

**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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

**CAPÍTULO III: ESTADO DEL ARTE 16**

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

**CAPÍTULO IV: MODELADO DE NEGOCIO ……………………………………………………20**

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

# INTRODUCCIÓN

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

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

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

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

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

La presente tesis está dividida en las siguientes partes:

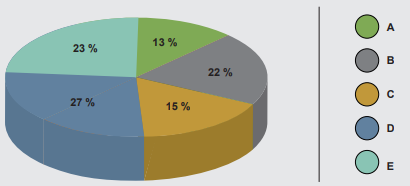
* En el capítulo I, muestra la visión del proyecto, donde se describen y detallan los antecedentes que se tomaron como referencia, la problemática a definir, así como el objetivo general y específicos. Seguidamente se menciona la justificación del tema y el alcance a desarrollarse.

# CAPÍTULO I: VISIÓN DEL PROYECTO

## 1.1 Antecedentes del Problema

### 1.1.1 El Negocio

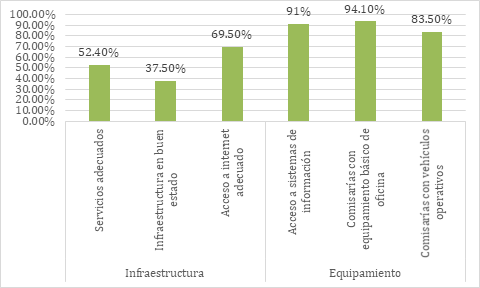
Las comisarías de la Policía Nacional del Perú son la célula base de esta institución y suman un total de 1,318 a nivel nacional. Según su tipo pueden clasificarse en A, B, C, D, E de acuerdo a su capacidad de efectivos policiales, área construida y cobertura de habitantes. La A representa la de mayor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



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

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

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



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

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

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

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

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

### 1.1.2 Procesos del Negocio

Diagrama

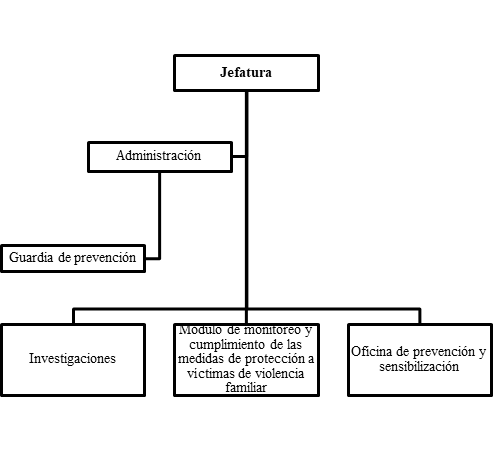
El contenido generado por IA puede ser incorrecto.

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

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

### 1.1.3 Organigrama

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

****

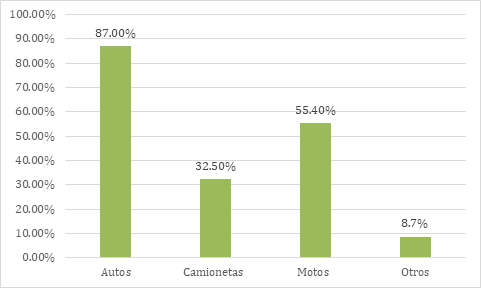
***Figura 1.4.*** *Organigrama de una comisaría PNP*

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

## 1.2 Formulación del Problema

### 1.2.1 Realidad Problemática

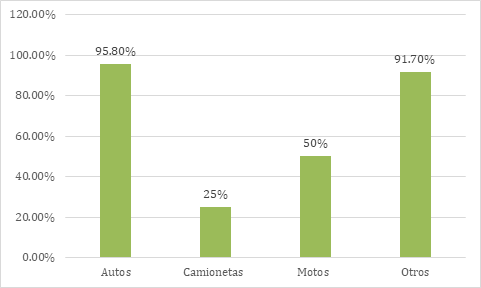
Los recursos en la institución policial resultan relevantes para mejorar la seguridad ciudadana, pero cuando se habla de vehículos policiales o también conocidos como patrulleros, se tiene la lamentable cifra de unidades que no pueden ser utilizadas debido a que se encuentran inoperativas. En ese sentido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



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

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

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



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

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

### 1.2.2 Descripción del Problema

#### 1.2.2.1 Problema Principal

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

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

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

#### 1.2.2.2 Problema Secundario

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

## 1.3 Objetivos del Proyecto

### 1.3.1 Marco Lógico

#### 1.3.1.1 Árbol del Problemas:

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

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

#### 1.3.1.2 Árbol de Objetivos:



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

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

### 1.3.2 Objetivo General

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

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

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

### 1.3.3 Objetivos Específicos

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

## 

## 1.4 Justificación del Proyecto

### 1.4.1 Justificación Académica

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

Gráfico, Gráfico de cajas y bigotes

El contenido generado por IA puede ser incorrecto.

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

***Fuente:*** *(COMEXPERÚ, 2025)*

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

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

### 1.4.2 Beneficios Tangibles

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

### 1.4.3 Beneficios Intangibles

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

## 1.5 Alcance del Proyecto

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

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

# CAPÍTULO II: MARCO TEÓRICO

# CAPÍTULO III: ESTADO DEL ARTE

**3.1.** **Artículos**

### 3.1.1 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)

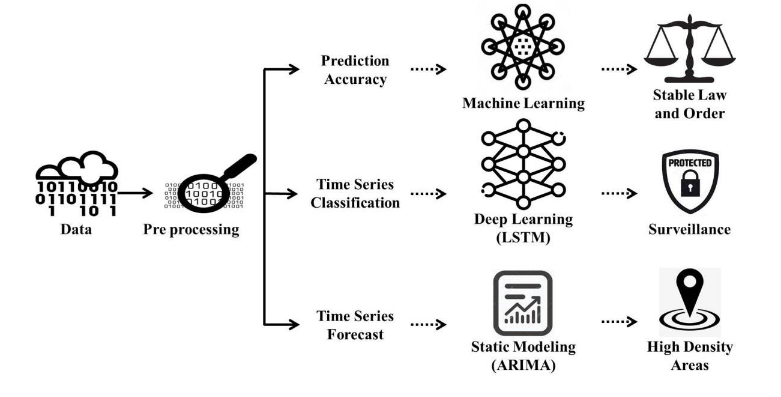
**(DOI: 10.1109/ACCESS.2021.3078117)**

**Citado 100 veces en Scopus**

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 1:*** *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, Naïve 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 al aporte de este artículo implementaremos en nuestro proyecto de tesis, una base empírica sobre la efectividad de los métodos de predicción y pronóstico del crimen mediante técnicas de aprendizaje automático y aprendizaje profundo. Aplicaremos estas metodologías en el desarrollo de una herramienta que utilizará datos históricos de criminalidad de Lima Metropolitana para identificar patrones y predecir áreas de alta criminalidad, lo que permitirá a las autoridades asignar recursos policiales de manera más eficiente y oportuna. Con el uso de modelos predictivos basados en Machine Learning, específicamente los algoritmos LSTM y ARIMA, en nuestra plataforma podremos anticipar futuras olas de criminalidad y direccionar las patrullas a las zonas de mayor riesgo. Este enfoque no solo mejorará la toma de decisiones en tiempo real, sino que también optimizará la prevención y la seguridad pública en la ciudad.

En el artículo, se menciona que 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.

### 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)**

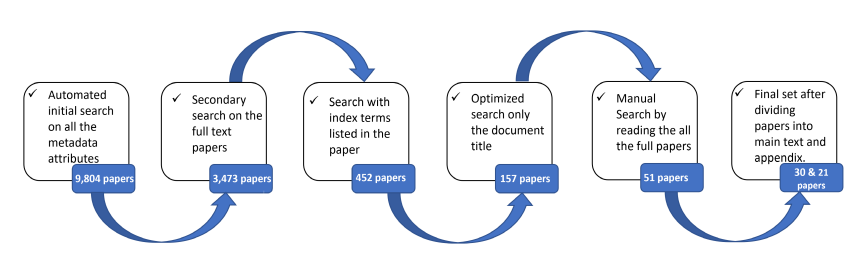
**Citado 59 veces en Scopus**

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 1:*** *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**

Este artículo es de gran utilidad para nuestra proyecto de tesis, ya que implementaremos un marco teórico sólido sobre el uso de técnicas avanzadas de aprendizaje automático (Machine Learning, ML) y aprendizaje profundo (Deep Learning, DL) en la predicción de delitos. En nuestro proyecto, implementaremos algoritmos de ML y DL para analizar patrones históricos de criminalidad en Lima Metropolitana y predecir la probabilidad de ocurrencia de crímenes en áreas específicas. Esto permitirá a las autoridades policiales asignar recursos de manera más eficiente, optimizando las patrullas y la respuesta ante incidentes. Al aplicar estos enfoques, se logrará una mayor precisión en la predicción de crímenes, lo que, a su vez, mejorará la seguridad pública y la asignación dinámica de recursos en función de las predicciones generadas por el modelo.

En el artículo, se menciona que 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

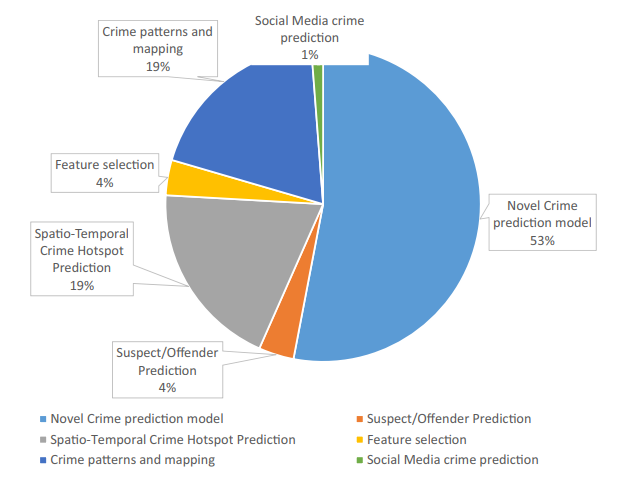
(Jenga et al., 2023)

**(DOI: 10.1007/s12652-023-04530-y)**

**Citado 55 veces en Scopus**

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 1:*** *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, Springer Link, Science Direct y Scopus. A través de búsquedas avanzadas con términos específicos como “crime prediction” y “machine learning”, se seleccionaron artículos que cumplían 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 este artículo, aplicaremos las técnicas de minería de datos y aprendizaje automático descritas en el artículo para desarrollar un sistema predictivo capaz de anticipar la ocurrencia de delitos en áreas específicas de Lima Metropolitana. Al implementar modelos de aprendizaje supervisado y no supervisado, podremos identificar patrones delictivos y predecir crímenes con mayor precisión. Esta capacidad será fundamental para mejorar la asignación de recursos policiales de manera más eficiente y focalizada, asegurando que las fuerzas del orden puedan intervenir de forma anticipada en las zonas de mayor riesgo. Este artículo, al sintetizar las metodologías más avanzadas y los enfoques en la predicción del crimen, aporta una base sólida para nuestra investigación, ya que proporciona una comprensión detallada de los métodos más efectivos, los cuales implementaremos y adaptamos a las características particulares de Lima Metropolitana.

En el artículo, se menciona 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.

### 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**

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

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

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

La investigación no debe terminar con la correlación de variables, si bien es un conocimiento importante, el que permitirá propuestas de soluciones es la explicativa. El motivo es que, en primer lugar, la correlación no significa causalidad y, en segundo lugar, porque, las ubicaciones a 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 recopilaron tuits de Twitter utilizando la biblioteca Snscrape, un programa de búsqueda de servicios de redes sociales en Python.

**C) Etapas**

Etapa 1: comprensión del patrón.

Etapa 2: Regresión Lasso.

Etapa 3: Clasificación con aprendizaje automático:

- Logistic Regression

- Decision Tree

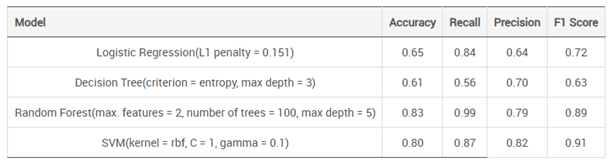
- Random Forest

- SVM

Etapa 4: Análisis de sentimientos.

**Tabla 1**

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

******

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

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

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

### 3.1.5 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://dx.doi.org/10.14569/IJACSA.2018.091222)**)**

**Citado 22 veces en Scopus**

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

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

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

**B) Metodología**

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

**C) Etapas**

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

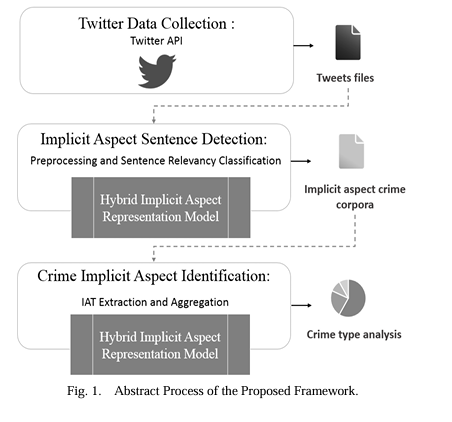
Etapa 3: Identificación de aspectos implícitos (IAI):

* Extracción de términos implícitos:
* Agregación de términos

Etapa 4: Análisis de sentimientos.

***Figura 1***

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

******

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

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

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

En ese sentido, este artículo se centra en conceptos clave como WordNet y sus relaciones semánticas, indispensables para enriquecer el vocabulario de delitos implícitos; el análisis de sentimiento basado en aspectos implícitos, que permite descubrir indicios criminales ocultos en los tuits; y el esquema de ponderación TF-ICF junto con los clasificadores 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](https://www.scopus.com/record/display.uri?eid=2-s2.0-85092078489&origin=resultslist&sort=plf-f&src=s&sid=7315aca7d5b4f3b3a43b9e35cdc03c0a&sot=b&sdt=b&s=TITLE-ABS-KEY%28SURVEY+ON+CRIME+ANALYSIS+AND+PREDICTION+USING+DATA+MINING+TECHNIQUES%29&sl=139&sessionSearchId=7315aca7d5b4f3b3a43b9e35cdc03c0a&relpos=12)

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

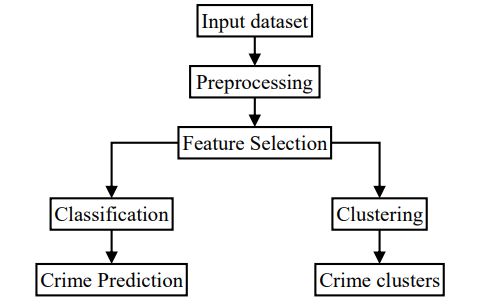
Saravanan P.; Selvaprabhu J.; Arun Raj L.; Abdul Azeez Khan A.; Javubar Sathick K. (2021)

**DOI: 10.1007/978-981-15-7241-8\_31**

**Citado 34 veces en Scopus**

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 1:*** *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 al aporte de este artículo, implementaremos una base teórica y práctica sobre el uso de técnicas de minería de datos para analizar y predecir el comportamiento criminal. En nuestro proyecto, implementaremos enfoques de minería de datos, aprendizaje supervisado y no supervisado para identificar patrones delictivos y predecir crímenes en Lima Metropolitana. Al integrar estas metodologías, se podrá anticipar la ocurrencia de delitos en áreas específicas, permitiendo una asignación más eficiente de los recursos policiales, basándose en predicciones de zonas de alto riesgo. Además, el artículo destaca la relevancia de estos enfoques para mejorar la capacidad de las agencias policiales para prevenir y responder a los delitos de manera más eficaz, lo cual es crucial para la implementación exitosa de nuestro aplicativo.

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 Machine learning in crime prediction

### Aprendizaje automático en la predicción de delitos

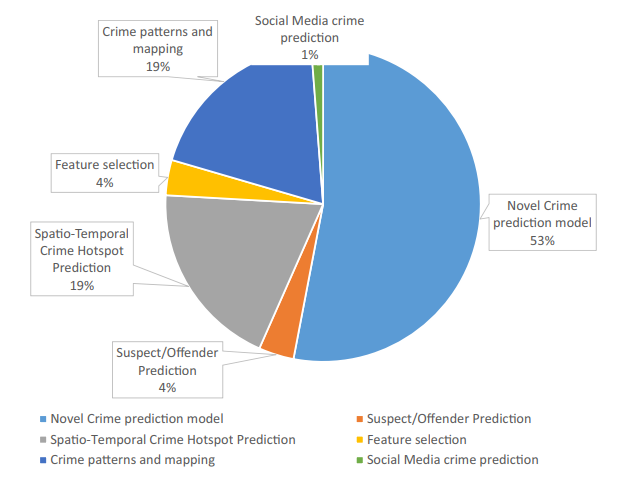
Karabo Jenga; Cagatay Catal; Gorkem Kar (2023)

**(DOI: 10.1007/s12652-023-04530-y)**

**Citado 49 veces en Scopus**

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 1:*** *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, Springer Link, Science Direct y Scopus. A través de búsquedas avanzadas con términos específicos como “crime prediction” y “machine learning”, se seleccionaron artículos que cumplían 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**

El artículo es de gran relevancia para el desarrollo de nuestra tesis, se aplicarán las técnicas de minería de datos y aprendizaje automático descritas en el artículo para desarrollar un sistema predictivo capaz de anticipar la ocurrencia de delitos en áreas específicas de Lima Metropolitana. Al implementar modelos de aprendizaje supervisado y no supervisado, podremos identificar patrones delictivos y predecir crímenes con mayor precisión. Esta capacidad será fundamental para mejorar la asignación de recursos policiales de manera más eficiente y focalizada, asegurando que las fuerzas del orden puedan intervenir de forma anticipada en las zonas de mayor riesgo. Este artículo, al sintetizar las metodologías más avanzadas y los enfoques en la predicción del crimen, aporta una base sólida para nuestra investigación, ya que proporciona una comprensión detallada de los métodos más efectivos, los cuales implementaremos y adaptamos a las características particulares de Lima Metropolitana.

En el artículo, se menciona 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.

### 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**

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.

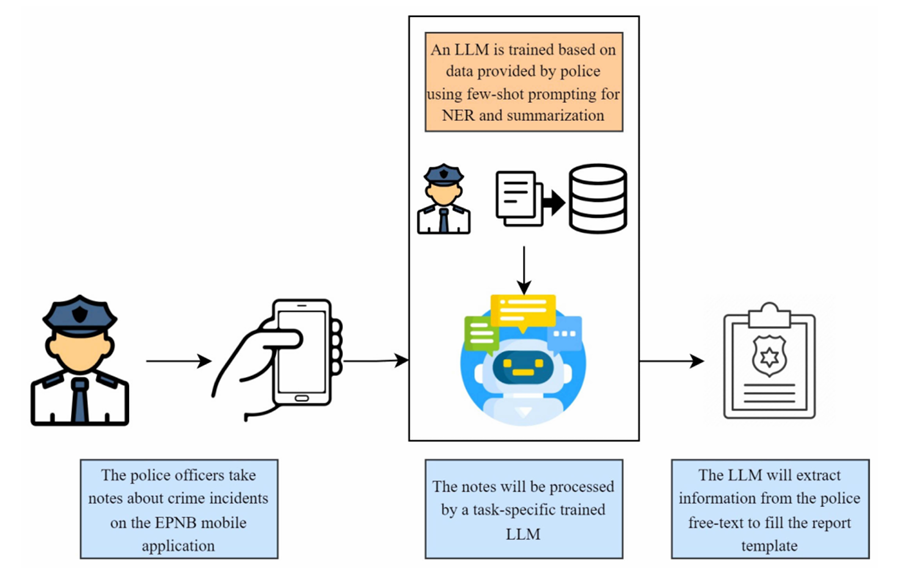
Etapa 2: Criterios de inclusión y snowballing para ampliar la búsqueda.

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 1***

*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**

**Este artículo aporta significativamente a nuestra tesis para mejorar la asignación de recursos policiales en Lima Metropolitana" al proporcionar una revisión sistemática de las aplicaciones del aprendizaje automático (ML) y el procesamiento del lenguaje natural (NLP) en la vigilancia policial inteligente. En nuestro aplicativo implementaremos algoritmos de clasificación y regresión como SVM, Random Forest y KDE, así como técnicas de NLP para el análisis de informes narrativos de crímenes que faciliten la extracción automática de información clave y la clasificación de datos en grandes volúmenes de documentos de texto no estructurado, permitiendo la gestión eficiente de grandes volúmenes de datos para la identificación de patrones delictivos en Lima Metropolitana.** 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.

### 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

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

### 

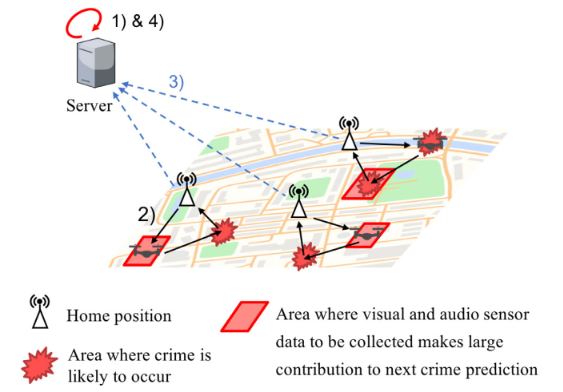
### El marco se basa en un modelo predictivo que guía el despliegue de los UAVs no solo para disuadir incidentes delictivos, sino también para recolectar datos históricos relevantes que retroalimentan y mejoran la eficacia futura del sistema. Esta estrategia permite una optimización continua en la asignación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

**B) Metodología**

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

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

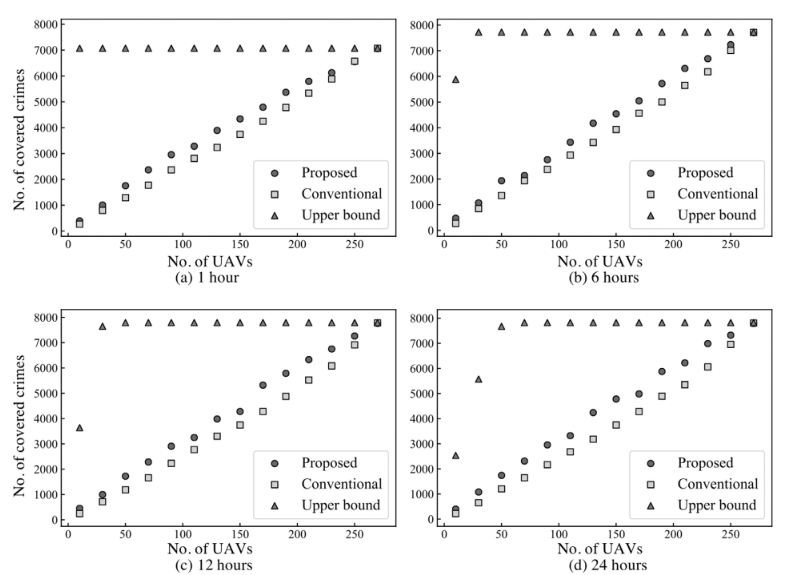
**Figura 1:** Arquitectura del sistema asumida en este estudio.



**C) Etapas**

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

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



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

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

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

### 3.1.10 Predictive criminology: a near future or a fiction in the making?

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

Pérez Salazar, (2024)

### (DOI: 10.14718/NovumJus.2024.18.3.13)

**Citado 16 veces en Scopus**

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

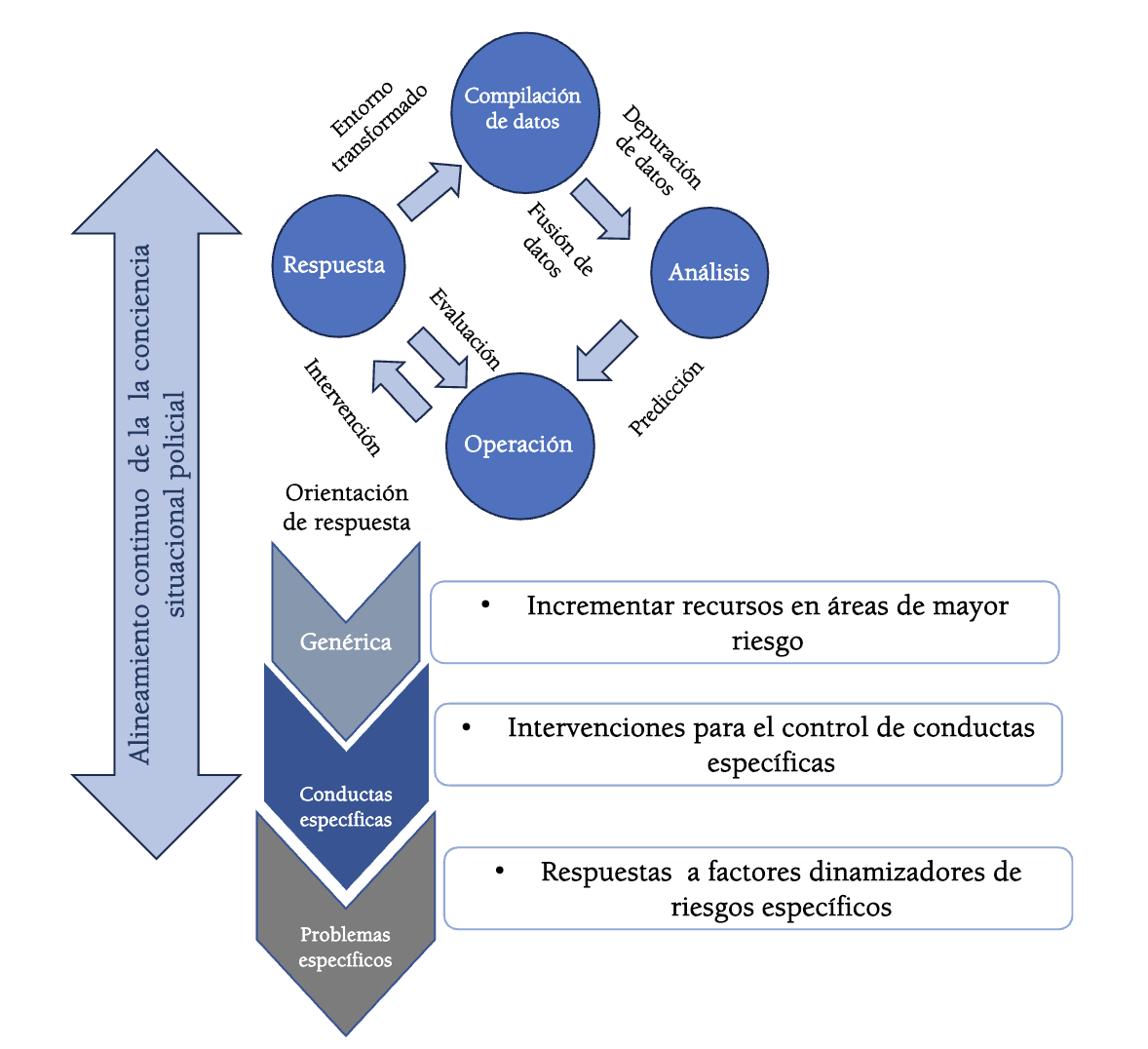
### Pérez Salazar señala que si bien estos sistemas ofrecen beneficios operativos y contribuyen a la prevención del delito, presentan riesgos en cuanto a sesgos algorítmicos y falta de interpretación contextual. Aun así, se plantea que el uso de estos modelos puede fortalecer la capacidad del Estado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

**B) Metodología**

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

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

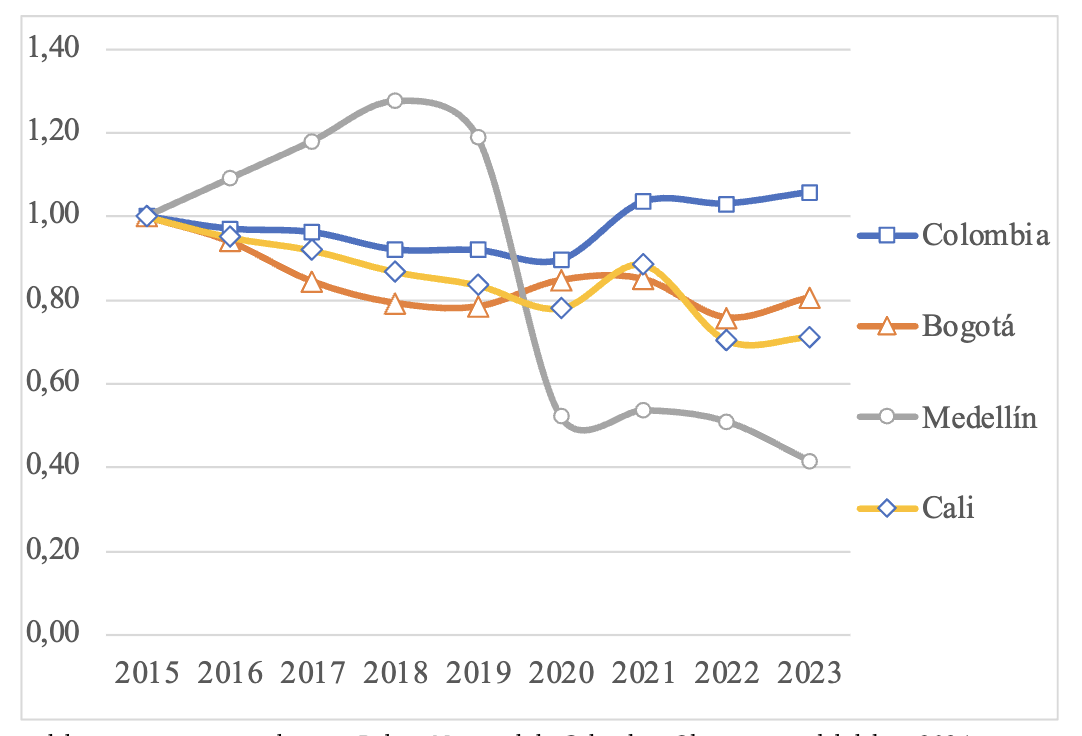
**Figura 1:** Ciclo de gestión del conocimiento en el enfoque de policía predictiva



**C) Etapas**

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

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



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

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

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

### 1.1.11 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, Kaylyn Jackson; Schiff, Daniel S.; Adams, Ian T.; McCrain, Joshua; Mourtgos, Scott M. (2023)

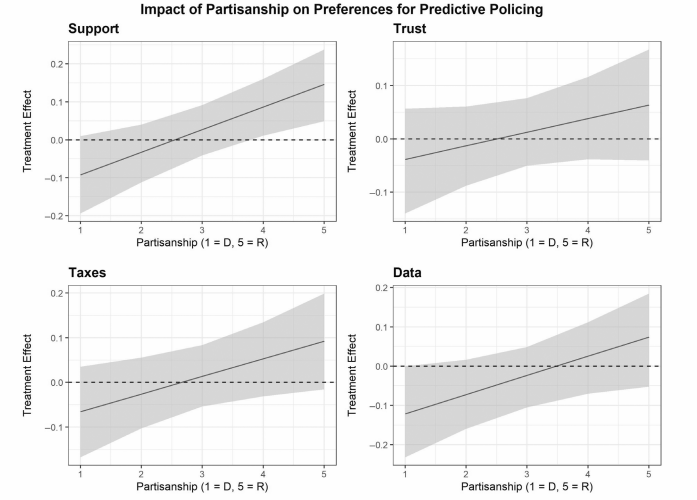
**(DOI: 10.1111/puar.13754)**

**Citado 11 veces en Scopus**

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:**

**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**

Este artículo es de gran utilidad para nuestra tesis propuesta al proporcionar un análisis para la implementar como influyen en la percepción pública y la efectividad de la implementación de herramientas de inteligencia artificial (IA) en el ámbito policial. En nuestro proyecto, aplicaremos los principios de la confianza pública y la proximidad burocrática al diseñar el sistema de IA que se utilizará para prever delitos en Lima. Con base en los hallazgos del artículo, implementaremos un enfoque que considere la proximidad de las autoridades locales, como la Policía Nacional del Perú, para aumentar la aceptación pública del sistema predictivo. Esto contribuirá a mejorar la asignación de recursos policiales, asegurando que los esfuerzos sean más efectivos y estén alineados con la confianza de la comunidad en las instituciones locales, optimizando así la distribución de patrullas y recursos en zonas de alta criminalidad.

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.

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

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

Lena Podoletz; (2022)

**DOI: 10.1007/s00146-022-01435-w**

**Citado 29 veces en Scopus**

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.

1. **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.

1. **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.

1. **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**

Este artículo aporta una presente proyecto tesis, la implementación de diseño y alcance de nuestro aplicativo predictivo, especialmente en lo referente a los límites éticos del uso de inteligencia artificial en contextos de seguridad pública. A partir de los cuestionamientos planteados sobre el uso de tecnologías invasivas como la inteligencia artificial emocional, nuestro proyecto priorizará el uso de modelos predictivos basados exclusivamente en datos históricos verificables, evitando así cualquier forma de inferencia subjetiva sobre el comportamiento o estado emocional de las personas. Con ello, se busca desarrollar una herramienta confiable y técnicamente fundamentada, que permita anticipar zonas de riesgo de manera objetiva, contribuyendo a mejorar la planificación policial sin comprometer los derechos fundamentales de los ciudadanos.

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.

### 1.1.13 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)**

**Citado 12 veces en Scopus**

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.

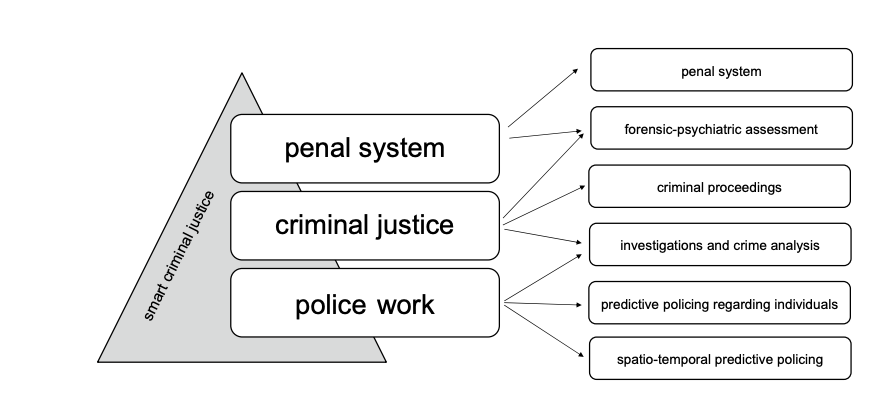
El trabajo se estructura en torno a tres ejes: la identificación de las aplicaciones algorítmicas actualmente implementadas; los motivos institucionales y políticos que impulsaron su adopción; y la evaluación de su impacto percibido en la práctica cotidiana de operadores policiales y penitenciarios. 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 1: 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**

Este artículo proporciona un marco valioso para el desarrollo de nuestro proyecto de tesis, ya que expone cómo vamos a implementar el uso de algoritmos en el ámbito penal puede optimizar procesos como la prevención del delito, el análisis criminal y la gestión de recursos institucionales. A partir de esta referencia, en nuestro proyecto se aplicarán modelos de aprendizaje automático para detectar patrones de criminalidad en Lima Metropolitana y, con base en ello, generar recomendaciones precisas para una distribución más eficiente de los recursos policiales. Esta evidencia empírica y metodológica será útil para estructurar nuestro sistema predictivo bajo parámetros realistas, asegurando tanto su funcionalidad operativa como su coherencia con estándares éticos y legales, sin sustituir la toma de decisiones humanas, pero sí fortaleciéndose con apoyo tecnológico basado en datos verificables.

En el artículo, se menciona que 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 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)**

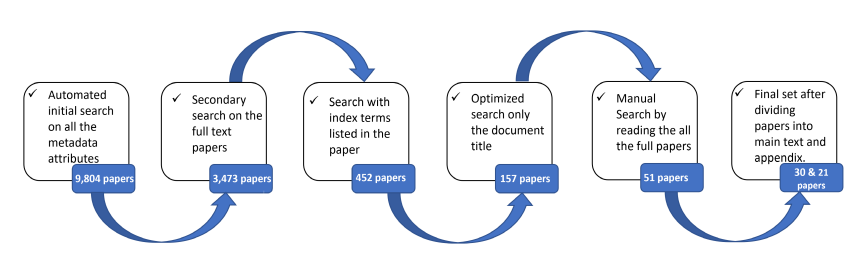
**Citado 59 veces en Scopus**

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 1:*** *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**

Este artículo es de gran utilidad para nuestra tesis, ya que vamos a implementar un marco teórico sólido sobre el uso de técnicas avanzadas de aprendizaje automático (Machine Learning, ML) y aprendizaje profundo (Deep Learning, DL) en la predicción de delitos. En nuestro proyecto, también implementaremos algoritmos de ML y DL para analizar patrones históricos de criminalidad en Lima Metropolitana y predecir la probabilidad de ocurrencia de crímenes en áreas específicas. Esto permitirá a las autoridades policiales asignar recursos de manera más eficiente, optimizando las patrullas y la respuesta ante incidentes. Al aplicar estos enfoques, se logrará una mayor precisión en la predicción de crímenes, lo que, a su vez, mejorará la seguridad pública y la asignación dinámica de recursos en función de las predicciones generadas por el modelo.

En el artículo, se menciona que 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.

### 1.1.15 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)

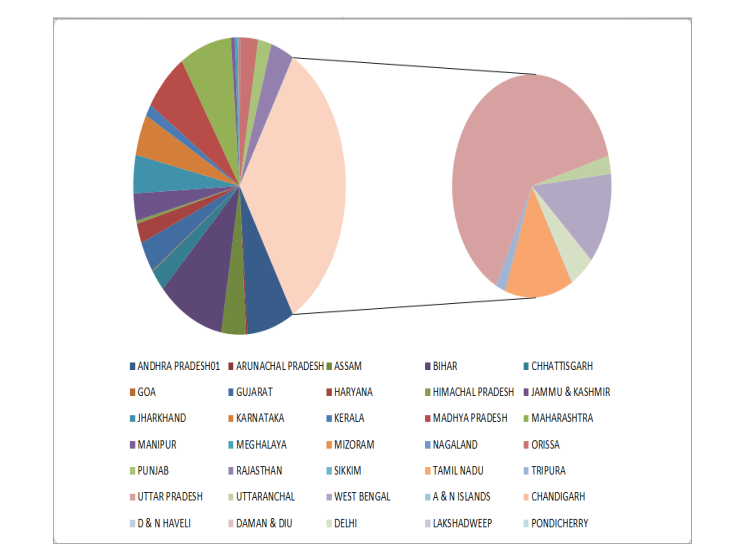
**(DOI: 10.1109/ACCESS.2021.3075140)**

**Citado 96 veces en Scopus**

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 1:*** *Distribución de la delincuencia en las regiones seleccionadas.*

*.*

**B) Metodología**

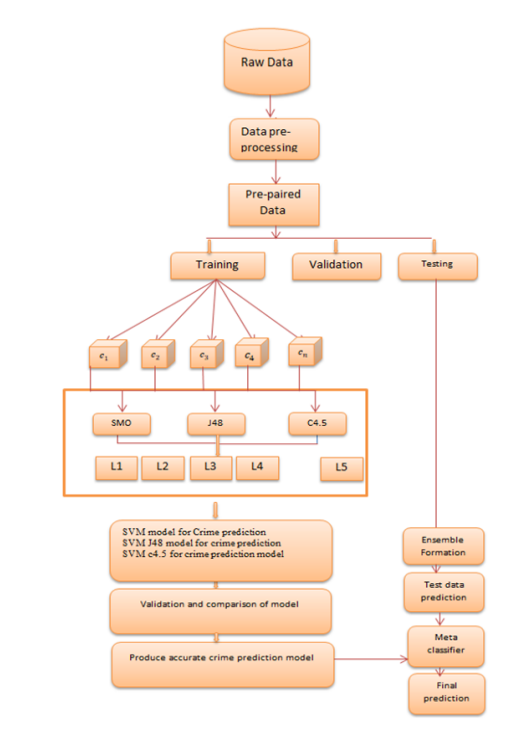
El enfoque metodológico adoptado es **investigación empírica comparativa**, en la que se realiza un análisis detallado de varios algoritmos de machine learning aplicados a la predicción de crímenes. Se recopiló un conjunto de datos de crímenes violentos en India, utilizando fuentes como el **National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

1. **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.
2. **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.
3. **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)**.
4. **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 2:*** *Clasificador propuesto para el modelo de predicción de delitos basado en conjuntos.*

*.*

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

Gracias al aporte de este artículo implementaremos nuestra tesis, la metodología de generalización apilada (stacked generalization) y aprendizaje en conjunto (ensemble learning) propuesta en el estudio será fundamental para nuestra implementación. En nuestro proyecto, aplicaremos el enfoque de SVM stacking y otros algoritmos de machine learning para predecir la ocurrencia de crímenes en diferentes áreas de Lima Metropolitana. Al incorporar técnicas avanzadas como estas, lograremos mejorar la precisión de nuestras predicciones, lo que permitirá a las autoridades asignar recursos de manera más eficiente en áreas de alta criminalidad. La alta precisión alcanzada por el modelo propuesto en el artículo (99,5%) servirá como un referente para evaluar y optimizar nuestro sistema de predicción en tiempo real.

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.

# 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

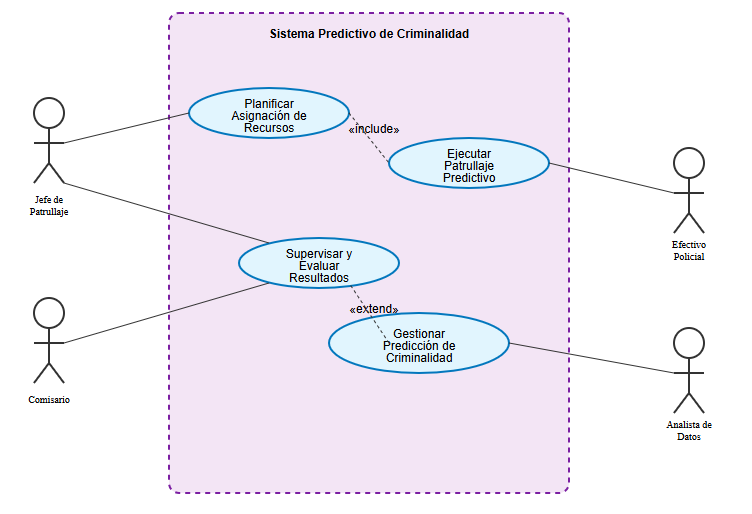
* Solo el jefe de patrullaje puede autorizar cambios en las rutas asignadas.
* Los resultados de patrullaje deben ser reportados al comisario dentro de las 24 horas.
* Las recomendaciones del sistema deben ser validadas por personal con rango mínimo de suboficial.

### 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*

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

**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.

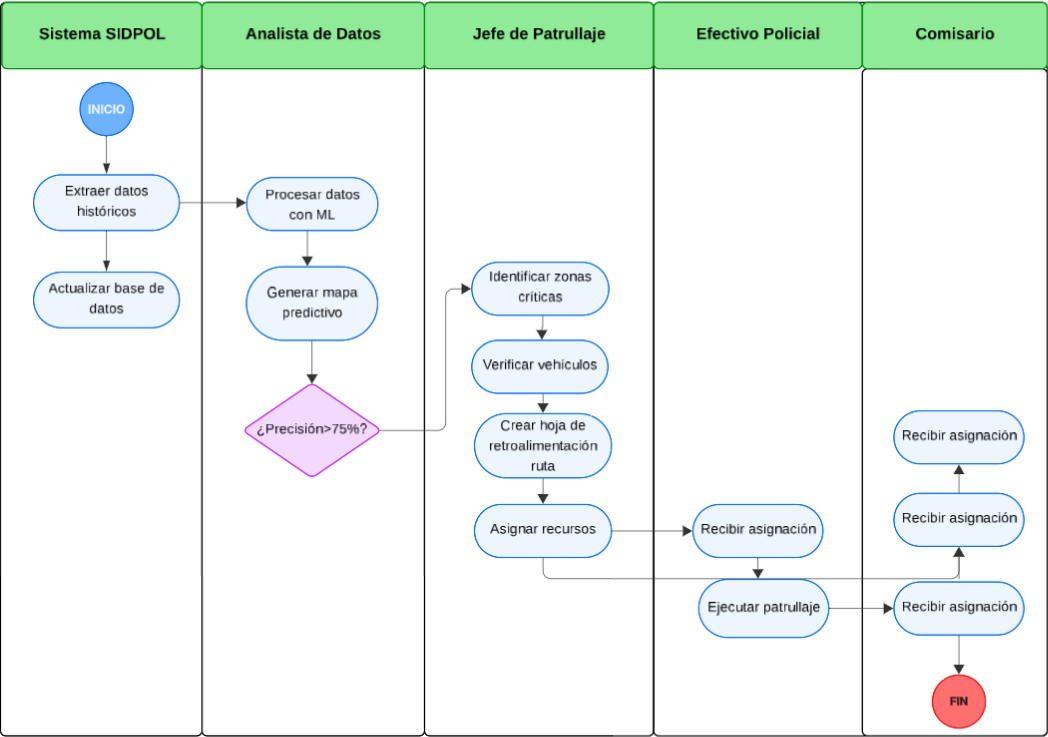
**Actores Secundarios**:

* Comisario: Supervisor general que recibe informes y validar 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**

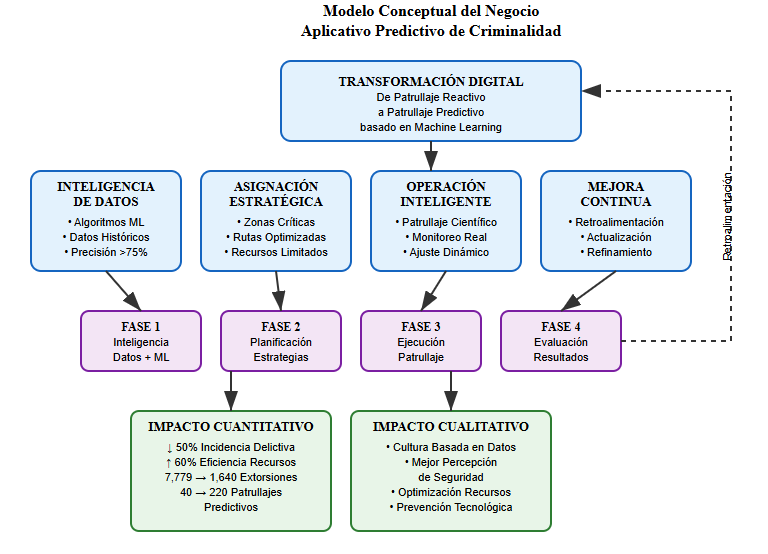
**Asignación Predictiva de Recursos Policiales**



***Figura 1.4.*** *Diagrama de actividades del negocio*

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

**4.3.3 Modelo Conceptual**



***Figura 1.5.*** *Modelo Conceptual*

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

# REFERENCIA BIBLIOGRÁFICA

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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