

**UNIVERSIDAD TÉCNICA DE MANABÍ**  
**FACULTAD DE CIENCIAS BÁSICAS**  
**INGENIERÍA ESTADÍSTICA**

**ANTEPROYECTO**

Miguel Alberto Almeida Loor  
Vicente Valentín Vera Carranza

Portoviejo – Ecuador  
2026

## Índice general

<b>2. Marco teórico</b>	<b>2</b>
2.1. Antecedentes de investigación . . . . .	2
2.1.1. Predicción espaciotemporal del delito: evolución, tendencias y resultados empíricos . . . . .	2
2.1.2. Aprendizaje automático y aprendizaje profundo en predicción del delito: aportes y limitaciones . . . . .	3
2.1.3. Modelos probabilísticos para conteos con dependencia temporal: avances, alcance y brechas . . . . .	3
2.1.4. Modelos probabilísticos espaciotemporales para crimen y el lugar de SPINGARCH . . . . .	4
2.1.5. Vacíos identificados y relación directa con el anteproyecto . . . . .	5
2.2. Referentes teóricos . . . . .	6
2.2.1. Delito como proceso espaciotemporal discreto: criminología del lugar y consecuencias estadísticas . . . . .	6
2.2.2. Modelamiento probabilístico de conteos dinámicos: Poisson condicional, sobredispersión e INGARCH . . . . .	6
2.2.3. Extensión espaciotemporal: matrices de pesos y estructura SPINGARCH	7
2.2.4. Aprendizaje supervisado y gradiente boosting: fundamentos de XGBoost	8
2.2.5. Motivación teórica del enfoque híbrido SPINGARCH–XGBoost . . . .	8
2.2.6. Interpretabilidad, explicabilidad y consideraciones éticas en modelos predictivos del delito . . . . .	9
2.3. Definiciones conceptuales . . . . .	10
2.3.1. Conceptos clave y definición operacional preliminar . . . . .	10
2.3.2. Cierre del marco teórico . . . . .	13

## Índice de figuras

## Índice de cuadros

**Modelamiento predictivo espaciotemporal del delito en Guayaquil mediante un enfoque  
híbrido basado en SPINGARCH y XGBoost**

## 2. Marco teórico

Este marco teórico organiza, sistematiza y discute la evidencia académica y los fundamentos conceptuales necesarios para sostener el anteproyecto *Modelamiento predictivo espaciotemporal del delito en Guayaquil mediante un enfoque híbrido basado en SPINGARCH y XGBoost*. El objetivo no es reproducir la literatura de forma extensa, sino seleccionar lo estrictamente pertinente para: (i) identificar avances metodológicos en predicción del delito, (ii) delimitar vacíos científicos aplicables al caso de estudio y (iii) sustentar la lógica del enfoque híbrido propuesto desde una base probabilística e interpretativa.

### 2.1. Antecedentes de investigación

#### 2.1.1. Predicción espaciotemporal del delito: evolución, tendencias y resultados empíricos

La investigación contemporánea coincide en que el delito urbano no se distribuye de manera homogénea ni aleatoria, sino que tiende a concentrarse en una fracción pequeña de microterritorios y a persistir en el tiempo. Esta regularidad se formaliza en la criminología del lugar mediante la denominada *ley de concentración del crimen*, que documenta la estabilidad de los focos delictivos y justifica el análisis territorializado de la seguridad urbana [1]. La implicación metodológica de este hallazgo es directa: un modelo predictivo que ignore la estructura espacial y temporal del delito arriesga subestimar la heterogeneidad de riesgo y, en consecuencia, producir predicciones poco útiles para la planificación preventiva.

En la última década, la literatura ha transitado desde el mapeo descriptivo de *hotspots* hacia esquemas predictivos que buscan anticipar el riesgo futuro en unidades espaciales finas (cuadrículas, barrios o sectores administrativos) y en periodos temporales discretos (días, semanas o meses). Una síntesis robusta de esta transición se recoge en la revisión sistemática de Kounadi *et al.* [2], donde se reporta que las aproximaciones dominantes incluyen modelos estadísticos (regresiones, series temporales, modelos Bayesianos), métodos espaciales (suavizamientos y kernel), así como técnicas de aprendizaje automático. Un patrón recurrente en los trabajos revisados es que la precisión predictiva suele mejorar al incorporar explícitamente dependencia temporal (memoria del proceso) y dependencia espacial (interacción vecinal), aunque persisten dificultades para la generalización fuera de muestra y para mantener interpretabilidad cuando aumenta la complejidad del modelo.

Un aspecto metodológico menos visible, pero relevante para comparabilidad y validez externa, es el tratamiento de la exposición o *población en riesgo*. En particular, los estimadores del denominador (población residencial vs. población “ambiental” o flotante) pueden alterar sustantivamente tasas y predicciones cuando la ciudad presenta movilidad intraurbana intensa. Khalfa y Hardyns [3] muestran que diferentes aproximaciones de población en riesgo influyen en las

tasas delictivas y en los resultados predictivos, lo cual sugiere que, aun cuando el estudio se formule en términos de conteos, debe justificarse conceptualmente la unidad espacial y el esquema de exposición utilizado en el análisis.

### **2.1.2. Aprendizaje automático y aprendizaje profundo en predicción del delito: aportes y limitaciones**

El auge de grandes volúmenes de datos urbanos y la disponibilidad de infraestructura computacional han impulsado el uso de modelos de aprendizaje automático (ML) y aprendizaje profundo (DL) para predicción del delito. Una revisión sistemática reciente [4] evidencia que técnicas como bosques aleatorios, gradiente boosting, redes recurrentes y modelos basados en grafos se han aplicado con resultados prometedores, especialmente cuando se integran fuentes de datos heterogéneas (variables socioeconómicas, movilidad, clima, uso de suelo y variables temporales). No obstante, la misma literatura advierte que el desempeño depende críticamente del preprocesamiento, la resolución espacial, el esquema de validación (por ejemplo, validación temporal estricta) y la estabilidad del fenómeno bajo cambios estructurales.

Dentro de los enfoques de DL, destacan los modelos que representan la ciudad como una red de relaciones espaciales y funcionales. Por ejemplo, Hou *et al.* [12] proponen un modelo basado en grafos con mecanismos de atención para capturar relaciones complejas entre regiones y su evolución temporal. Este tipo de métodos suele mejorar la predicción en escenarios de alta interacción espacial; sin embargo, su implementación exige mayor cantidad de datos, mayor ajuste de hiperparámetros y, con frecuencia, reduce la trazabilidad de las decisiones del modelo. En un problema socialmente sensible como la seguridad, la ganancia marginal de precisión no necesariamente compensa la pérdida de interpretabilidad si el resultado se orienta a uso público o institucional.

La literatura también muestra evidencia empírica específica en contextos urbanos latinoamericanos. Escobedo *et al.* [5] comparan modelos de regresión para predecir crimen contra la propiedad en distritos de Lima con alta incidencia, resaltando que la elección del modelo afecta tanto la precisión como la estabilidad. En este tipo de escenarios, caracterizados por ruido en los registros, subregistro potencial y variaciones abruptas, los modelos flexibles tienden a capturar patrones no lineales, pero requieren mecanismos de control de sobreajuste y evaluación robusta.

### **2.1.3. Modelos probabilísticos para conteos con dependencia temporal: avances, alcance y brechas**

Desde la perspectiva estadística, el delito agregado por unidad espacial y tiempo se modela naturalmente como una variable discreta de conteo. En este marco, los modelos de series tem-

porales para conteos constituyen un cuerpo teórico central. La revisión sistemática de Liu *et al.* [8] sintetiza el desarrollo de los modelos INGARCH (Integer-valued Generalized Autoregressive Conditional Heteroskedasticity), resaltando tres aportes relevantes: (i) la especificación de una distribución condicional discreta, (ii) la modelación explícita de autocorrelación a través de la intensidad condicional y (iii) la capacidad de inducir sobredispersión incondicional aun con distribución condicional Poisson. Estas propiedades los vuelven especialmente atractivos para series de crimen donde la varianza observada excede la media y donde existe dependencia temporal.

Los avances recientes extienden estos modelos hacia estructuras multivariadas y espaciotemporales. Jahn *et al.* [9] proponen aproximaciones lineales para INGARCH en conteos espaciotemporales, con el propósito de mejorar factibilidad computacional y permitir análisis a mayor escala espacial. Por su parte, Maletz *et al.* [10] desarrollan un marco de autorregresión de conteos espaciotemporal, orientado a capturar dependencia cruzada y estructura temporal de manera integrada. Aunque estos desarrollos consolidan una base probabilística para la predicción, una limitación frecuente es que la relación entre covariables y respuesta se introduce mediante formas funcionales relativamente rígidas (a menudo aditivas y lineales), lo cual puede ser insuficiente para capturar umbrales, interacciones de alto orden o efectos no lineales asociados a condiciones urbanas complejas.

#### **2.1.4. Modelos probabilísticos espaciotemporales para crimen y el lugar de SPINGARCH**

En el dominio de crimen urbano, existen enfoques probabilísticos alternativos que modelan eventos en espacio y tiempo, tales como modelos Bayesianos espaciotemporales y procesos puntuales. Por ejemplo, Hu *et al.* [6] presentan un modelo Bayesiano espaciotemporal para predicción de crimen, destacando la representación conjunta de efectos espaciales y temporales y la cuantificación explícita de incertidumbre. Complementariamente, Shirota y Gelfand [7] proponen procesos de Cox log-Gaussianos con tiempo circular para datos de eventos criminales, lo que permite flexibilidad en la intensidad del proceso y captura de estacionalidad. Estos enfoques son conceptualmente ricos, pero pueden requerir mayor complejidad inferencial y depender fuertemente de supuestos de estructura latente.

Como alternativa parsimoniosa e interpretable en el contexto de conteos agregados, los modelos SPINGARCH (Spatial INGARCH) incorporan explícitamente interacción espacial mediante matrices de pesos, manteniendo una estructura de intensidad condicional con interpretación probabilística. Escudero *et al.* [11] desarrollan un modelo espacialmente correlacionado con estructura tipo GARCH para conteos de crimen, mostrando que la dependencia territorial pue-



de integrarse sin abandonar la naturaleza discreta de la variable respuesta. Aun así, persiste un desafío metodológico: la estructura paramétrica, aunque interpretable, puede no capturar adecuadamente patrones residuales no lineales vinculados a covariables contextuales múltiples o a cambios estructurales.

### 2.1.5. Vacíos identificados y relación directa con el anteproyecto

A partir de la síntesis anterior, se identifican cuatro vacíos metodológicos relevantes:

- **Separación entre rigor probabilístico y flexibilidad predictiva:** los modelos probabilísticos para conteos (INGARCH/SPINGARCH y variantes) modelan dependencia e incertidumbre, pero suelen imponer relaciones funcionales limitadas [8, 11]. En contraste, muchos métodos de ML capturan no linealidades, pero no especifican explícitamente una distribución generadora del conteo ni preservan interpretabilidad probabilística [4].
- **Dependencia espaciotemporal como requisito, no como accesorio:** buena parte de trabajos basados en ML incorporan la componente espacial como variables derivadas o embeddings, sin garantizar una estructura de dependencia coherente (por ejemplo, estacionariedad o condiciones de estabilidad) [12].
- **Escasez de evidencia híbrida para conteos delictivos con énfasis interpretativo:** aunque existen aproximaciones híbridas en otros dominios, aún es limitada la evidencia que combine, de forma explícita, modelos espaciotemporales de conteo con algoritmos de boosting dentro de un diseño reproducible y evaluable.
- **Necesidad de explicabilidad y uso responsable:** en problemas de alto impacto social, la “caja negra” es metodológicamente riesgosa; por ello, se requieren enfoques interpretables o explicables que permitan auditar resultados y discutir implicaciones [15].

En consecuencia, el anteproyecto se posiciona en el cruce entre estos vacíos: propone un enfoque híbrido que preserve una base probabilística espaciotemporal (SPINGARCH) y, simultáneamente, incorpore flexibilidad no paramétrica mediante XGBoost, complementado con técnicas de explicabilidad para sostener un uso público responsable.

## 2.2. Referentes teóricos

### 2.2.1. Delito como proceso espaciotemporal discreto: criminología del lugar y consecuencias estadísticas

El punto de partida conceptual del estudio es tratar el delito urbano como un proceso espaciotemporal discreto: en un territorio dividido en unidades espaciales  $i = 1, \dots, N$  y en periodos de tiempo  $t = 1, \dots, T$ , se observan conteos  $Y_{i,t}$  que representan la cantidad de eventos delictivos en la región  $i$  durante el periodo  $t$ . Esta conceptualización es consistente con la evidencia empírica sobre concentración territorial del crimen y persistencia temporal de *hotspots* [1]. En términos estadísticos, ello implica que la dependencia (autocorrelación temporal) y la interacción territorial (dependencia espacial) no son anomalías, sino propiedades estructurales del fenómeno.

Adicionalmente, el concepto de “riesgo” delictivo se relaciona con la exposición: un mismo conteo puede implicar riesgos diferentes si cambia la población o la actividad en la zona. En ciudades con alta movilidad, la población residencial puede no representar adecuadamente la población efectivamente expuesta en cada zona y periodo, afectando tasas, comparaciones y predicciones [3]. Por ello, incluso cuando el foco del modelamiento sea el conteo, el diseño conceptual debe explicitar la unidad espacial y temporal, y justificar qué interpretación del riesgo se pretende (conteo esperado, tasa por exposición o intensidad).

### 2.2.2. Modelamiento probabilístico de conteos dinámicos: Poisson condicional, sobredispersión e INGARCH

La familia de modelos INGARCH se basa en especificar una distribución condicional discreta para el conteo en función de la información pasada. Un caso base es el modelo Poisson condicional:

$$Y_t \mid \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t), \quad (2.1)$$

donde  $\mathcal{F}_{t-1}$  denota la información disponible hasta  $t - 1$  y  $\lambda_t = \mathbb{E}(Y_t \mid \mathcal{F}_{t-1})$  es la intensidad condicional. La dinámica temporal se introduce mediante una ecuación autorregresiva de la intensidad, por ejemplo en un INGARCH( $p, q$ ):

$$\lambda_t = \omega + \sum_{k=1}^p \alpha_k Y_{t-k} + \sum_{\ell=1}^q \beta_\ell \lambda_{t-\ell}, \quad (2.2)$$

con  $\omega > 0$  y  $\alpha_k, \beta_\ell \geq 0$ . Esta formulación separa explícitamente: (i) efectos de corto plazo vinculados a eventos recientes (a través de  $Y_{t-k}$ ) y (ii) memoria estructural de la intensidad (a

través de  $\lambda_{t-\ell}$ ).

Un resultado clave es que, aun con distribución condicional Poisson, la estructura dinámica induce sobredispersión incondicional, propiedad observada frecuentemente en series delictivas [8]. La misma literatura señala extensiones para sobredispersión severa, como versiones binomiales negativas, que incrementan flexibilidad al desacoplar varianza condicional y media. En el anteproyecto, estos fundamentos son relevantes porque justifican el uso de modelos de intensidad condicional como base estocástica para el conteo delictivo, preservando interpretación probabilística y permitiendo predicción *fuera de muestra* bajo una dinámica explícita.

### 2.2.3. Extensión espaciotemporal: matrices de pesos y estructura SPINGARCH

Para datos urbanos territoriales, es necesario generalizar el enfoque anterior hacia un proceso multivariado de conteos. Sea  $Y_{i,t}$  el número de delitos observados en la región  $i$  durante el periodo  $t$ . Bajo un marco SPINGARCH, se especifica una distribución condicional discreta para cada región, típicamente:

$$Y_{i,t} \mid \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_{i,t}), \quad (2.3)$$

donde  $\lambda_{i,t}$  es la intensidad condicional espaciotemporal. La interacción territorial se incorpora mediante una matriz de pesos espaciales  $W = (w_{ij})$  que codifica vecindad o proximidad (contigüidad, distancia o  $k$ -vecinos). Un esquema parsimonioso de dinámica puede expresarse como:

$$\lambda_{i,t} = \omega + \alpha Y_{i,t-1} + \beta \lambda_{i,t-1} + \gamma \sum_{j=1}^N w_{ij} Y_{j,t-1}. \quad (2.4)$$

En esta ecuación,  $\alpha$  captura persistencia temporal local (efecto de corto plazo),  $\beta$  modela inercia o memoria de la intensidad, y  $\gamma$  cuantifica contagio espacial: el conteo en regiones vecinas influye en el riesgo actual de la región  $i$ . Este componente es crucial para describir desplazamiento de focos delictivos y difusión territorial de la actividad criminal. La separación entre efectos temporales y espaciales, además, facilita interpretación y discusión con teorías del lugar (por ejemplo, estabilidad de hotspots) [1].

En el marco del anteproyecto, SPINGARCH se adopta como pilar estadístico por dos razones: (i) respeta la naturaleza discreta del delito agregado y (ii) modela explícitamente dependencia espaciotemporal de forma interpretable. Sin embargo, la propia evidencia empírica aplicada a conteos de crimen muestra que, aun con estructura espacial, pueden persistir patrones no lineales asociados a covariables urbanas complejas [11]. Esta limitación constituye el punto de entrada para incorporar métodos algorítmicos complementarios.

#### 2.2.4. Aprendizaje supervisado y gradiente boosting: fundamentos de XGBoost

Desde una perspectiva estadística, el aprendizaje supervisado busca aproximar una función desconocida  $f(\cdot)$  que relaciona covariables  $x$  con una respuesta  $y$ , minimizando un riesgo empírico regularizado. En el caso de gradiente boosting, la función se construye como un ensamble aditivo:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad (2.5)$$

donde cada  $f_k$  pertenece a una clase funcional restringida (por ejemplo, árboles de decisión). El algoritmo ajusta iterativamente nuevos componentes para corregir errores residuales, interpretándose como un descenso por gradiente funcional sobre una función de pérdida [13].

XGBoost formaliza este enfoque mediante una función objetivo regularizada:

$$\mathcal{L} = \sum_{i=1}^n \ell(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k), \quad (2.6)$$

donde  $\ell(\cdot)$  mide error predictivo y  $\Omega(\cdot)$  penaliza complejidad del ensamble (por ejemplo, número de hojas y norma de pesos en árboles). La regularización explícita contribuye a controlar sobreajuste y estabilizar el desempeño *fuera de muestra*, lo cual resulta valioso cuando los registros criminales presentan ruido y colas pesadas.

En dominios espaciales, se han propuesto variantes que buscan capturar heterogeneidad local. Por ejemplo, Grekousis [14] presenta una extensión denominada *Geographical-XGBoost*, enfocada en regresión espacialmente local. En el contexto del crimen, estas ideas sugieren que la relación entre covariables y riesgo puede variar territorialmente, motivando la inclusión de variables espaciales (o mecanismos de segmentación) en el componente ML del enfoque híbrido.

#### 2.2.5. Motivación teórica del enfoque híbrido SPINGARCH–XGBoost

El núcleo conceptual del enfoque híbrido es que el delito observado combina: (i) una estructura estocástica espaciotemporal (dependencia y sobredispersión) y (ii) componentes funcionales complejos asociados a covariables contextuales (no linealidades, interacciones, umbrales). Los modelos SPINGARCH están diseñados para el primer componente, mientras que XGBoost es un aproximador potente para el segundo. La integración se justifica, entonces, como una complementariedad entre un *modelo generativo* (con distribución explícita y parámetros interpretables) y un *modelo predictivo* (con alta capacidad funcional).

En términos operativos, la hibridación puede interpretarse como un esquema en el cual el SPINGARCH captura la dinámica base del conteo (memoria temporal y contagio espacial), y el com-

ponente XGBoost se entrena para explicar variación adicional asociada a covariables o a patrones residuales no capturados por la estructura paramétrica. Esta separación modular permite: (a) evaluar el aporte marginal de cada componente, (b) preservar interpretabilidad probabilística de la dinámica y (c) mejorar precisión al incorporar flexibilidad no paramétrica.

Además, el uso de modelos basados en grafos en predicción delictiva [12] refuerza la idea de que la interacción espacial real es compleja y puede incluir relaciones funcionales no lineales. Sin embargo, en lugar de reemplazar la estructura probabilística por un modelo opaco, el enfoque híbrido mantiene una base estocástica explícita y utiliza ML como complemento, reduciendo el riesgo de sobreinterpretación algorítmica y facilitando auditoría metodológica.

#### 2.2.6. Interpretabilidad, explicabilidad y consideraciones éticas en modelos predictivos del delito

Los modelos predictivos del delito se ubican en un dominio de alto impacto social: los resultados pueden influir en la asignación territorial de recursos, en la priorización de intervenciones y en percepciones públicas sobre comunidades. En este marco, la interpretabilidad no es un adorno metodológico, sino una condición para uso responsable. Rudin [15] argumenta que, en decisiones de alto riesgo, la preferencia por modelos interpretables puede ser superior a “explicar” cajas negras, especialmente cuando las decisiones deben ser auditables y justificables.

No obstante, en escenarios donde se utilizan modelos complejos por necesidad predictiva, es posible incorporar técnicas de explicabilidad (XAI). Fan *et al.* [16] ejemplifican el uso de aprendizaje automático interpretable en el análisis espacial de relación entre detenciones policiales y crimen callejero, mostrando que la interpretabilidad puede integrarse en la evaluación sustantiva del fenómeno y no solo en la presentación de resultados.

En el componente XGBoost del anteproyecto, la explicabilidad se abordará mediante valores SHAP (Shapley Additive exPlanations), los cuales se basan en teoría de juegos cooperativos para asignar a cada covariable una contribución aditiva a la predicción [17]. En su forma general, una predicción se descompone como:

$$f(x) = \phi_0 + \sum_{j=1}^p \phi_j, \quad (2.7)$$

donde  $\phi_0$  es el valor base (predicción esperada) y  $\phi_j$  representa el aporte de la covariable  $j$ . Esta descomposición favorece explicaciones locales (por observación) y globales (por patrón promedio), permitiendo discutir qué factores aumentan o disminuyen el riesgo estimado en determinadas zonas y periodos.

Finalmente, el marco teórico asume que la utilidad del modelo es *analítica* y no determinista:

las predicciones deben interpretarse como estimaciones probabilísticas sujetas a incertidumbre y no como justificación automática de intervenciones. Esto implica reportar resultados agregados, transparentar supuestos y reconocer potenciales sesgos en los datos (por ejemplo, sesgos de registro), alineando el estudio con un enfoque institucional responsable y verificable.

## 2.3. Definiciones conceptuales

Esta sección consolida los conceptos esenciales del estudio y propone definiciones operacionales preliminares (susceptibles de ajuste según la disponibilidad y calidad de los datos del cantón Guayaquil).

### 2.3.1. Conceptos clave y definición operacional preliminar

Concepto	Definición conceptual y operacional preliminar
Delito urbano (evento)	Hecho delictivo registrado por una institución competente en un espacio y tiempo determinados. En este estudio se trabaja con registros agregados (conteos) por unidad territorial y periodo, sin inferencia individual.
Conteo delictivo $Y_{i,t}$	Variable aleatoria discreta que representa el número de eventos delictivos observados en la unidad espacial $i$ durante el periodo $t$ . Operacionalmente, $Y_{i,t}$ se construirá mediante agregación de incidentes georreferenciados (o asignados territorialmente) a una partición espacial definida.
Unidad espacial $i$	Partición territorial utilizada para discretizar el espacio urbano (por ejemplo, barrios, sectores o una cuadrícula regular). Operacionalmente, la unidad será definida de acuerdo con: (i) disponibilidad de datos, (ii) coherencia con gestión pública y (iii) estabilidad geométrica (evitar unidades demasiado pequeñas con exceso de ceros o demasiado grandes que oculten concentración).
Periodo temporal $t$	Intervalo de tiempo discreto para agregación (día, semana o mes). Operacionalmente, se seleccionará un periodo que balancee granularidad (captura de dinámica) y estabilidad estadística (evitar series excesivamente esporádicas).

Concepto	Definición conceptual y operacional preliminar
Hotspot espaciotemporal	Zona y periodo con intensidad delictiva relativamente alta respecto al promedio urbano. Conceptualmente se asocia con la concentración del crimen y su persistencia [1]. Operacionalmente puede identificarse mediante percentiles del conteo esperado o del riesgo estimado.
Dependencia temporal	Propiedad por la cual la ocurrencia pasada de eventos afecta la ocurrencia futura. En series de conteo, se modela mediante términos autorregresivos en la intensidad condicional (por ejemplo, $\alpha Y_{t-1}$ y $\beta \lambda_{t-1}$ en INGARCH/SPINGARCH) [8].
Dependencia espacial	Propiedad por la cual la actividad delictiva en una zona se asocia con la actividad de zonas vecinas (por proximidad y movilidad). En SPINGARCH se incorpora mediante una matriz de pesos $W$ y un término de contagio espacial $\gamma \sum_j w_{ij} Y_{j,t-1}$ [11].
Matriz de pesos espaciales $W = (w_{ij})$	Estructura que codifica vecindad o conexión territorial entre unidades espaciales. Operacionalmente, $W$ puede construirse por contigüidad (adyacencia), distancia (umbral) o $k$ vecinos más cercanos; además, puede normalizarse por filas para interpretar $\sum_j w_{ij} Y_{j,t-1}$ como un promedio ponderado vecinal.
Intensidad condicional $\lambda_{i,t}$	Media condicional del conteo: $\lambda_{i,t} = \mathbb{E}(Y_{i,t} \mid \mathcal{F}_{t-1})$ . Conceptualmente representa el nivel esperado de delito dada la historia. Operacionalmente se estima mediante el ajuste del modelo y se usa como base para predicción [8, 11].
Sobredispersión	Situación donde $\text{Var}(Y_{i,t}) > \mathbb{E}(Y_{i,t})$ . En datos delictivos es frecuente por heterogeneidad territorial y cambios estructurales. Los modelos INGARCH/SPINGARCH pueden inducir sobredispersión incondicional aun con distribución condicional Poisson [8].
Modelo INGARCH	Modelo de series temporales para conteos que define una distribución condicional discreta y una dinámica autorregresiva para la intensidad. Operacionalmente, se ajusta a series de conteos para capturar memoria y producir pronósticos [8].

Concepto	Definición conceptual y operacional preliminar
Modelo SPINGARCH	Extensión espaciotemporal de INGARCH que incorpora interacción territorial mediante $W$ . Operacionalmente, modela $Y_{i,t}$ como conteos con intensidad dinámica dependiente del historial local y vecinal, permitiendo predicción por zona y periodo [11].
XGBoost	Algoritmo de gradiente boosting basado en árboles, que optimiza una función de pérdida regularizada para producir predicciones robustas [13]. En este estudio se utilizará para capturar relaciones no lineales entre covariables y el componente delictivo que no sea explicado adecuadamente por la estructura paramétrica.
Covariables $x_{i,t}$	Conjunto de variables explicativas asociadas a la unidad $i$ en el tiempo $t$ (por ejemplo, variables temporales, espaciales y contextuales). Operacionalmente, se construirán a partir de fuentes disponibles y se evaluará su contribución mediante XAI.
Enfoque híbrido	Estrategia metodológica que combina un modelo probabilístico (SPINGARCH) y un modelo de aprendizaje automático (XGBoost) en un esquema modular. Operacionalmente, se evaluará el aporte de cada componente y su desempeño conjunto frente a alternativas.
Interpretabilidad	Capacidad de explicar el comportamiento del modelo de manera trazable y auditable. En el componente probabilístico se logra mediante parámetros con significado (memoria temporal y contagio espacial). En el componente ML se apoya en técnicas de explicabilidad [15].
SHAP ( $\phi_j$ )	Valores de contribución aditiva derivados de teoría de juegos que asignan a cada covariable su aporte promedio marginal a la predicción [17]. Operacionalmente, se utilizarán para explicar predicciones de XGBoost a nivel local y global.
Desempeño predictivo	Grado en que las predicciones se aproximan a los valores observados fuera de muestra. Operacionalmente, se evaluará con métricas de error (p. ej., MAE y RMSE) y con validación temporal/espaciotemporal coherente.



Concepto	Definición conceptual y operacional preliminar
Uso responsable	Principio según el cual las predicciones se interpretan como soporte analítico y no como decisiones automáticas. Operacionalmente implica: reporte agregado, transparencia de supuestos, comunicación clara y reconocimiento de incertidumbre.

### 2.3.2. Cierre del marco teórico

En conjunto, el marco teórico muestra que la predicción del delito requiere respetar la naturaleza discreta del fenómeno, incorporar dependencia temporal y espacial, y balancear precisión con interpretabilidad. La evidencia revisada sustenta el valor de modelos de conteo dinámicos (INGARCH/SPINGARCH) para capturar el proceso generador, y reconoce el potencial de métodos de ML como XGBoost para modelar relaciones no lineales. El vacío metodológico se ubica en la integración coherente de ambos mundos bajo criterios de explicabilidad y uso público responsable; este es precisamente el aporte que el anteproyecto busca evaluar en el contexto de Guayaquil.

## Bibliografía

- [1] D. Weisburd, “The law of crime concentration and the criminology of place,” *Criminology*, vol. 53, no. 2, pp. 133–157, 2015. Available: <https://doi.org/10.1111/1745-9125.12070>
- [2] O. Kounadi, A. Ristea, A. Araujo Jr., and M. Leitner, “A systematic review on spatial crime forecasting,” *Crime Science*, vol. 9, p. 7, 2020. Available: <https://doi.org/10.1186/s40163-020-00116-7>
- [3] R. Khalfa and W. Hardyns, “On how population-at-risk estimates influence crime rates and predictions: Comparing residential and ambient-like estimates,” *Applied Geography*, vol. 185, 2025. Available: <https://doi.org/10.1016/j.apgeog.2025.103780>
- [4] V. Mandalapu, L. Elluri, P. Vyas, and N. Roy, “Crime prediction using machine learning and deep learning: A systematic review and future directions,” *IEEE Access*, vol. 11, pp. 60153–60170, 2023. Available: <https://doi.org/10.1109/ACCESS.2023.3286344>
- [5] M. Escobedo, C. Tapia, J. Gutierrez, and V. Ayma, “Comparing regression models to predict property crime in high-risk Lima districts,” *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 3, pp. 62–68, 2024.
- [6] T. Hu, X. Zhu, L. Duan, and W. Guo, “Urban crime prediction based on spatio-temporal Bayesian model,” *PLOS ONE*, vol. 13, no. 10, p. e0206215, 2018. Available: <https://doi.org/10.1371/journal.pone.0206215>
- [7] S. Shirota and A. E. Gelfand, “Space and circular time log Gaussian Cox processes with application to crime event data,” *Annals of Applied Statistics*, vol. 11, no. 2, pp. 481–507, 2017.
- [8] M. Liu, F. Zhu, J. Li, and C. Sun, “A systematic review of INGARCH models for integer-valued time series,” *Entropy*, vol. 25, no. 6, p. 922, 2023. Available: <https://doi.org/10.3390/e25060922>
- [9] M. Jahn, C. H. Weiß, and H.-Y. Kim, “Approximately linear INGARCH models for spatio-temporal counts,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 72, no. 2, pp. 476–497, 2023. Available: <https://doi.org/10.1093/jrssc/qlad018>

- [10] S. Maletz, K. Fokianos, and R. Fried, “Spatio-temporal count autoregression,” *Data Science in Science*, vol. 3, no. 1, p. 2425171, 2024. Available: <https://doi.org/10.1080/26941899.2024.2425171>
- [11] I. Escudero, J. M. Angulo, and J. Mateu, “A spatially correlated model with generalized autoregressive conditionally heteroskedastic structure for counts of crimes,” *Entropy*, vol. 24, no. 7, p. 892, 2022. Available: <https://doi.org/10.3390/e24070892>
- [12] M. Hou, X. Hu, J. Cai, X. Han, and S. Yuan, “An integrated graph model for spatial-temporal urban crime prediction based on attention mechanism,” *ISPRS International Journal of Geo-Information*, vol. 11, no. 5, p. 294, 2022. Available: <https://doi.org/10.3390/ijgi11050294>
- [13] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, 2016, pp. 785–794. Available: <https://doi.org/10.1145/2939672.2939785>
- [14] G. Grekousis, “Geographical-XGBoost: A new ensemble model for spatially local regression based on gradient-boosted trees,” *Journal of Geographical Systems*, vol. 27, pp. 169–195, 2025. Available: <https://doi.org/10.1007/s10109-025-00465-4>
- [15] C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” *Nature Machine Intelligence*, vol. 1, pp. 206–215, 2019. Available: <https://doi.org/10.1038/s42256-019-0048-x>
- [16] Z. Fan, G. Song, J. Long, L. Cai, and J. Chen, “Exploring the spatial relationship between police stops and street crime based on interpretable machine learning,” *Geographical Research*, vol. 43, no. 11, pp. 3072–3087, 2024. Available: <https://doi.org/10.11821/dlyj020240051>
- [17] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems 30*, 2017, pp. 4765–4774.