

# Revisión sistemática de la predicción y predicción de delitos

Juan C. Montaña<sup>1</sup>

<sup>1</sup> MINSC, Pontificia Universidad Javeriana,  
Colombia, Bogotá D.C.

<sup>1</sup>juanc\_montana@javeriana.edu.co

**Resumen.** Este artículo tiene el propósito de realizar una revisión sistemática a la literatura de la prevención y predicción de delitos. El objetivo de este artículo es realizar una revisión sistemática de la literatura en la prevención y predicción de delitos, para comprender y evaluar el estado del arte en cuanto conceptos y métodos dado el ritmo sin precedentes de los estudios empíricos publicados. Tiene muchas aplicaciones, como el pronóstico espacial del número de delitos, el tipo de actividad delictiva, la siguiente ubicación de un delito en una serie u otra información relacionada con el delito. Mediante este análisis se obtuvieron oportunidades de investigación que son presentadas en el apartado de trabajos futuros.

**Palabras Clave:** Prevención de delitos, predicción de delitos, inseguridad.

## 1 Introducción

Los datos sobre delitos se clasifican según el tipo de delito. Puede ser útil para identificar las áreas delictivas en función de la categoría del delito y muy útil para proteger las áreas delictivas para disminuir la tasa de delitos. Los crímenes están influenciados por organizaciones y otros lugares ocurren con frecuencia en una sociedad. El uso principal de analizar las áreas delictivas es identificar las áreas delictivas después de analizar los conglomerados, las ocurrencias delictivas con frecuencia durante diferentes años. La predicción de delitos se utiliza para identificar las áreas en las que se cometieron delitos y reducir esos delitos, esto puede ser muy útil para disminuir la tasa de delitos. Basado en los datos estructurados analizar los delitos.

El robo es un tipo de delitos contra la propiedad común en todo el mundo, que generalmente se refiere al acto de posesión ilegal de bienes públicos o de otras personas. En los últimos años, a pesar de las tasas de delincuencia generalmente a la baja, prevenir y combatir la delincuencia sigue siendo un desafío. Los gobiernos locales y la policía a menudo prestan más atención al asesinato, asalto y otros delitos violentos [1], por lo que se asignan recursos limitados para prevenir o luchar contra los delitos contra la propiedad. Por lo tanto, la tasa de delitos de robo parece mucho más alta que la de delitos violentos. Además, las estrategias de prevención eficaces contra nuevos tipos de robo (como el robo de vehículos eléctricos) son bastante inadecuadas. El análisis de riesgos cuantitativo y preciso es vital para prevenir robos cuando los recursos policiales son limitados. Además, existen fuertes correlaciones entre las tasas de criminalidad y otras variables como la ubicación geográfica de una comunidad (con áreas de bajo y alto riesgo), el tiempo y el clima durante el año (patrones estacionales) [2].

La IA puede estar involucrada en delitos de diversas formas. Más obviamente, la IA podría emplearse como una herramienta para el delito, haciendo uso de sus capacidades para facilitar acciones contra objetivos del mundo real: predecir el comportamiento de personas o instituciones para descubrir y explotar vulnerabilidades; generar contenido falso para usarlo en el chantaje o para mancillar la reputación; realizar hazañas que los perpetradores humanos no pueden o no quieren hacer por sí mismos, por razones de peligro, tamaño físico, velocidad de reacción, etc. Aunque los métodos son nuevos, los delitos en sí pueden ser de tipo tradicional: robo, extorsión, intimidación, terror [3].

La previsión de delitos se refiere al proceso básico de predecir delitos antes de que ocurran. Se necesitan herramientas para predecir un delito antes de que ocurra. Actualmente, la policía utiliza herramientas para ayudar en tareas específicas, como escuchar la llamada telefónica de un sospechoso o usar una cámara corporal para registrar alguna actividad ilegal inusual [4].

El objetivo de este artículo es responder la pregunta de investigación asociada a cómo generar recomendaciones basadas en reglas de asociación, aplicables y entendibles por los ciudadanos, que permitan mitigar el riesgo de ser víctimas de delitos de hurto. En este artículo se presenta una revisión sistemática de la literatura, que permita identificar oportunidades, conceptos, técnicas y métodos que se han utilizado para el análisis y predicción de delitos para responder la pregunta de investigación planteada

anteriormente. La estructura de este artículo consiste en presentar la metodología a utilizar, los resultados obtenidos, la discusión de los resultados encontrados y por último las conclusiones.

## 2 Metodología

Teniendo en cuenta la pregunta investigadora mencionada anteriormente, se realizó una revisión sistemática del estado del arte. Este proceso está compuesto de 3 fases. En la primera fase se crean las ecuaciones de búsqueda mediante palabras clave incluyendo AND o OR para diferentes conceptos o términos, para ser utilizadas en las bases de datos y obtener los artículos. Posteriormente, en la segunda fase se realiza una selección y filtrado de los artículos a partir de los criterios de selección. Finalmente, en la tercera fase una síntesis fue desarrollada para construir la base final de artículos.

### 2.1 Búsqueda de ecuaciones

Se realizó la construcción de ecuaciones de búsqueda a partir de términos y keywords de cada tema específico listados en la Tabla 1. Se utilizan operadores lógicos AND y OR para crear conjuntos disjuntos y no disjuntos, para que los artículos resultantes de la búsqueda coincidan con estas expresiones en su título, abstract o keywords. Se realizaron iteraciones sobre la ejecución de las ecuaciones en las distintas bases de datos para encontrar nuevos keywords que permitieran potenciar las ecuaciones de búsqueda y encontrar nuevos artículos.

**Tabla 1: Ecuaciones de búsqueda**

#	Topic	EQ
1	Crime prevention	predict crime OR prevent crime OR crime analysis OR crime forecasting
2	Artificial Intelligence	machine learning OR data mining OR clustering association rules OR fuzzy logic
3	Recommendations	recommender model OR recommender system OR recommendation methods
4	Risk	risk analysis OR high risk areas OR hotspots OR danger zone

### 2.2 Selección de artículos

A partir de los artículos resultantes de la anterior fase se realizó un filtro adicional para garantizar la novedad y actualidad. Por lo tanto se seleccionaron los artículos de los últimos 10 años (2010 – 2020). Finalmente se aplicaron los criterios de selección que se encuentra en la Tabla 2, basados en la novedad, facilidad de realizar experimentos, beneficio, extensibilidad, curva de aprendizaje y escalabilidad.

**Tabla 2: Criterios de selección**

#	Criteria	Description
1	Novedad	Es algo novedoso que no se ha visto antes o que se conoce hace poco tiempo y es muy llamativo
2	Experimentos	Probabilidad de que se tengan buenos resultados
3	Beneficio	Utilidad del producto o aporte que se podría dar a los grupos que vaya dirigido
4	Extensibilidad	Capacidad y facilidad de que pueda ser retomado por otro investigador para realizar un aporte o mejora
5	Curva de aprendizaje	Afinidad del investigador basado en sus conocimientos y habilidades, para el diseño y solución del problema
6	Escalabilidad	Capacidad de replicar la solución o investigación en contextos más grandes

Posteriormente a la aplicación de los criterios de selección los artículos resultantes fueron 15. Los cuales fueron utilizados para realizar la revisión sistemática de la literatura, extraer y analizar la información.

## 2.3 Análisis y estructura

Para evaluar y estudiar los resultados del contenido de los artículos seleccionados se clasifica en los siguientes tópicos (Tabla 3):

- Características: Información relacionada al espacio (área de estudio), tiempo (periodo de muestra y periodo en meses), datos de delito (tipo de delito y tamaño de muestra), factores y variables.
- Métodos y técnicas: Información relacionada a los métodos, técnicas y algoritmos empleados y comparados.
- Métricas y resultados: Métricas utilizadas en los experimentos para la evaluación, estrategia de validación y posibles trabajos futuros.

## 3 Resultados

En esta sección se presentan los resultados después de la RSL de los artículos resultantes que pueden ser vistos en una descripción general en la Tabla 3, la sección 3.1 discute acerca de las características y datos del delito, la sección 3.2 menciona los métodos y técnicas relacionadas a los análisis, pronósticos y predicciones y la sección 3.3 discute las métricas y resultados de los experimentos y comparaciones.

### 3.1 Características

En esta sección se discutirá acerca de las características de los artículos seleccionados. Para empezar se considera 5 características de información básica del estudio (Muestra, Área de estudio, Periodo de la muestra, Tipo de delito, Tamaño de la muestra y Factores más importantes). Un lector que desee modificar o replicar el enfoque metodológico presentado en la investigación, necesitará los mismos elementos de datos para evaluar si dicho enfoque es adecuado para el estudio de seguimiento y los objetivos de la investigación del autor.

El factor de división son las entradas de los modelos de predicción, es decir, las variables independientes. Hay algunos modelos de predicción que aceptan como entrada la latitud, la longitud y la marca de tiempo de los eventos delictivos (datos sin procesar), mientras que otros necesitan aplicar agregaciones o transformaciones explícitas antes de alimentar sus modelos. Feature engineering hace referencia al proceso de elaboración, escalado y selección de covariables o características para explicar mejor una variable de predicción que a menudo requiere un conocimiento considerable de un dominio específico.

Un ejemplo es la agregación de eventos delictivos en series espacio-temporales, que pueden descomponerse en rezagos autorregresivos y utilizarse como características. Este feature engineering también se puede aplicar a metodologías de datos auxiliares que no estén directamente relacionadas con el delito. Por ejemplo, en [5] cuentan puntos de referencia en la cuadrícula contando el número de elementos en cada celda (agregación espacial) y crean una nueva característica para cada tipo de punto de referencia, mientras que en [6] definen que una parte de su algoritmo es una capa de incrustación de región para procesar solo la ubicación sin procesar de los puntos de referencia de la ciudad.

El factor de división por entradas de método puede ser información útil para un investigador potencial que desee realizar pronósticos espaciales y consulte esta sección de nuestro artículo. El procesamiento de datos requiere conocimientos de dominio y es una tarea costosa (en el tiempo), especialmente cuando se trata de grandes conjuntos de datos espacio-temporales. Por lo tanto, puede ser preferible evitar el proceso de ingeniería de características algunos investigadores. Por otro lado, se puede preferir utilizar datos para derivar sus variables con patrones particulares. Llamamos a los métodos que tienen un enfoque interno para agregar eventos delictivos en variables espaciotemporales "independientes de la ingeniería de características" y dependiente de la ingeniería".

**Tabla 3: Descripción general de los 15 artículos seleccionados**

No	Authors	Space	Time	Crime Data			Forecasting				
		Study Area	Sampling period	Type	Sample	Most important Factors	Method	Inference	Task	Metric	Results
1	Zhang et al. (2019)	San Francisco, USA	Undefined	Theft Assault	720k	User Location	Recommendation model	Hotspots	Binary Classification	MAE	1408 1.645
2	Giménez et al. (2018)	Bogotá, Colombia	2012-2013	Homicide Assault Theft	Undefined	Banks Bus Stations Tourist Areas	RTM	Hotspots	Binary Classification	PAI	7.42 8.22 9.96
3	Cobos et al. (2020)	Colombia	1960-2019	Theft	350K	# year theft # city theft #department theft	SRV	# Of Crimes	Regression	RMSE	1795
4	Gelvez et al. (2021)	Bucaramanga, Colombia	2016-2019	Homicide Assault Theft	1865K	Latitude Longitude Month	TF-IDF	Hotspots	Binary Classification	Accuracy	0.5 - 0.6
5	Araujo Junior et al. (2017)	Natal, Brazil	2006-2016	Undefined	Undefined	Location Type Suspects Characteristics	Undefined	# Of Crimes	Regression	MSE RAS RPS	207.63
6	Hart and andbergen (2014)	Arlington, USA	2007-2008	AA, aggravated assault R, robbery CB, commercial burglary MVT, motor vehicle theft.	6295	Geocode point	KDE	Hotspots	Binary Classification	Hit rate PAI RRI	75% 92%
7	Hu et al. (2018)	Baton Rouge, USA	2011	Residential burglary	3706	Victim Arrestee Suspect Witness	SKDE	Hotspots	Binary Classification	PAI	3.91
8	Huang et al. (2018)	New York, USA	2014	Burglary Robbery Felony Assault Grand Larceny	103K	Crime reports Point of Interest Public-Service Compliants	Deep Neural Network	Category of crime	Binary Classification	Macro F1 Micro F1	Improve 4.5% and 6.1%
9	Kadar et al. (2017)	New York, USA	2011-2015	grandlarcenies robberies burglaries assaults	174K	Area Population Venues Checkins	Spatial Lag Linear Spatial Lag Linear Spatial Lag	# Of Crimes	Regression	R2	56%
10	Lin et al. (2018)	Taoyuan City, Taiwan	2015-2018	Motor Vehicle theft	8580	Undefined	Deep Learning Random Forest Naive Bayes	Hotspots	Binary Classification	Accuracy F1-Score Precision Recall	Undefined
11	Zhanjun et al. (2020)	Philadelphia, United States	2016	Unarmed robbery	1612	Longitude Latitude Timestamps	Near repeat crime	Hotspots	Binary Classification	Undefined	Undefined
12	Zhang et al. (2019)	Chicago and NSW, USA	2012-2017 2008-2012	365 types of crime 65 types of crime	1456713 Undefined	Crime District Month	Fuzzy association rules	Crime relations	Discover relations	N/A	N/A
13	Englin (2015)	Chicago, USA	2001-2014	All types of crime	4556343	Latitude Longitude Date District Wards Location Description	Indirect Association rules	Crime relations	Show significant relationships	N/A	N/A
14	Malik et al. (2014)	Tippecance, USA	2004-2014	All types of crime	310000	Geographic Location Offense type Agency Datetime	Seasonal Trend decomposition based on Loess (STL)	Hotspots	Binary Classification	Accuracy	95%
15	Mohler (2014)	Chicago, USA	2007-2012	Homicide Robbery Assault Weapons viol. Battery Sex	78852	Undefined	Marked point process model	Hotspots	Binary Classification	Accuracy	Improve 17%

En otras palabras, estos métodos necesitan explícitamente agregaciones para derivar variables espacio-temporales a partir de los datos brutos independientemente del algoritmo de pronóstico. La mayoría de los artículos reportados tienen una forma explícita de transformar sus eventos delictivos, así como los datos auxiliares, en características para alimentar sus algoritmos (es decir, dependientes de la ingeniería de

características). Aunque hemos encontrado muchas formas diferentes de agregación de datos en características, tanto espacial como temporalmente, el procedimiento de asignación de características a menudo se informa de manera insuficiente, lo que dificulta la reproducción de la metodología propuesta. Aún así, los flujos de trabajo o marcos bien definidos seguidos de métodos dependientes de la ingeniería de características se detallaron en [7] y [8]. Sintetizaron sus métodos de pronóstico en (1) agregar datos brutos espacialmente, siguiendo una metodología de mapeo del delito (por ejemplo, contando eventos dentro de las celdas de la cuadrícula), (2) generar series de tiempo y sus características, (3) ajustar un modelo de pronóstico usando un algoritmo y (4) visualizar los resultados.

En los métodos dependientes de la ingeniería de características, la agregación y la generación de series de tiempo se realizan por separado como pasos de procesamiento antes de ajustar un modelo, mientras que esto no es necesario para los métodos independientes de la ingeniería de características

En cuanto al área de estudio la mayoría de los artículos seleccionados están enfocados a casos de estudios de ciudades y mayormente en USA. Sin embargo en el contexto de Colombia se encontraron 3 artículos. Solamente el estudio de Zhang et al. en [6] utiliza 2 áreas de estudio, el resto de las investigaciones se basan en una sola área de estudio. Los datos más antiguos de crímenes son del estudio de Cobos et al. en [7] tomando datos de delitos en Colombia desde 1960 hasta 2019. Los periodos de muestra van desde 5 meses [9] hasta 1416.

En cuanto a los tipos de delito hay 2 artículos [7], [8] que estudian todos los tipos de delito, luego esta el delito de hurto y el delito de agresión como los más frecuentes. La muestra de datos más grande la tiene [8] con poco más de 4 millones de registros de delitos. En cuanto a los datos la gran mayoría están enfocados a un análisis geoespacial por tal razón es imprescindible la latitud y longitud del delito, sin embargo hay otros estudios como el de Giménez et al. [10] que considera un total de 17 factores de riesgo ambiental. Adicionalmente Hu et al. en [9] utiliza el API de google maps para codificar los datos y extraer nueva información complementaria para registros que no tenga información de la dirección en la que ocurrió el delito.

## Métodos y técnicas

En el contexto de Colombia, Giménez propuso un estudio utiliza el modelo de terreno de riesgo (RTM) [10] para identificar qué lugares tienen un riesgo mayor o menor de victimización por delitos violentos. Así, y con el objetivo de producir información procesable para mejorar la seguridad pública de Bogotá, este estudio ofrece un relato integral de los principales factores ambientales de vulnerabilidad y exposición al delito. Cobos et al. [7] presentaron un modelo de aprendizaje de máquina basado en Máquinas de Soporte Vectorial para la regresión (SRV) ajustado, para la predicción de la tendencia de hurtos en Colombia. Gelvez et al. [11] utilizó un modelo de procesamiento de señales para grafos y una adaptación del modelo de vectorización de texto TF-IDF al caso espacio temporal, se identificó que los mejores resultados de la predicción del delito se dieron al utilizar los modelos con relaciones espaciales de grafos por semanas.

KDE es un método bastante popular para identificar los puntos críticos (hotspots) de delito. Sin embargo, computando un KDE es muy pesado computacionalmente hablando cuando se trata de grandes conjuntos de datos y la calidad de la estimación resultante depende en gran medida de parámetros que son difíciles de configurar manualmente. Hart y Zanderberg propusieron un modelo de estimación de densidad del núcleo (KDE) [12] para examinar los efectos de la configuración de los parámetros definidos por el usuario (por ejemplo, el método de interpolación, el tamaño de la celda de la cuadrícula y el ancho de banda) en la precisión predictiva de los mapas de puntos críticos del delito. Se observa una variación al modelo simple de KDE, en forma de Espacio Temporal KDE (STKDE) presentada por Hu et al. en [9].

Aunque los algoritmos de aprendizaje profundo tienen la misma formulación que los algoritmos tradicionales de aprendizaje automático, presentan una estructura interna mucho más compleja que afecta su uso. La estructura de capa profunda hace que el presupuesto computacional sea principalmente necesario durante el entrenamiento. Además, la necesidad de muestras también es mayor que para los otros enfoques. Entre los artículos reportados, los tres que han utilizado este tipo de algoritmo argumentan que tiene el mejor desempeño general. Esto incluye las redes neuronales profundas (DNN) instaladas por Lin et al. [5] y el DeepCrime framework de Huang et al. [6], incluso presenta una arquitectura neuronal dedicada a un enfoque independiente de características, con una capa recurrente dedicada a codificar las dependencias temporales directamente de los sucesos delictivos. Aún así, ninguno de estos artículos ha discutido el rendimiento del tiempo computacional frente a otros algoritmos, ni tamaños de muestra suficientes para obtener modelos precisos. Se argumenta que no hay

una guía clara sobre cuándo se debe realizar un enfoque de redes neuronales profundas, aunque en los últimos años ha comenzado a surgir evidencia de su efectividad.

Un problema común en los métodos actuales de predicción de delitos es que, dadas las unidades temporales espaciales detalladas, los datos sobre delitos se volverían muy escasos y la predicción no funcionaría correctamente. Sin embargo, al modelar la predicción del delito como un problema de recomendación, se puede hacer uso de la abundante selección de métodos en los sistemas de recomendación que consideran inherentemente la escasez de datos. Zhang et al. [8] propuso modelar la predicción del delito como un problema de recomendación. Mediante este modelado, se puede tener una granularidad temporal espacial más fina, si se utiliza técnicas en sistemas de recomendación para mitigar el problema de la escasez de datos. Presento el método para modelar factores espaciales y temporales como usuarios y elementos, y discutí varias técnicas de recomendación que se pueden aplicar. Las técnicas incluyen métodos de filtrado colaborativo, así como métodos de predicción de calificación basados en el contexto.

Otro enfoque desde la perspectiva de la minería de datos es el de encontrar patrones en los datos de delitos como el presentado por Zhang et al. en donde utiliza la lógica difusa para establecer un una tasa de delitos alta, media o baja para posteriormente generar patrones frecuentes mediante el algoritmo FP-Growth y así generar las reglas de asociación difusas. Como resultado hay muchas reglas interesantes y sorprendentes, que merecen ser estudiadas más a fondo por expertos en el dominio. Un análisis similar presentado por Englin en [13]. para extraer relaciones significativas de datos públicos no limpios mediante el empleo y la extensión de la extracción de reglas de asociación indirecta en los atributos disponibles. También menciona que la minería de reglas de asociación indirecta para datos sobre delitos tiene el potencial de proporcionar relaciones interesantes entre los datos, pero requiere más manipulación de datos y reglas que las que se proporcionaron para este trabajo.

## 3.2 Métricas y resultados

Como se menciona en la selección de los artículos incluyen modelos novedosos o modelos conjuntos propuestos por autores de otros trabajos previos. Las métricas de evaluación de tales modelos son, en general, bien conocidas en criminología, ciencia GIS, matemáticas o estadística. Sin embargo, es importante mencionar que pocos autores destacan la necesidad de combinar o utilizar diversas métricas de evaluación. No podemos hacer una comparación de todos los resultados de la evaluación en los 15 artículos debido a varias razones, tales como diferentes unidades espaciales y temporales, áreas de estudio o métodos de predicción aplicados. Sin embargo, podemos discutir ciertas similitudes entre ellos.

La elección de una métrica de evaluación depende en gran medida del resultado principal de la predicción, como los recuentos (p. Ej., Para una distribución de Poisson), valores o tasas normalizados (p. Ej., Para una distribución gaussiana continua) o clasificación binaria (delito está ausente o presente). Las métricas de evaluación más frecuentes utilizadas en los artículos seleccionados son la Exactitud (Accuracy,  $n = 4$ ), seguida del Índice de Precisión de la Exhatitud (PAI,  $n = 3$ ), el Puntaje F1 ( $n = 2$ ) y luego se tiene la Precisión y Exhaustividad ( $n = 1$ ), el error cuadrático medio (MSE,  $n = 1$ ), el error cuadrático medio de la raíz (RMSE,  $n = 1$ ), R-cuadrado ( $n = 1$ ), el índice de tasa de recuperación (RRI,  $n = 1$ ). Algunas métricas adicionales se utilizan sólo una vez, a saber, la RMSE promedio (RMSE), la Puntuación de precisión de regresión (RAS), la Puntuación de precisión de regresión (RPS), la prueba, el error absoluto medio (MAE), el error de porcentaje absoluto medio (MAPE), macro-F1, micro-F1.

Analizamos las tres principales métricas de evaluación (PA, PAI, F1-Score) en relación con su distribución entre los elementos de datos de la disciplina, el tipo de algoritmo de pronóstico propuesto, la inferencia de pronóstico, la tarea de pronóstico, la unidad espacial y la unidad temporal. Curiosamente, encontramos que los informáticos utilizan exclusivamente el PA, mientras que los criminólogos prefieren aplicar el PAI. Además, mientras que el PA y el F1-Score se han probado preferiblemente para predicciones a corto plazo (es decir, menor o igual a 3 meses), el PAI se ha utilizado para predicciones tanto a corto como a largo plazo. No se detectó ningún otro patrón obvio entre los otros elementos de información con respecto al uso y preferencia de estas métricas de evaluación.

Para validar estos resultados la gran mayoría de los artículos seleccionados utilizan el 70% del dataset para entrenamiento y el 30% para pruebas. Algunos trabajos utilizan la estrategia de validación de comparar sus modelos con otros modelos existentes [8]. Huang et al. [6] que compararon su modelo de DNN con SVR, ARIMA, LR, MLP, TriMine, Wide&Deep y GRU. Por otro lado trabajos como [5], [14] realizan una comparación de sus propios modelos con varios casos por ejemplo la unidad inferencia, tipos de delitos y variación de hiperparámetros.

## 4 Discusión y resultados

La regresión se utiliza en gran medida para predecir el número de delitos y el desempeño se mide mediante varias medidas de error, como MSE o RMSE. En [14] Araujo et al. proponen dos nuevas métricas de evaluación, la Regression Precision Score (RAS), que representa el porcentaje de éxito en la predicción de una muestra, y la Regression Precision Score (RPS), que define la precisión de RAS. El RPS mide el MSE de las muestras de éxito normalizadas por la varianza de la muestra de entrenamiento. Sin embargo, el número de delitos no es la única inferencia considerada en los modelos de regresión. Otros autores predicen el porcentaje de delincuencia en conglomerados e infieren tasas de delincuencia a partir de datos históricos.

La relevancia de las métricas de evaluación puede estar sesgada por varias razones. Un ejemplo puede ser el desequilibrio de clases. Un modelo puede tener una gran precisión mientras predice muy bien las ubicaciones sin delitos. En contraste, las ubicaciones con delitos no están bien pronosticadas. Algunos autores intentan mejorar la relación negativo-positivo entre las células delictivas y no delictivas ajustando el peso de los puntos calientes y los puntos fríos [5]. Otra dependencia son los diferentes tipos de agregación que tienen lugar durante el modelado por atributos de tiempo, espacio o tipos de delitos. Por ejemplo, si bien la mayoría de los documentos informan trabajar con tipos de delitos desagregados, algunos de ellos consideran agregar tipos de delitos a, por ejemplo, “delitos violentos”, sin especificar qué tipos se incluyen.

Hay un enorme aumento en los estudios de predicción espacial del delito. Del conjunto de los 15 artículos seleccionados, se publicaron 11 en los últimos 5 años. Esto muestra el creciente interés de los académicos de diversas disciplinas en este tipo de investigación. El tipo de delito que más se ha estudiado es el robo residencial. No está claro por qué es más probable que se estudien este tipo de delito en particular y los delitos contra la propiedad, en general.

Existen variedades de oportunidades de investigación para responder la pregunta planteada en el presente artículo, algunos ejemplos de estas son listadas a continuación:

- Según Giménez, un modelo de predicción solamente presenta un diagnóstico y luego las autoridades se deben encargar de hallar la cura [10].
- La terminología utilizada suele ser muy especializada debido a que los trabajos son realizados por científicos que trabajan en áreas de criminología, ciencias de la computación, geo ciencias, etc. [15]. Lo cual hace que un lector que desee modificar o replicar la metodología utilizada en estos trabajos no pueda hacerlo de la mejor manera.
- Muchos de los detalles significativos en los trabajos no son reportados en su totalidad [15], tales como características de las muestras, detalles de los métodos utilizados y detalles de los experimentos realizados. Debido a esto la reproducibilidad y extensibilidad se ve afectada ya que otro investigador no podría utilizar estos trabajos como punto de comparación o extenderlos.
- En cuanto a datos e información muchos de los trabajos presentados se quedan cortos a nivel de información puesto que solamente utilizan datos relacionados al tiempo (Fecha y hora) y espacio (Latitud y Longitud) por tal razón surge la oportunidad de tener una base más sólida que permita tener resultados muchos más completos. Acorde a los modelos de minería de datos basados en reglas de asociación tiene el potencial de proporcionar relaciones interesantes entre los datos, pero requiere más manipulación de datos y reglas que las que se proporcionaron para este trabajo.
- En los trabajos relacionados a las reglas de asociación se menciona que habría sido interesante extender el algoritmo y limpiar el conjunto de datos en un conjunto que tenga atributos con un número reducido de valores posibles porque esto permitiría producir más reglas potenciales. Al agrupar descriptores de delitos comunes o descriptores de ubicación, el soporte para estos valores habría sido mayor, lo que habría resultado en más reglas con más de esos atributos [13], que merecen ser estudiadas por expertos.
- El enfoque de la predicción de delitos como un sistema de recomendaciones solamente se centra en generar un MAE que sea mejor al comparar con otros modelos, pero no se detalla nada relacionado a las recomendaciones lo cual abre una oportunidad interesante para resolver la pregunta de investigación planteada.
- En los trabajos en los cuales se utiliza reglas de asociación están limitados a solo presentar las reglas de asociación para su posterior análisis, pero no son utilizadas para realizar clasificación de delitos similar al enfoque presentado en [16] y tampoco para crear recomendaciones a partir de las reglas de asociación como en [17].

## 5 Conclusiones y trabajos futuros

Este trabajo se centro en conducir una revisión de la literatura concerniente a la predicción y prevención del delito. Se aplico una metodología para buscar artículos en diferentes bases de datos e identificar 15 artículos (Tabla 3) para su posterior análisis. Se analizaron los artículos por el características, métodos y métricas y evaluación. Se encontraron bastantes enfoques, la gran mayoría de ellos se centra en identificar los puntos calientes de delitos en particulares ciudades del mundo, otros a predecir el número de delitos y otros en un enfoque más de minería de datos que permita identificar relaciones y patrones interesantes.

Después de realizar el análisis se encontraron varias limitaciones que representa una motivación a realizar un trabajo futuro y responden la pregunta de investigación planteada. La primera limitación, según Giménez, es que un modelo de predicción solamente presenta un diagnóstico y luego las autoridades se deben encargar de hallar la cura [10]. La gran mayoría de autores presentan sus estudios para que más adelante otros individuos sean los encargados de utilizarlos para prevenir crímenes, generalmente las autoridades locales. Es allí donde se identifica la carencia de un enfoque orientado a los ciudadanos. La segunda limitación es la terminología como lo menciona Kounadi et al. en su revisión sistemática utilizando PRISMA [15], la cual puede dificultar la comprensión debido a que los científicos que trabajan este tema utilizan un lenguaje especializado (criminología, informática, geo ciencias, etc.). La tercera también mencionada por los mismos autores es que los detalles significativos de los experimentos no son reportados en su totalidad, como la unidad de análisis, el tamaño de la muestra, lo que hace que un estudio sea difícil de reproducir o comparar sus resultados con un posible estudio futuro.

La mayoría de los autores enfocan sus investigaciones en encontrar los puntos calientes o predecir los números de delitos. Solamente [6] se enfoca en clasificar el tipo de delito. Por otro lado se encontró un enfoque que modela la predicción del delito como un problema de recomendación, el cual solamente es comparado por la métrica MAE, pero no se muestra nada relacionado a las recomendaciones. En cuanto los análisis realizados mediante reglas de recomendación existen limitaciones en cuanto a la cantidad de datos, algoritmos utilizados y que solamente se centran en generar las reglas de asociación. Por lo tanto existe la oportunidad de ir más allá y utilizar estas reglas para 1) Construir un clasificador asociativo y 2) Crear recomendaciones para la mitigación y reducción de riesgos de ser víctima de delito de hurto.



## Referencias

- [1] M. Friedman, A. C. Grawert, and J. Cullen, "Crime trends: 1990-2016," pp. 1–36, 2017, [Online]. Available: <https://www.brennancenter.org/publication/crime-trends1990-2016>.
- [2] X. Hu, J. Wu, P. Chen, T. Sun, and D. Li, "Impact of climate variability and change on crime rates in Tangshan, China," *Sci. Total Environ.*, vol. 609, Dec. 2017, doi: 10.1016/j.scitotenv.2017.07.163.
- [3] M. Caldwell, J. T. A. Andrews, T. Tanay, and L. D. Griffin, "AI-enabled future crime," *Crime Sci.*, vol. 9, no. 1, Dec. 2020, doi: 10.1186/s40163-020-00123-8.
- [4] N. Shah, N. Bhagat, and M. Shah, "Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention," *Visual Computing for Industry, Biomedicine, and Art*, vol. 4, no. 1. 2021, doi: 10.1186/s42492-021-00075-z.
- [5] Y. L. Lin, T. Y. Chen, and L. C. Yu, "Using Machine Learning to Assist Crime Prevention," *Proc. - 2017 6th IIAI Int. Congr. Adv. Appl. Informatics, IIAI-AAI 2017*, pp. 1029–1030, 2017, doi: 10.1109/IIAI-AAI.2017.46.
- [6] C. Huang, J. Zhang, Z. Yu, and N. V. Chawla, "DeepCrime: Attentive hierarchical recurrent networks for crime prediction," in *International Conference on Information and Knowledge Management, Proceedings*, 2018, pp. 1423–1432, doi: 10.1145/3269206.3271793.
- [7] H. Ordóñez, C. Cobos, and V. Bucheli, "Machine learning model for predicting theft trends in Colombia," *RISTI - Rev. Iber. Sist. e Tecnol. Inf.*, vol. 2020, no. E29, pp. 494–506, 2020.
- [8] Y. Zhang, P. Siriaraya, Y. Kawai, and A. Jatowt, "Time and location recommendation for crime prevention," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11496 LNCS, pp. 47–62, doi: 10.1007/978-3-030-19274-7\_4.
- [9] Y. Hu, F. Wang, C. Guin, and H. Zhu, "A spatio-temporal kernel density estimation framework for predictive crime hotspot mapping and evaluation," *Appl. Geogr.*, vol. 99, no. August, pp. 89–97, 2018, doi: 10.1016/j.apgeog.2018.08.001.
- [10] A. Giménez-Santana, J. M. Caplan, and G. Drawve, "Risk Terrain Modeling and Socio-Economic Stratification: Identifying Risky Places for Violent Crime Victimization in Bogotá, Colombia," *Eur. J. Crim. Policy Res.*, vol. 24, no. 4, pp. 417–431, Dec. 2018, doi: 10.1007/s10610-018-9374-5.
- [11] J. D. Gélvez-ferreira, M. Paula, and N. Rodríguez, "Archivos de economía," 2021.
- [12] T. Hart and P. Zandbergen, "Kernel density estimation and hotspot mapping: Examining the influence of interpolation method, grid cell size, and bandwidth on crime forecasting," *Policing*, vol. 37, no. 2, pp. 305–323, 2014, doi: 10.1108/PIJPSM-04-2013-0039.
- [13] R. Englin, "Indirect association rule mining for crime data analysis," 2015.
- [14] A. A. Junior, N. Cacho, A. C. Thome, A. Medeiros, and J. Borges, "A predictive policing application to support patrol planning in smart cities," *2017 Int. Smart Cities Conf. ISC2 2017*, pp. 0–5, 2017, doi: 10.1109/ISC2.2017.8090817.
- [15] O. Kounadi, A. Ristea, A. Araujo, and M. Leitner, "A systematic review on spatial crime forecasting," *Crime Sci.*, vol. 9, no. 1, pp. 1–22, 2020, doi: 10.1186/s40163-020-00116-7.
- [16] A. Gharbi, C. De Runz, S. Faiz, and H. Akdag, "Towards association rules as a predictive tool for geospatial areas evolution," *GISTAM 2016 - Proc. 2nd Int. Conf. Geogr. Inf. Syst. Theory, Appl. Manag.*, no. Gistam, pp. 201–206, 2016, doi: 10.5220/0005914202010206.
- [17] T. Osadchiy, I. Poliakov, P. Olivier, M. Rowland, and E. Foster, "Recommender system based on pairwise association rules," *Expert Syst. Appl.*, vol. 115, no. September, pp. 535–542, 2019, doi: 10.1016/j.eswa.2018.07.077.