mental podría cometer, utilizando un conjunto de datos clínicos y sociodemográficos recopilados de 1240 pacientes forenses. Todos los participantes habían sido declarados legalmente No Criminalmente Responsables (NCR) o No Aptos para Juicio (UST) debido a su condición mental, y provenían de diez instituciones psiquiátricas en la provincia de Ontario, Canadá.

Los resultados obtenidos muestran que los modelos fueron capaces de realizar predicciones individualizadas con una precisión superior a las herramientas convencionales. Por ejemplo, los modelos lograron diferenciar con altos niveles de sensibilidad y especificidad los delitos sexuales de los violentos y no violentos, alcanzando valores de precisión balanceada superiores al 80% en algunos casos. Entre los predictores más significativos se encuentran el diagnóstico de parafilia, los trastornos del control de impulsos, la falta de ingresos económicos, antecedentes de abuso de sustancias y la presencia de deterioro cognitivo.

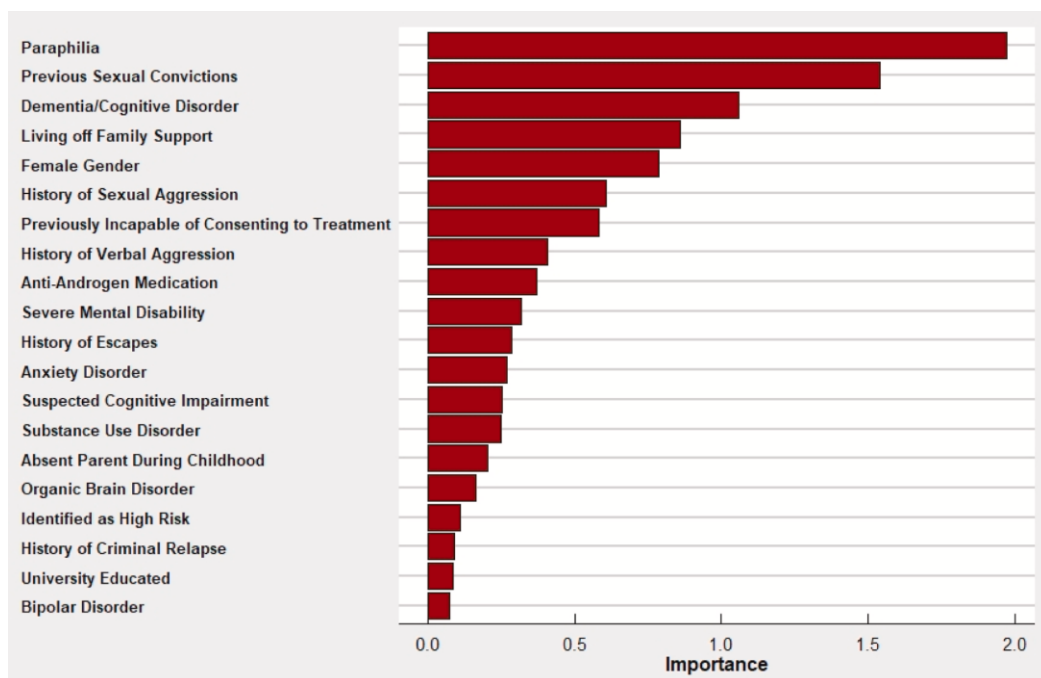


Figura 30: Variable importance plot of sexual vs violent offenses

## B) Metodología

El estudio empleó un diseño cuantitativo y transversal, con un enfoque basado en ciencia de datos, orientado a la aplicación de técnicas de aprendizaje automático para la predicción del tipo de delito en individuos con diagnóstico psiquiátrico. La muestra estuvo

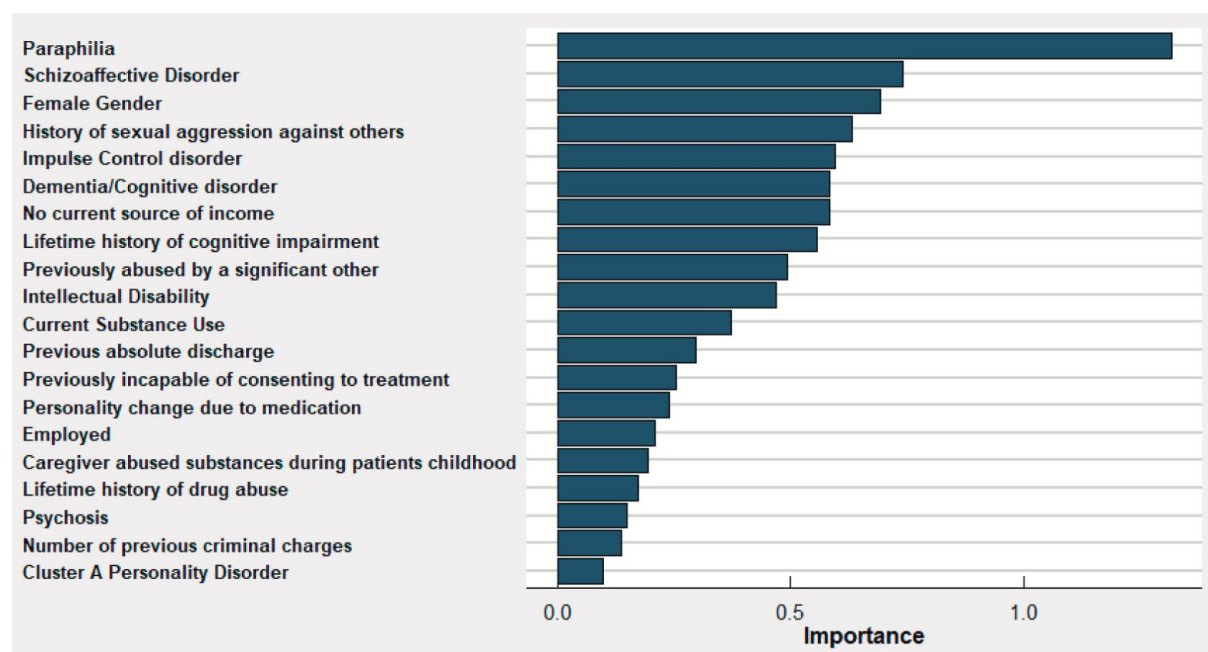


compuesta por 1240 pacientes forenses que fueron declarados No Criminalmente Responsables (NCR) o No Aptos para Juicio (UST) entre los años 2014 y 2015, bajo la jurisdicción del Ontario Review Board en Canadá. Los participantes provenían de diez instituciones psiquiátricas y fueron categorizados en tres grupos en función del delito más grave cometido: delitos violentos, no violentos o sexuales.

La base de datos inicial incluyó 246 variables clínicas, sociodemográficas, históricas y de factores de riesgo conocidos. Tras aplicar un umbral de tolerancia al porcentaje de datos faltantes ( $\leq 15\%$ ), se conservaron 156 variables que fueron sometidas a procesos de codificación binaria (one-hot encoding) para su análisis computacional. El objetivo fue predecir, mediante modelos supervisados, el tipo de delito cometido por cada individuo, utilizando únicamente información registrada antes del hecho delictivo.

- **Datos y variables:** Se partió de 246 variables clínicas, demográficas e históricas; se seleccionaron 156 con menos de 15% de datos faltantes.
- **Modelado y análisis:** Se entrenaron tres algoritmos: **Elastic Net**, **Random Forest** y **Support Vector Machine (SVM)**.
- **Desbalance de clases:** Se emplearon estrategias como submuestreo, sobremuestreo y ponderación de clases para equilibrar las categorías de delitos.

**Figura 31: Variable importance plot of sexual vs nonviolent offenses**



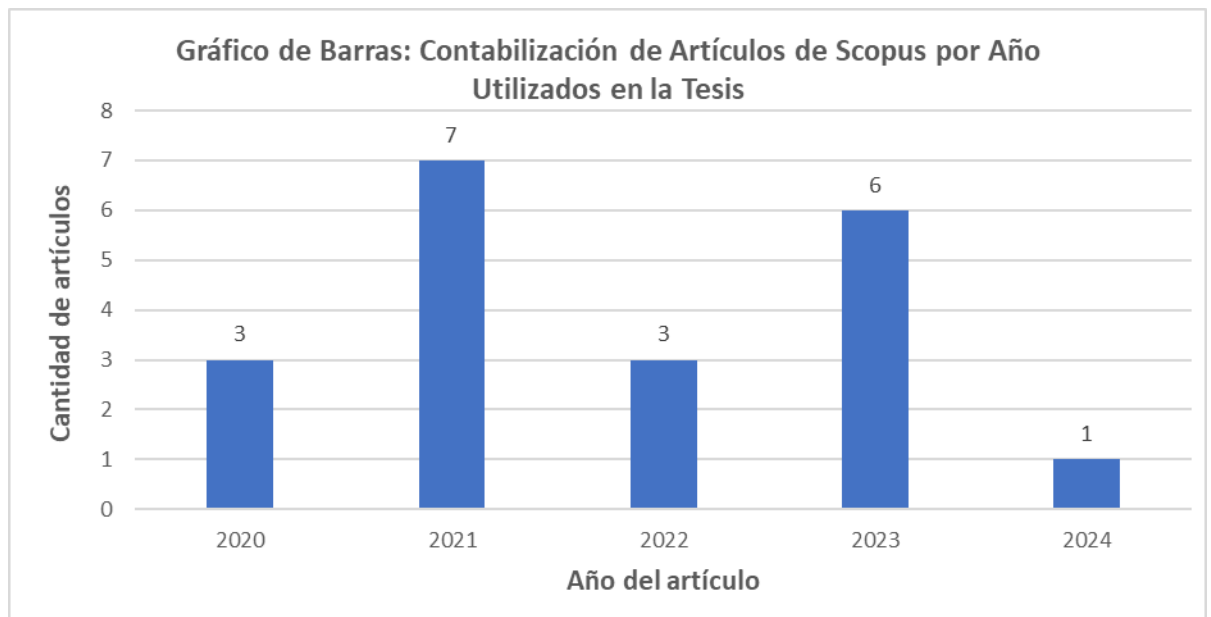
### **Utilidad del artículo para el proyecto de tesis**

Gracias a los aportes de este artículo, que son de gran utilidad para nuestra tesis propuesta, ya que desarrollaremos un módulo de clasificación de delitos utilizando algoritmos de aprendizaje automático como Random Forest, Elastic Net y SVM. Estos modelos permitirán predecir el tipo de delito cometido a nivel individual, basándose en variables clínicas y demográficas, como trastornos de control de impulsos, abuso de sustancias y antecedentes penales. Otra aportación de gran utilidad es que diseñaremos un módulo de selección de características utilizando técnicas como la eliminación recursiva de características (RFE) y la selección de características por importancia, lo que mejorará la precisión y la interpretabilidad de los modelos predictivos. Este enfoque basado en datos históricos de delitos y características delictivas facilitará una predicción más precisa, permitiendo un modelo más personalizado para la toma de decisiones en la gestión de la seguridad pública.

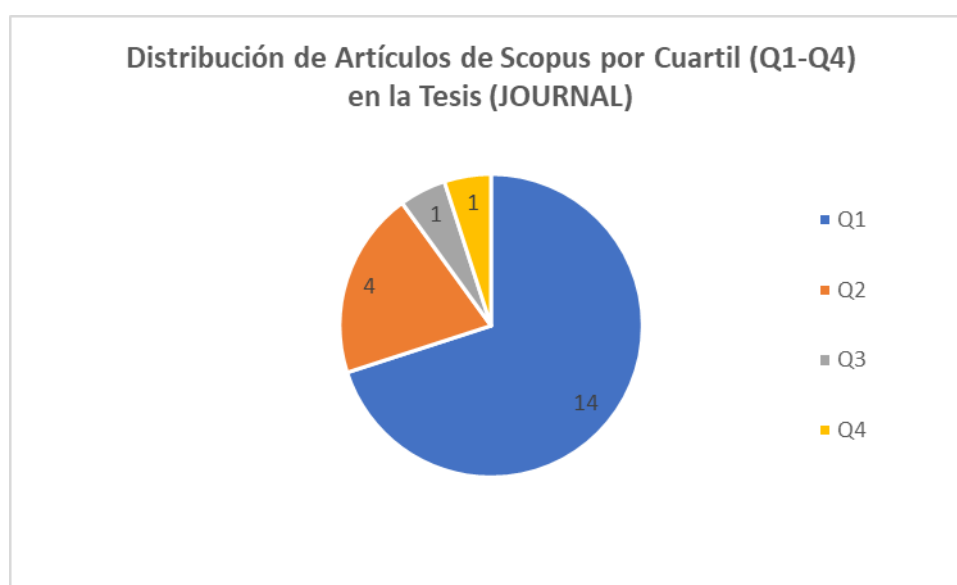
En el artículo, se menciona que la predicción del comportamiento delictivo puede ser significativamente más precisa mediante el uso de técnicas modernas como el aprendizaje automático (machine learning), superando las limitaciones de los métodos tradicionales que priorizan factores estáticos como antecedentes penales. Esta observación es especialmente pertinente para nuestra investigación, ya que también buscamos dejar atrás las estimaciones generales y adoptar un enfoque más dinámico y personalizado. Al incorporar modelos de machine learning entrenados con datos históricos delictivos de Lima, se generará predicciones más específicas y actualizadas, ajustadas a la realidad urbana de nuestra ciudad. Esto permitirá anticipar patrones de criminalidad y, en consecuencia, optimizar el despliegue de efectivos policiales en los sectores más vulnerables.

### 3.2 Análisis de la Distribución de los Artículos en la Tesis: Año y Cuartil

#### 3.2.1 Gráfico de Barras: Distribución de Artículos por Año



### 3.2.2 Gráfico Circular: Distribución de Artículos por Cuartil



## CAPÍTULO IV : MODELADO DE NEGOCIO

### 4.1 Reglas del Negocio

El modelo de negocio que nosotros proponemos consiste en la interacción y conexión entre los recursos policiales y el sistema predictivo de criminalidad a través de una plataforma basada en Machine Learning. Esta plataforma permitirá optimizar la asignación de patrulleros y personal policial a zonas con mayor probabilidad delictiva, a través de la predicción de delitos en función de factores como la ubicación, tipo de delito y hora del día.

Estas reglas del negocio deben ser elaboradas conforme a las normativas y protocolos establecidos por las autoridades competentes en seguridad pública, garantizando su alineación con el plan de seguridad del Estado y las directrices específicas de la policía nacional.

#### **RN01 - Gestión de Datos Históricos**

- Los datos históricos de criminalidad deben tener una antigüedad mínima de 2 años para garantizar la precisión del modelo predictivo.

- Los datos deben incluir: fecha, hora, ubicación (coordenadas), tipo de delito y distrito.
- La información debe ser actualizada semanalmente desde las bases de datos de denuncias policiales.

## **RN02 - Algoritmos de Machine Learning**

- El sistema debe utilizar al menos dos algoritmos de machine learning para comparar resultados (ej: Random Forest y Regresión Logística).
- La precisión del modelo predictivo debe ser superior al 75% para ser considerado confiable.
- El modelo debe ser re entrenado mensualmente con nuevos datos.

## **RN03 - Asignación de Recursos Vehiculares**

- Solo se pueden asignar vehículos que estén operativos y disponibles.
- La asignación debe priorizar zonas con probabilidad delictiva superior al 60%.
- Un vehículo patrullero no puede ser asignado a más de 3 zonas críticas por turno.

## **RN04 - Generación de Rutas de Patrullaje**

- Las rutas deben optimizar tiempo y distancia considerando el tráfico vehicular.
- Cada ruta no debe exceder las 8 horas de patrullaje continuo.
- Las rutas deben incluir puntos de control obligatorios cada 2 horas.

## **RN05 - Autorización y Supervisión**

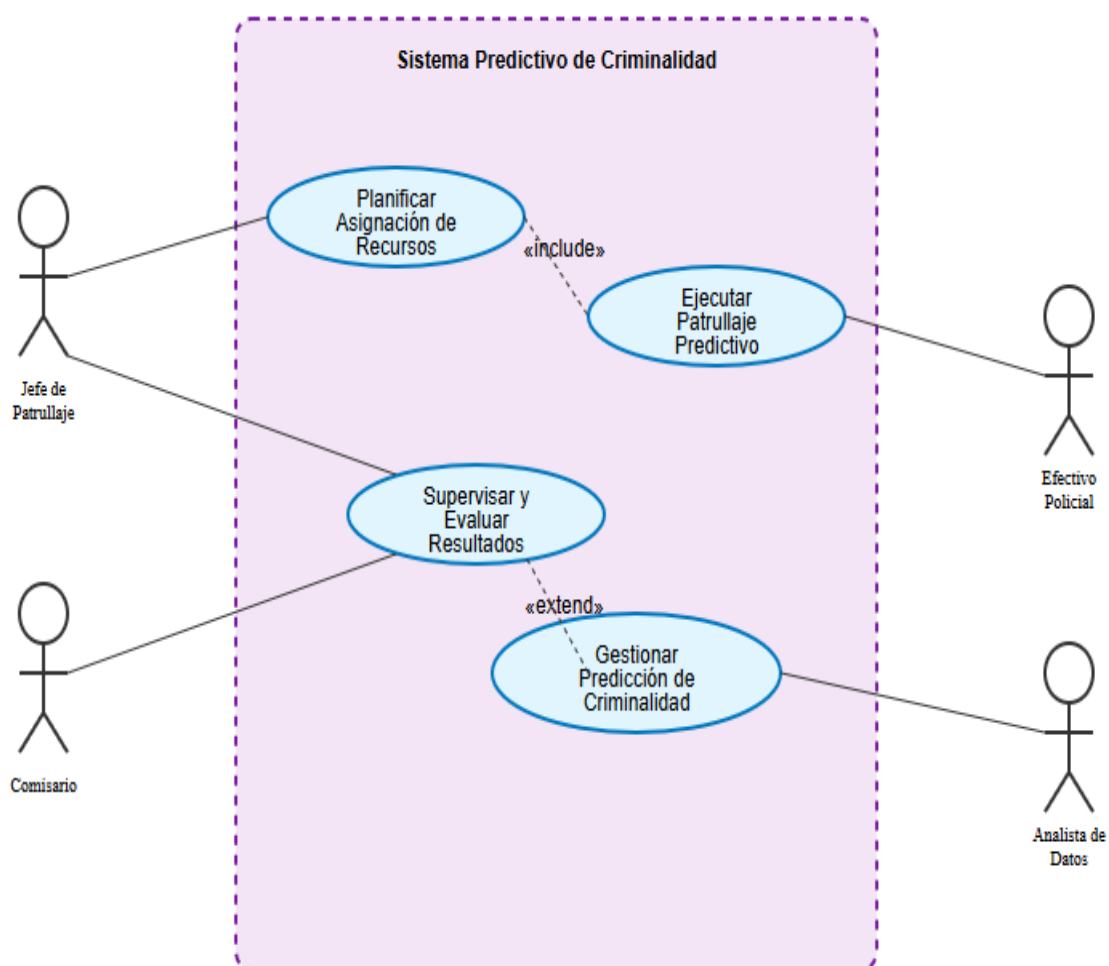
- Solo el jefe de patrullaje puede autorizar cambios en las rutas asignadas.
- Los resultados de patrullaje deben ser reportados al comisario dentro de las 24 horas.
- Las recomendaciones del sistema deben ser validadas por personal con rango mínimo de suboficial.

## **RN06 - Seguridad de la Información**

- Los datos predictivos son clasificados como información sensible.
- El acceso al sistema requiere autenticación de doble factor.
- Los logs de actividad deben conservarse por un período mínimo de 1 año.

## 4.2 Caso de uso de negocio

### 4.2.1 Diagrama de caso uso de negocio



**Figura 1.3.** Diagrama de caso de Uso

### 4.2.2 Actores del negocio

Los trabajadores del negocio en el uso de la plataforma son los siguientes:

**Actor Principal:**

- Jefe de Patrullaje: Responsable de la planificación, asignación y supervisión de recursos vehiculares para patrullaje preventivo.

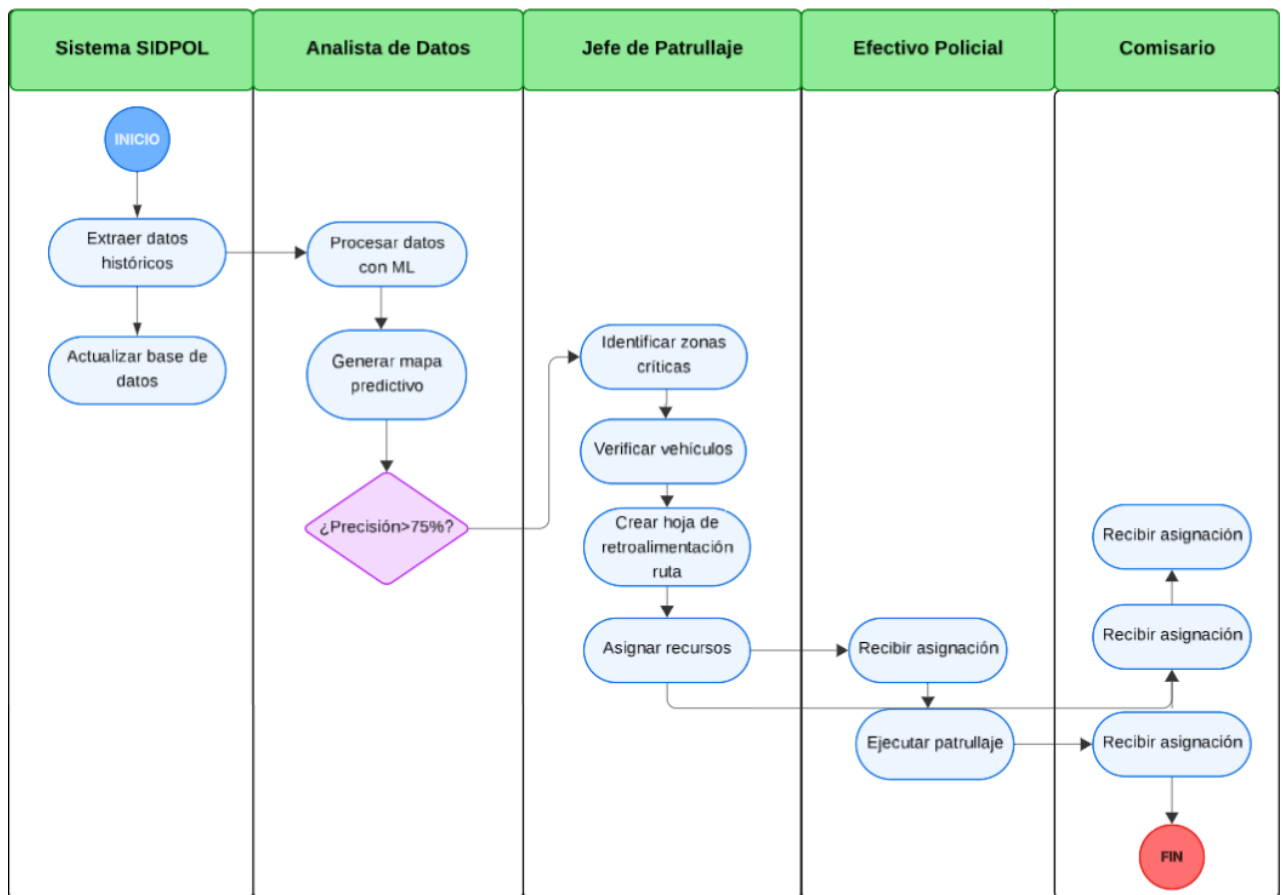
**Actores Secundarios:**

- Comisario: Supervisor general que recibe informes y valida estrategias de patrullaje.
- Efectivo Policial: Ejecuta las rutas de patrullaje asignadas y reporta incidencias.
- Analista de Datos: Mantiene y actualiza los modelos predictivos del sistema.
- Sistema SIDPOL: Sistema externo que proporciona datos de denuncias y estadísticas delictivas.

#### **4.2.3 Diagrama de actividades del negocio**

**Diagrama de Actividades del Negocio**

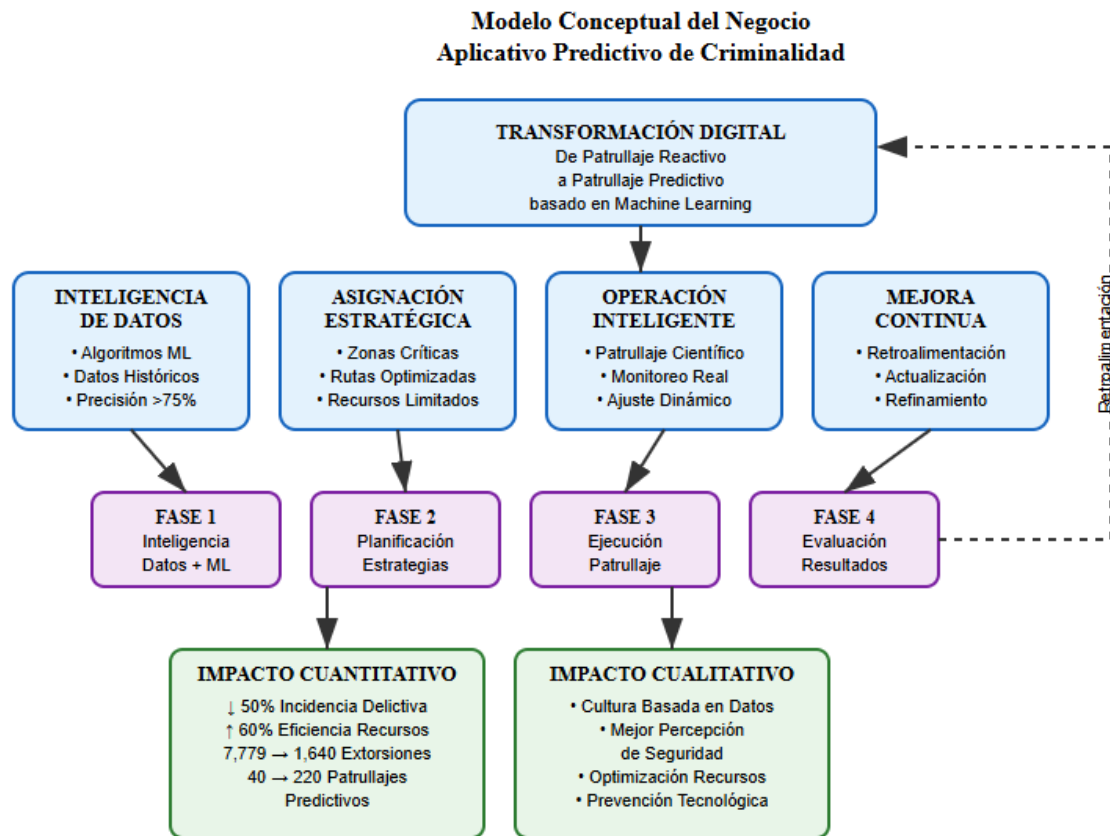
**Asignación Predictiva de Recursos Policiales**



**Figura 31.** Diagrama de actividades del negocio

### 4.3.3 Modelo Conceptual





**Figura 32. Modelo Conceptual**

## CAPÍTULO V : REQUERIMIENTO DEL PROYECTO

### 5.1.1 Relación de Requerimientos Funcionales del Sistema

#### RF01 - Gestión de Datos Históricos de Criminalidad

- **Descripción:** El sistema debe permitir la importación, almacenamiento y gestión de datos históricos de criminalidad de al menos 2 años.
- **Prioridad:** Alta
- **Entradas:** Archivos CSV/Excel con datos de denuncias del SIDPOL
- **Salidas:** Base de datos normalizada con información delictiva

#### RF02 - Análisis Predictivo con Machine Learning

- **Descripción:** El sistema debe implementar algoritmos de ML (Random Forest, Regresión Logística) para predecir zonas y horarios de alta probabilidad delictiva.
- **Prioridad:** Alta
- **Entradas:** Datos históricos procesados
- **Salidas:** Predicciones de criminalidad con precisión >75%

### **RF03 - Gestión de Recursos Vehiculares**

- **Descripción:** El sistema debe mantener un registro actualizado de vehículos operativos y su estado.
- **Prioridad:** Alta
- **Entradas:** Lista de vehículos, estado operativo
- **Salidas:** Inventario de recursos disponibles

### **RF04 - Asignación Automática de Patrulleros**

- **Descripción:** El sistema debe asignar automáticamente vehículos a zonas de alta probabilidad delictiva basado en las predicciones.
- **Prioridad:** Alta
- **Entradas:** Predicciones de criminalidad, recursos disponibles
- **Salidas:** Asignaciones optimizadas de patrulleros

### **RF05 - Generación de Rutas de Patrullaje**

- **Descripción:** El sistema debe generar rutas optimizadas considerando distancia, tiempo y tráfico vehicular.
- **Prioridad:** Alta
- **Entradas:** Zonas asignadas, datos de tráfico
- **Salidas:** Rutas de patrullaje optimizadas

### **RF06 - Gestión de Usuarios y Perfiles**

- **Descripción:** El sistema debe manejar diferentes tipos de usuarios con permisos específicos.
- **Prioridad:** Media
- **Entradas:** Credenciales de usuario
- **Salidas:** Acceso autorizado según perfil

### **RF07 - Generación de Reportes**

- **Descripción:** El sistema debe generar reportes de eficiencia, incidencias y cumplimiento de rutas.
- **Prioridad:** Media
- **Entradas:** Datos de patrullaje ejecutado
- **Salidas:** Reportes estadísticos y dashboards

### **RF08 - Monitoreo en Tiempo Real**

- **Descripción:** El sistema debe permitir el seguimiento en tiempo real de las patrullas asignadas.
- **Prioridad:** Media
- **Entradas:** Ubicación GPS de patrulleros
- **Salidas:** Mapa de seguimiento en tiempo real

### **RF09 - Notificaciones y Alertas**

- **Descripción:** El sistema debe enviar notificaciones automáticas sobre cambios en asignaciones o incidencias.
- **Prioridad:** Baja
- **Entradas:** Eventos del sistema
- **Salidas:** Notificaciones push/email

#### RF10 - Integración con SIDPOL

- **Descripción:** El sistema debe integrarse con el Sistema de Denuncias Policiales para obtener datos actualizados.
- **Prioridad:** Alta
- **Entradas:** API del SIDPOL
- **Salidas:** Datos sincronizados

### 5.1.2 Especificación de Requerimientos

Especificación RF01 - Gestión de Datos Históricos de Criminalidad

Campo	Descripción
<b>Código</b>	RF01
<b>Nombre</b>	Gestión de Datos Históricos de Criminalidad
<b>Descripción</b>	El sistema debe permitir la carga, procesamiento y almacenamiento de datos históricos de criminalidad con una antigüedad mínima de 2 años, incluyendo validación de integridad y limpieza de datos.
<b>Actores</b>	Analista de Datos, Jefe de Patrullaje
<b>Precondiciones</b>	- Usuario autenticado con permisos de administración - Datos disponibles en formato CSV/Excel - Conexión a base de datos activa
<b>Postcondiciones</b>	- Datos históricos almacenados correctamente - Datos validados y limpios - Log de importación generado
<b>Flujo Principal</b>	1. Usuario selecciona archivo de datos 2. Sistema valida formato y estructura 3. Sistema procesa y limpia datos 4. Sistema almacena en base de datos 5. Sistema genera reporte de importación
<b>Flujos Alternativos</b>	1a. Archivo con formato incorrecto - Sistema muestra mensaje de error - Usuario debe corregir formato 3a. Datos duplicados encontrados - Sistema solicita confirmación para sobrescribir

<b>Excepciones</b>	<ul style="list-style-type: none"> <li>- Error de conexión a base de datos</li> <li>- Archivo corrupto</li> <li>- Espacio insuficiente en disco</li> </ul>
--------------------	--

#### Especificación RF02 - Análisis Predictivo con Machine Learning

<b>Campo</b>	<b>Descripción</b>
<b>Código</b>	RF02
<b>Nombre</b>	Análisis Predictivo con Machine Learning
<b>Descripción</b>	El sistema debe ejecutar algoritmos de Machine Learning para generar predicciones de criminalidad con precisión superior al 75%, utilizando datos históricos procesados.
<b>Actores</b>	Analista de Datos, Sistema
<b>Precondiciones</b>	- Datos históricos cargados - Modelos ML entrenados - Recursos computacionales disponibles
<b>Postcondiciones</b>	- Predicciones generadas con precisión >75% - Zonas de riesgo identificadas - Probabilidades calculadas por horario
<b>Flujo Principal</b>	1. Sistema carga datos históricos 2. Sistema ejecuta algoritmos ML 3. Sistema genera predicciones 4. Sistema calcula métricas de precisión 5. Sistema almacena resultados
<b>Flujos Alternativos</b>	4a. Precisión inferior al 75% - Sistema reajusta parámetros - Sistema re-entrena modelo
<b>Excepciones</b>	- Datos insuficientes para predicción - Error en algoritmos ML - Recursos computacionales agotados

#### 5.1.3 Requerimientos No Funcionales del Sistema

##### RNF01 - Rendimiento

- **Descripción:** El sistema debe procesar predicciones de criminalidad en menos de 5 minutos para un dataset de 100,000 registros.
- **Prioridad:** Alta
- **Métrica:** Tiempo de respuesta  $\leq$  5 minutos

##### RNF02 - Disponibilidad

- **Descripción:** El sistema debe estar disponible 24/7 con un uptime mínimo del 99.5%.

- **Prioridad:** Alta
- **Métrica:** Uptime  $\geq 99.5\%$

#### **RNF03 - Escalabilidad**

- **Descripción:** El sistema debe soportar hasta 500 usuarios concurrentes sin degradación del rendimiento.
- **Prioridad:** Media
- **Métrica:** Capacidad de 500 usuarios simultáneos

#### **RNF04 - Seguridad**

- **Descripción:** El sistema debe implementar autenticación de doble factor y encriptación de datos sensibles.
- **Prioridad:** Alta
- **Métrica:** Cumplimiento de estándares ISO 27001

#### **RNF05 - Usabilidad**

- **Descripción:** El sistema debe ser intuitivo con una curva de aprendizaje máxima de 2 horas para usuarios finales.
- **Prioridad:** Media
- **Métrica:** Tiempo de aprendizaje  $\leq 2$  horas

#### **RNF06 - Compatibilidad**

- **Descripción:** El sistema debe ser compatible con navegadores web modernos (Chrome, Firefox, Edge) y dispositivos móviles.
- **Prioridad:** Media
- **Métrica:** Compatibilidad con navegadores principales

#### **RNF07 - Mantenibilidad**

- **Descripción:** El código debe seguir estándares de programación y documentación para facilitar el mantenimiento.
- **Prioridad:** Baja
- **Métrica:** Cobertura de documentación  $\geq 80\%$

#### **RNF08 - Portabilidad**

- **Descripción:** El sistema debe poder ejecutarse en diferentes sistemas operativos (Windows, Linux).
- **Prioridad:** Baja
- **Métrica:** Compatibilidad multiplataforma

#### **RNF09 - Confiabilidad**

- **Descripción:** El sistema debe tener una tasa de error inferior al 0.1% en las predicciones generadas.
- **Prioridad:** Alta

- **Métrica:** Tasa de error  $\leq 0.1\%$

## RNF10 - Recuperación

- **Descripción:** El sistema debe recuperarse automáticamente de fallos en menos de 15 minutos.
- **Prioridad:** Media
- **Métrica:** Tiempo de recuperación  $\leq 15$  minutos

### 5.3 Casos de Uso del Sistema

#### 5.3.1 Diagrama de Actores del Sistema

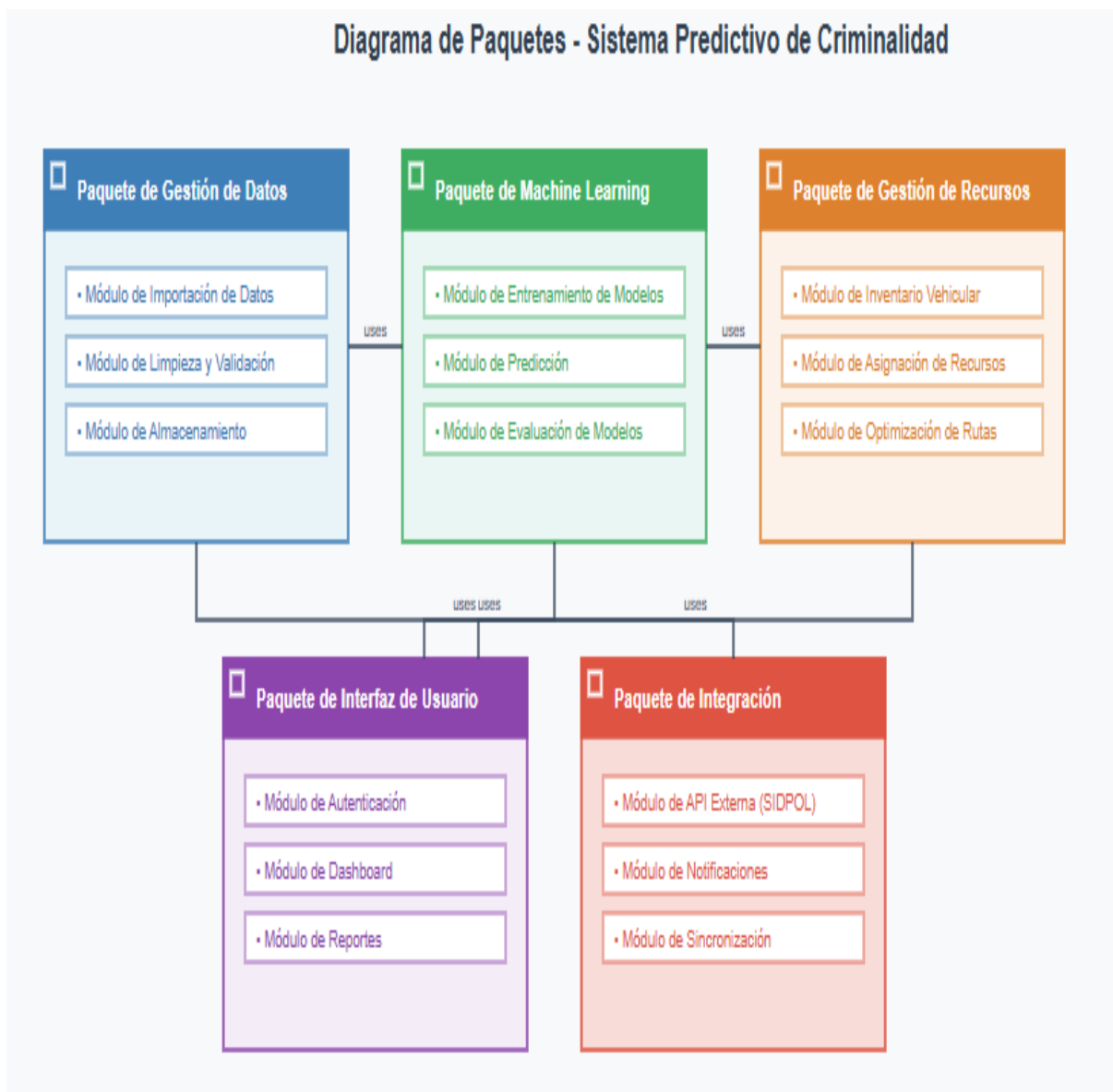
##### 5.3.1.1 Especificación de los Actores del Sistema

Actor	Descripción	Responsabilidades
Jefe de Patrullaje	Responsable principal de la planificación y asignación de recursos de patrullaje	<ul style="list-style-type: none"> <li>- Revisar y aprobar asignaciones automáticas</li> <li>- Modificar rutas según criterio operativo</li> <li>- Supervisar cumplimiento de rutas</li> <li>- Generar reportes de gestión</li> </ul>
Comisario	Supervisor general de la comisaría	<ul style="list-style-type: none"> <li>- Revisar reportes de eficiencia</li> <li>- Aprobar cambios estratégicos</li> <li>- Supervisar cumplimiento de objetivos</li> <li>- Validar resultados del sistema</li> </ul>
Efectivo Policial	Personal operativo que ejecuta las patrullas	<ul style="list-style-type: none"> <li>- Consultar rutas asignadas</li> <li>- Reportar incidencias durante patrullaje</li> <li>- Actualizar estado de cumplimiento</li> <li>- Confirmar llegada a puntos de control</li> </ul>
Analista de Datos	Especialista en análisis de datos y Machine Learning	<ul style="list-style-type: none"> <li>- Mantener modelos predictivos</li> <li>- Actualizar algoritmos ML</li> <li>- Validar precisión de predicciones</li> <li>- Generar reportes técnicos</li> </ul>
Administrador del Sistema	Responsable técnico del	<ul style="list-style-type: none"> <li>- Gestionar usuarios y</li> </ul>

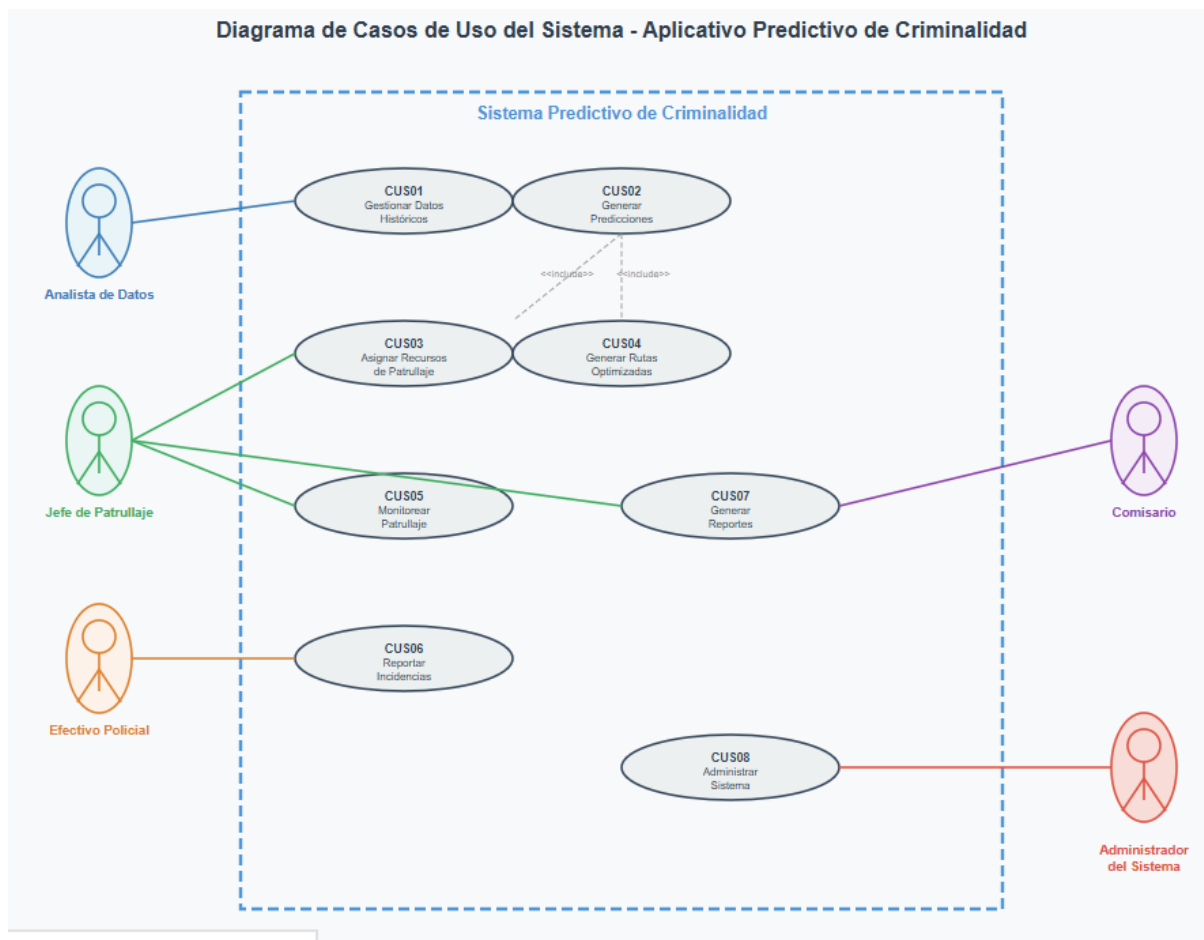
	sistema	permisos - Configurar parámetros del sistema - Realizar mantenimiento técnico - Gestionar copias de seguridad
--	---------	--

### 5.3.2 Diagrama de Paquetes

El sistema se organizará en los siguientes paquetes principales:



### 5.3.3 Diagrama de Casos de Uso del Sistema



### 5.3.4 Especificaciones de CUS

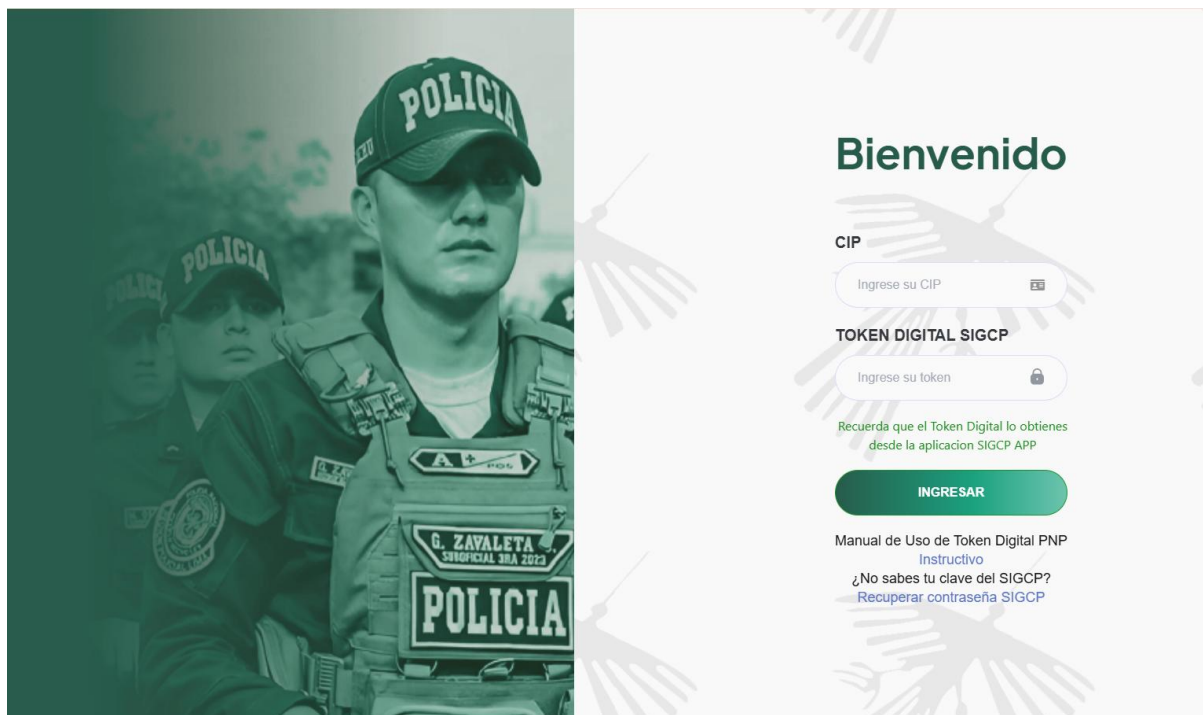
#### CUS01 - Gestionar Datos Históricos

Campo	Descripción
Código	CUS01
Nombre	Gestionar Datos Históricos
Descripción	El analista de datos puede importar, validar y almacenar datos históricos de criminalidad en el sistema
Actor Principal	Analista de Datos
Actores Secundarios	Sistema
Precondiciones	<ul style="list-style-type: none"> <li>- Usuario autenticado</li> <li>- Permisos de administración de datos</li> <li>- Datos disponibles en formato compatible</li> </ul>
Postcondiciones	<ul style="list-style-type: none"> <li>- Datos históricos almacenados</li> <li>- Datos validados y limpios</li> </ul>



	- Log de importación generado
<b>Flujo Principal</b>	<ol style="list-style-type: none"> <li>1. Analista selecciona opción "Gestionar Datos"</li> <li>2. Sistema muestra interfaz de importación</li> <li>3. Analista carga archivo de datos</li> <li>4. Sistema valida formato y estructura</li> <li>5. Sistema procesa y limpia datos</li> <li>6. Sistema almacena en base de datos</li> <li>7. Sistema genera reporte de importación</li> <li>8. Sistema notifica finalización exitosa</li> </ol>
<b>Flujos Alternativos</b>	<ol style="list-style-type: none"> <li>4a. Formato de archivo incorrecto <ul style="list-style-type: none"> <li>- Sistema muestra mensaje de error</li> <li>- Sistema solicita archivo correcto</li> </ul> </li> <li>5a. Datos duplicados encontrados <ul style="list-style-type: none"> <li>- Sistema solicita confirmación</li> <li>- Analista decide acción a tomar</li> </ul> </li> </ol>
<b>Excepciones</b>	<ul style="list-style-type: none"> <li>- Error de conexión a base de datos</li> <li>- Archivo corrupto o inaccesible</li> <li>- Espacio insuficiente en disco</li> </ul>

## AVANCE DE NUESTRO PROTOTIPO



## REFERENCIA BIBLIOGRÁFICA

- Aziz, R. M., Hussain, A., Sharma, P., & Kumar, P. (2022). Machine Learning-based Soft Computing Regression Analysis Approach for Crime Data Prediction. *Karbala International Journal of Modern Science*, 8(1), 1–19. <https://doi.org/10.33640/2405-609X.3197>
- Hälterlein, J. (2021). Epistemologies of predictive policing: Mathematical social science, social physics and machine learning. *Big Data and Society*, 8(1). <https://doi.org/10.1177/20539517211003118>
- Hannach, H. El, & Benkhalifa, M. (2018). WordNet based Implicit Aspect Sentiment Analysis for Crime Identification from Twitter. In *IJACSA International Journal of Advanced Computer Science and Applications* (Vol. 9, Issue 12). [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- He, J., & Zheng, H. (2021). Prediction of crime rate in urban neighborhoods based on machine learning. *Engineering Applications of Artificial Intelligence*, 106. <https://doi.org/10.1016/j.engappai.2021.104460>
- Jenga, K., Catal, C., & Kar, G. (2023). Machine learning in crime prediction. *Journal of Ambient Intelligence and Humanized Computing*, 14(3), 2887–2913. <https://doi.org/10.1007/s12652-023-04530-y>
- Kshatri, S. S., Singh, D., Narain, B., Bhatia, S., Quasim, M. T., & Sinha, G. R. (2021). An Empirical Analysis of Machine Learning Algorithms for Crime Prediction Using Stacked Generalization: An Ensemble Approach. *IEEE Access*, 9, 67488–67500. <https://doi.org/10.1109/ACCESS.2021.3075140>
- Mandalapu, V., Elluri, L., Vyas, P., & Roy, N. (2023). Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions. *IEEE Access*, 11, 60153–60170. <https://doi.org/10.1109/ACCESS.2023.3286344>
- Miyano, K., Shinkuma, R., Shiode, N., Shiode, S., Sato, T., & Oki, E. (2020). Multi-UAV Allocation Framework for Predictive Crime Deterrence and Data Acquisition. *Internet of Things (Netherlands)*, 11. <https://doi.org/10.1016/j.iot.2020.100205>
- Parthasarathy, S., Rahman, B. S. A., Jegananathan, S., & Sathick, J. (2020). *Survey on Crime Analysis and Prediction Using Data Mining and Machine Learning Techniques*. <https://www.researchgate.net/publication/349277409>
- Podoletz, L. (2023). We have to talk about emotional AI and crime. *AI and Society*, 38(3), 1067–1082. <https://doi.org/10.1007/s00146-022-01435-w>
- Rosés, R., Kadar, C., & Malleson, N. (2021). A data-driven agent-based simulation to predict crime patterns in an urban environment. *Computers, Environment and Urban Systems*, 89. <https://doi.org/10.1016/j.compenvurbsys.2021.101660>
- Saeed, R. M., & Abdulmohsin, H. A. (2023). A study on predicting crime rates through machine learning and data mining using text. In *Journal of Intelligent Systems* (Vol. 32, Issue 1). De Gruyter Open Ltd. <https://doi.org/10.1515/jisys-2022-0223>
- Safat, W., Asghar, S., & Gillani, S. A. (2021). Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques. *IEEE Access*, 9, 70080–70094. <https://doi.org/10.1109/ACCESS.2021.3078117>
- Salazar, B. P. (2024). Predictive Criminology: A Near Future or a Distant Fiction? *Novum Jus*, 18(3), 343–396. <https://doi.org/10.14718/NovumJus.2024.18.3.13>
- 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>
- Sarzaeim, P., Mahmoud, Q. H., Azim, A., Bauer, G., & Bowles, I. (2023). A Systematic Review of Using Machine Learning and Natural Language Processing in Smart Policing. In *Computers* (Vol. 12, Issue 12). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/computers12120255>
- Schiff, K. J., Schiff, D. S., Adams, I. T., McCrain, J., & Mourtgos, S. M. (2023). Institutional factors driving citizen perceptions of AI in government: Evidence from a survey experiment on policing. *Public Administration Review*.  
<https://doi.org/10.1111/puar.13754>
- Shah, N., Bhagat, N., & Shah, M. (2021). Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention. In *Visual Computing for Industry, Biomedicine, and Art* (Vol. 4, Issue 1). Springer.  
<https://doi.org/10.1186/s42492-021-00075-z>
- Simmler, M., Brunner, S., Canova, G., & Schedler, K. (2023). Smart criminal justice: exploring the use of algorithms in the Swiss criminal justice system. *Artificial Intelligence and Law*, 31(2), 213–237. <https://doi.org/10.1007/s10506-022-09310-1>
- Watts, D., Moulden, H., Mamak, M., Upfold, C., Chaimowitz, G., & Kapczinski, F. (2021). Predicting offenses among individuals with psychiatric disorders - A machine learning approach. *Journal of Psychiatric Research*, 138, 146–154.  
<https://doi.org/10.1016/j.jpsychires.2021.03.026>