

Maestría en

Ciencia de Datos y Máquinas de Aprendizaje con mención Inteligencia Artificial

Trabajo previo a la obtención de título de Magister en Ciencia de Datos y Máquinas de Aprendizaje con mención Inteligencia Artificial

AUTORES:

Felipe Esteban Gallardo Delgado
Marlon Alexander Mayorga Maroto
Imelda Alexandra Vizcaino Imacaña

TUTORES:

Karla Estefanía Mora Cajas
Juan Fernando López Tito

TEMA

Análisis espacio-temporal de la seguridad ciudadana (robo, homicidios, violencia y delitos) en las provincias de Guayas, Pichincha y Manabí mediante modelos predictivos de aprendizaje automático durante el periodo enero 2022 a diciembre 2025.

Certificación de autoría

Nosotros, Felipe Esteban Gallardo Delgado, Marlon Alexander Mayorga Maroto, Imelda Alexandra Vizcaino Imacaña, declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido presentado anteriormente para ningún grado o calificación profesional y que se ha consultado la bibliografía detallada.

Cedemos nuestros derechos de propiedad intelectual a la Universidad Internacional del Ecuador (UIDE), para que sea publicado y divulgado en internet, según lo establecido en la Ley de Propiedad Intelectual, su reglamento y demás disposiciones legales.

**Felipe
Gallardo**

Firmado digitalmente
por Felipe Gallardo
Fecha: 2026.07.25
17:34:16 -05'00'

Firma

Felipe Esteban Gallardo Delgado

**Marlon
Mayorga**

Firmado digitalmente
por Marlon Mayorga
Fecha: 2026.07.26
09:03:12 -05'00'

Firma

Marlon Alexander Mayorga Maroto

**Imelda Alexandra
Vizcaino Imacaña**

Firmado digitalmente por
Imelda Alexandra Vizcaino
Imacaña
Fecha: 2026.07.25 17:23:25
-05'00'

Firma

Imelda Alexandra Vizcaino Imacaña

Autorización de Derechos de Propiedad Intelectual

Nosotros, Felipe Esteban Gallardo Delgado, Marlon Alexander Mayorga Maroto, Imelda Alexandra Vizcaino Imacaña en calidad de autores del trabajo de investigación titulado “Análisis espacio-temporal de la seguridad ciudadana (robo, homicidios, violencia y delitos) en las provincias de Guayas, Pichincha y Manabí mediante modelos predictivos de aprendizaje automático durante el periodo enero 2022 a diciembre 2025”, autorizamos a la Universidad Internacional del Ecuador (UIDE) para hacer uso de todos los contenidos que nos pertenecen o de parte de los que contiene esta obra, con fines estrictamente académicos o de investigación. Los derechos que como autores nos corresponden, lo establecido en los artículos 5, 6, 8, 19 y demás pertinentes de la Ley de Propiedad Intelectual y su Reglamento en Ecuador.

D. M. Quito, julio de 2026

**Felipe
Gallardo**

Firmado digitalmente
por Felipe Gallardo
Fecha: 2026.07.25
17:34:44 -05'00'

Firma

Felipe Esteban Gallardo Delgado

**Marlon
Mayorga**

Firmado digitalmente
por Marlon Mayorga
Fecha: 2026.07.26
09:03:45 -05'00'

Firma

Marlon Alexander Mayorga Maroto

**Imelda Alexandra
Vizcaino Imacaña**

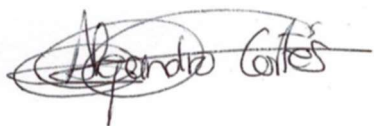
Firmado digitalmente por
Imelda Alexandra Vizcaino
Imacaña
Fecha: 2026.07.25 17:23:49
-05'00'

Firma

Imelda Alexandra Vizcaino Imacaña

Aprobación de dirección y coordinación del programa

Nosotros, **Director Alejandro Cortés López de EIG y Coordinadora Karla Estefanía Mora Cajas de UIDE**, declaramos que: **Felipe Esteban Gallardo Delgado, Marlon Alexander Mayorga Maroto, Imelda Alexandra Vizcaino Imacaña** son los autores exclusivos de la presente investigación y que ésta es original, auténtica y personal de ellos.



Alejandro Cortés López
Director/a de la
Maestría en Ciencia de datos y máquinas
de aprendizaje con mención en Inteligencia
Artificial



Karla Estefanía Mora Cajas
Coordinador/a de la
Maestría en Ciencia de datos y máquinas
de aprendizaje con mención en Inteligencia
Artificial

DEDICATORIA

Con profunda gratitud dedico este trabajo a mis padres, Sandra y Mauricio, cuyo apoyo incondicional ha sido fundamental para alcanzar mis metas profesionales y seguir persiguiendo mis sueños. Gracias por estar siempre presentes, por ayudarme a levantarme en los momentos difíciles y por demostrarme que, a pesar de la distancia, su amor y preocupación permanecen constantes.

A mis hermanos, Bryan y Marjury; sobrino Isaac, a quienes admiro profundamente por su fortaleza, perseverancia y capacidad de salir adelante con los recursos que tienen. Aunque no mantengamos una comunicación diaria, siempre he sentido su apoyo y cariño, los cuales me han impulsado a continuar luchando por mis metas.

A mi niño interior, que soñó con una vida estable pero llena de aventuras y que siempre, aunque fuera muy difícil se esforzó para lograr su sueño y alcanzar su meta.

Marlon Alexander Mayorga Maroto

A mis hijos Cristopher y Tomas quienes son mi fuente de inspiración en el desarrollo en mi vida profesional y académica. A mi compañero Pedro, quien me ha apoyado en el desarrollo de esta investigación. A mi familia que, acompañado con amor, paciencia mi vida académica. A todas las mujeres en las ciencias informáticas que encuentre espacios donde puedan ejercer su liderazgo y demostrar su capacidad.

Imelda Alexandra Vizcaino Imacaña

Dedico este trabajo, en primer lugar, a mi hijo, quien es mi mayor inspiración, mi fuerza y la principal motivación para continuar creciendo tanto en el ámbito personal como profesional.

Cada logro alcanzado lleva consigo el deseo de brindarle un mejor ejemplo de esfuerzo, perseverancia y superación.

A mis padres, por su amor incondicional, apoyo constante y por haber sido los pilares fundamentales en mi formación. Gracias por enseñarme el valor del trabajo, la disciplina y la humildad, principios que me han acompañado en cada etapa de mi vida.

Con profundo orgullo y gratitud, culmino este importante objetivo académico, consciente de que representa solo un paso más dentro de un camino de aprendizaje y crecimiento continuo. A ustedes, que han estado presentes en cada desafío y cada logro, les dedico este esfuerzo y les agradezco por creer siempre en mí.

Felipe Esteban Gallardo Delgado

AGRADECIMIENTOS

Agradezco profundamente a mi familia, quienes, a pesar de la distancia, siempre han confiado en que alcanzaré mis metas y han hecho de mis logros un motivo de orgullo. Gracias por ser una parte esencial de mi vida, por su amor incondicional y por mantenernos unidos como familia, incluso frente a las dificultades.

A mis amigos Alejandro y Jhonathan, con quienes he compartido una amistad sincera durante casi quince años. A pesar de la distancia, hemos permanecido juntos en los buenos y malos momentos, siempre creyendo en mis capacidades. Les agradezco por brindarme el ánimo necesario para seguir adelante y por celebrar conmigo cada logro alcanzado a lo largo de esta etapa.

A mis amigos de esta nueva etapa, con una mención especial para Mildreth, Juan, Jorge, David, Domenica y Sofía les agradezco profundamente por hacer que la transición de vivir en un nuevo país haya sido más llevadera.

Marlon Alexander Mayorga Maroto

Quiero expresar mi gratitud infinita a las personas e instituciones que hicieron posible la realización de este trabajo académico: a mis colegas de proyecto Marlon y Felipe con sus conocimientos, paciencia, dedicación y trabajo en equipo hemos concluido este proyecto. Al equipo de Universidad Internacional del Ecuador, por saber orientarme oportunamente en el desarrollo de la maestría.

Imelda Alexandra Vizcaino Imacaña

Quiero expresar mi más sincero agradecimiento a Dios por brindarme la fortaleza, la sabiduría y la perseverancia necesarias para culminar esta importante etapa de mi formación académica y profesional.

A mis padres, por su apoyo incondicional, sus enseñanzas y los valores que me han inculcado a lo largo de mi vida. Gracias por ser mi ejemplo de esfuerzo, responsabilidad y dedicación, y por acompañarme en cada uno de mis proyectos y metas.

A mi hijo Emilio y a mi esposa Michelle, por ser la mayor motivación de mi vida y el impulso para seguir creciendo personal y profesionalmente. Su amor, comprensión y apoyo incondicional me brindaron la fortaleza necesaria para afrontar cada desafío y culminar esta etapa con la certeza de que el esfuerzo y la perseverancia siempre tienen su recompensa.

Felipe Esteban Gallardo Delgado

RESUMEN

La investigación analiza la violencia y la criminalidad en Ecuador; se trata de un fenómeno relacionado con varios factores que, durante los últimos años, ha mostrado un aumento y ha afectado la seguridad ciudadana. El objetivo fue estudiar su evolución mediante la aplicación y comparación de modelos estadísticos, algoritmos de aprendizaje automático y técnicas de aprendizaje profundo; con ello se buscó generar información útil para apoyar la toma de decisiones y la planificación de políticas públicas de seguridad.

Para desarrollar el estudio se recopilaron datos del ECU 911, el Ministerio del Interior y la Fiscalía General del Estado, correspondientes al periodo comprendido entre enero de 2022 y diciembre de 2025. Las bases incluyeron registros diarios y mensuales, organizados por provincia, cantón, parroquia, categoría y subcategoría del delito; algunas fuentes también contenían información sobre sexo, género y etnia. A estos datos se añadieron indicadores del INEC relacionados con población, pobreza multidimensional, educación y empleo; después, toda la información fue depurada, integrada y normalizada para reducir errores y mantener una estructura común.

La comparación incluyó modelos como ARIMA y Exponential Smoothing, utilizados como referencias iniciales; también se probaron XGBoost y Temporal Fusion Transformer. Los resultados variaron según la variable estudiada, la escala territorial y el periodo de predicción; por esta razón, no corresponde presentar un modelo como superior en todos los casos, pues cada uno responde mejor a determinados tipos de datos y objetivos.

Los resultados muestran que los modelos predictivos pueden apoyar el análisis de la criminalidad cuando se utilizan junto con información territorial y social; las variables

socioeconómicas ayudaron a ampliar la representación de las provincias, aunque su aporte específico deberá revisarse con comparaciones más detalladas. La herramienta desarrollada puede servir como apoyo para organizar recursos, revisar tendencias y orientar políticas públicas; aun así, sus estimaciones deben acompañarse de análisis técnico y no asumirse como decisiones automáticas.

Palabras clave: robos, homicidios, violencia, criminalidad, Temporal Fusion Transformer, XGBoost, ARIMA, Exponential Smoothing.

ABSTRACT

This research analyzes violence and criminality in Ecuador a phenomenon linked to various factors that has intensified in recent years, impacting public security. The study aimed to examine its evolution by applying and comparing statistical models, machine learning algorithms, and deep learning techniques, thereby generating useful information to support decision-making and the planning of public security policies.

To conduct the study, data were collected from ECU 911, the Ministry of the Interior, and the Attorney General's Office, covering the period from January 2022 to December 2025. The datasets included daily and monthly records organized by province, canton, parish, and crime category and subcategory; some sources also contained information on sex, gender, and ethnicity. Indicators from the INEC regarding population, multidimensional poverty, education, and employment were incorporated into this data; subsequently, all information was cleaned, integrated, and standardized to minimize errors and maintain a common structure.

The comparison included models such as ARIMA and Exponential Smoothing, used as initial benchmarks; XGBoost and Temporal Fusion Transformer were also evaluated. The results varied depending on the variable studied, the spatial scale, and the prediction period; therefore, it is not appropriate to designate a single model as superior in all cases, as each performs better with specific data types and objectives.

The results demonstrate that predictive models can support crime analysis when used in conjunction with spatial and social data; socioeconomic variables helped broaden the representation of provinces, although their specific contribution requires further evaluation through more detailed comparisons. The developed tool can facilitate resource allocation, trend

analysis, and public policy guidance; however, its estimates should be accompanied by technical analysis rather than being considered a basis for automated decision-making.

Keywords: robberies, homicides, violence, criminality, Temporal Fusion Transformer, XGBoost, ARIMA, Exponential Smoothing.

CONTENIDO

CAPÍTULO 1: INTRODUCCIÓN AL ALCANCE DE PROYECTO	1
1.1 Introducción	1
1.1.1 Definición del proyecto.....	4
1.2 Justificación e importancia del trabajo de investigación.....	6
1.3 Alcance	9
1.4 Objetivos	11
1.4.1 Objetivo general	11
1.4.2 Objetivos específicos	11
CAPÍTULO 2: ESTADO DEL ARTE Y MARCO TEÓRICO	13
2.1 Estado del Arte	13
2.1.1 Estudios sobre predicción de criminalidad mediante Machine Learning.....	13
2.1.2 Aplicación de Deep Learning en el análisis criminal	15
2.1.3 Estudios relacionados con criminalidad en América Latina y el Ecuador	16
2.2 Marco Teórico	19
2.2.1 Criminalidad y Seguridad Ciudadana.....	19
2.2.2 Ciencia de Datos aplicada al análisis criminal	20
2.2.3 Análisis Espacio-Temporal de la Criminalidad	22
2.2.4 Modelos ARIMA, ExpSmooth, XGBoost y Temporal Fusion Transformer	23
2.2.5 Calidad de Datos y Preprocesamiento, IA en la seguridad ciudadana.....	24
CAPÍTULO 3: DESARROLLO DEL PROYECTO.....	26
3.1 Recopilación de datos.....	28
3.2 Selección de los conjuntos de datos	29
3.3 Análisis exploratorio de datos	30
3.4 Preprocesamiento de los datos	31
3.5 Análisis de modelo y visualización	32
3.6 Evolución y versiones del desarrollo	33
3.7 Modelos predictivos implementados.....	34
3.7.1 Preparación de datos para el modelamiento	34
3.7.2 Variables objetivo y variables derivadas.....	36
3.7.3 Comparación metodológica de modelos.....	37
3.7.4 Configuración de XGBoost	38

3.7.5 Configuración de ARIMA/SARIMAX.....	39
3.7.6 Configuración de Temporal Fusion Transformer	40
3.7.7 Índice de riesgo provincial y clustering.....	42
3.7.8 Consideraciones para la interpretación.....	43
CAPÍTULO 4: ANÁLISIS DE RESULTADOS	45
4.1 Aplicación de la metodología propuesta	46
4.1.1. Población y Muestreo	46
4.2 Cobertura de datos y unidad de análisis	49
4.3 Resultados del análisis descriptivo espacio-temporal	49
4.3.1 Tendencia de homicidios	49
4.3.2 Tendencia de Robos.....	51
4.3.3 Lectura integrada de las fuentes	54
4.4 Evaluación de los modelos predictivos	56
4.5 Proyección para 2026.....	58
4.6 Análisis de riesgo provincial	59
4.7 Limitaciones y consideraciones para la interpretación	60
4.8 Síntesis de resultados	61
CAPÍTULO 5: CONCLUSIONES Y RECOMENDACIONES.....	64
5.1 Conclusiones	64
5.2 Recomendaciones.....	65
REFERENCIAS BIBLIOGRÁFICAS.....	67
APÉNDICES.....	71
Apéndice A.....	71
Enlace al aplicativo	71
Apéndice B.....	71
Documentación técnica	71

ÍNDICE DE TABLAS

TABLA 1 ARQUITECTURA DEL SISTEMA	27
TABLA 2 ARCHIVOS GENERADOS PARA EL MODELAMIENTO PREDICTIVO	35
TABLA 3 VARIABLES OBJETIVO Y TRANSFORMACIONES UTILIZADAS	36
TABLA 4 RESUMEN COMPARATIVO DE CONFIGURACIÓN DE MODELOS	38
TABLA 5 CONFIGURACIÓN TÉCNICA DE XGBOOST	39
TABLA 6 CONFIGURACIÓN TÉCNICA DE ARIMA/SARIMAX.....	40
TABLA 7 CONFIGURACIÓN TÉCNICA DEL TEMPORAL FUSION TRANSFORMER	41
TABLA 8 CONFIGURACIÓN DEL ÍNDICE DE RIESGO	42
TABLA 9 HOMICIDIOS 2023 AL 2025 POR PROVINCIA.....	50
TABLA 10 TOTALES ACUMULADOS DEL PANEL INTEGRADO, 2022 – 2025	55
TABLA 11 XGBOOST PROMEDIOS DE VALIDACIONES 2023, 2024 Y 2025	56
TABLA 12 TFT Y ARIMA EN HOMICIDIOS, DESAPARECIDOS Y ARMAS	57
TABLA 13 XGBOOST PREDICCIONES 2026	59

ÍNDICE DE FIGURAS

FIGURA 1 <i>ARQUITECTURA FUNCIONAL DE LA SOLUCIÓN ANALÍTICA.</i>	29
FIGURA 2 <i>FLUJO DE PROCESAMIENTO Y ANÁLISIS DE DATOS.</i>	30
FIGURA 3 <i>FLUJO DE INTEGRACIÓN Y CONSTRUCCIÓN DE DATASETS.</i>	34
FIGURA 4 <i>UNIDAD DE ANÁLISIS Y SALIDA DE CADA MODELO</i>	37
FIGURA 5 <i>CÁLCULO DE ÍNDICE DE RIESGO PROVINCIAL</i>	42
FIGURA 6 <i>EVOLUCIÓN ANUAL DE HOMICIDIOS EN GUAYAS, MANABÍ Y PICHINCHA 2022 – 2025</i>	46
FIGURA 7 <i>TASA DE HOMICIDIOS EN GUAYAS, MANABÍ Y PICHINCHA 2025</i>	48
FIGURA 8 <i>PROVINCIAS CON MAYOR NÚMERO DE HOMICIDIOS, 2025</i>	50
FIGURA 9 <i>ROBOS POR PROVINCIA</i>	51
FIGURA 10 <i>TIPOS DE ROBOS MÁS FRECUENTES</i>	52
FIGURA 11 <i>ANÁLISIS MENSUAL POR TIPO DE ROBO</i>	53
FIGURA 12 <i>DISTRIBUCIÓN MENSUAL DE HOMICIDIOS POR PROVINCIA, 2022 – 2025</i>	58
FIGURA 13 <i>COMPOSICIÓN DE ARMAS INCAUTADAS POR PROVINCIA, 2022 – 2025</i>	60
FIGURA 14 <i>PREDICCIÓN DE HOMICIDIOS PARA 2026 CON ESCENARIOS XGBOOST</i>	62
FIGURA 15 <i>CLUSTERING TERRITORIAL</i>	63

CAPÍTULO 1: INTRODUCCIÓN AL ALCANCE DE PROYECTO

Este capítulo presenta las bases conceptuales y metodológicas de la investigación, junto con una revisión general de la seguridad ciudadana en el Ecuador; primero se explica el problema relacionado con los patrones espaciales y temporales de los delitos de alto impacto¹ y otros hechos vinculados con la violencia, utilizando registros institucionales de la plataforma de datos abiertos.

También se explica la importancia académica y social del estudio, así como el uso de herramientas de ciencia de datos e inteligencia artificial para generar predicciones; el análisis se concentra en Guayas, Pichincha y Manabí. Luego se presentan los objetivos, la delimitación temporal y las decisiones metodológicas que orientan el desarrollo del modelo predictivo.

1.1 Introducción

La seguridad ciudadana representa una de las áreas de mayor atención dentro de la gestión pública ecuatoriana; el aumento de los delitos de alto impacto, las emergencias y los eventos vinculados con la violencia ha mostrado una distribución territorial desigual, por ello, los registros generados por el ECU 911, el Ministerio del Interior y la Fiscalía General del Estado pueden utilizarse para examinar estos fenómenos mediante procedimientos técnicos de integración, depuración y análisis sistemático; bajo esta consideración, la analítica de datos facilita el reconocimiento de patrones temporales, diferencias provinciales y variaciones que podrían pasar inadvertidas en una revisión únicamente descriptiva, pues el estudio articulado de las dimensiones

¹ Delitos de alto impacto corresponden a homicidios, asesinatos, secuestros, extorsiones, robos con violencia y robo de vehículos, tráfico ilícito de drogas, delincuencia organizada y actividades relacionadas con organizaciones criminales. Estos debido a su gravedad, su incidencia y el impacto que generan en la seguridad ciudadana.

espaciales y temporales favorece la planificación de respuestas institucionales ante incidentes ocurridos en zonas urbanas y metropolitanas (Gómez y Ramírez, 2023).

El problema investigado surgió por la escasa articulación entre las fuentes institucionales relacionadas con la seguridad ciudadana; aunque las entidades públicas producen registros sobre robos, homicidios, denuncias de delitos, armas, detenciones y desapariciones, estos datos permanecían distribuidos en bases con estructuras distintas, niveles desiguales de detalle y unidades geográficas o temporales poco homogéneas; tal dispersión limita la identificación oportuna de zonas con mayor exposición delictiva, dificulta las comparaciones provinciales y reduce las posibilidades de aplicar modelos predictivos orientados a respaldar decisiones sustentadas en evidencia; además, la distribución espacial de los delitos podría guardar relación con la densidad poblacional, la pobreza multidimensional, el desempleo, la desigualdad económica y el crecimiento urbano acelerado (Chainey y Ratcliffe, 2015; Medina, 2010).

La delimitación territorial comprendió las provincias de Guayas, Pichincha y Manabí; su selección respondió a su peso poblacional, su actividad institucional y su incidencia dentro de los registros nacionales de delitos de alto impacto; sin embargo, el componente predictivo incorporó un panel territorial más amplio, conformado por varias bases previamente procesadas, con el propósito de exponer los algoritmos a una mayor diversidad geográfica y temporal; pese a esta ampliación, la interpretación de los resultados se concentró en las tres provincias establecidas como unidades principales de análisis.

La dimensión temporal abarcó los años 2022 a 2025; los análisis descriptivos y comparativos utilizaron registros correspondientes a ese periodo, siempre que las bases presentaran niveles adecuados de integridad y consistencia; por otra parte, los valores

correspondientes a 2026 fueron tratados como proyecciones exploratorias obtenidas mediante los modelos predictivos, razón por la cual no representan registros observados completos; esta distinción mantuvo correspondencia entre la delimitación del estudio, los objetivos planteados, los procedimientos aplicados y los resultados presentados.

La investigación se sustentó en aportes procedentes de la criminología ambiental, el análisis espacial del delito y la ciencia de datos aplicada a fenómenos sociales; la actividad delictiva no suele distribuirse de forma uniforme, sino que puede concentrarse en determinados lugares y periodos debido a condiciones urbanas, sociales, económicas y operativas (Chainey y Ratcliffe, 2005; Medina, 2010); asimismo, la combinación de variables delictivas, territoriales y socioeconómicas ofrece una interpretación más amplia de las variaciones asociadas con la inseguridad, sin atribuir los resultados a una sola causa.

El estudio siguió una orientación cuantitativa y aplicada; se trabajó mediante un diseño no experimental y longitudinal, acompañado de alcances exploratorio, descriptivo y predictivo; la fase exploratoria reunió fuentes heterogéneas e identificó comportamientos iniciales en los registros, mientras que el componente descriptivo examinó tendencias, concentraciones y diferencias entre provincias; a su vez, el componente predictivo empleó series temporales y algoritmos de aprendizaje automático para estimar posibles escenarios futuros; la evaluación correspondió a problemas de regresión y utilizó métricas como el error absoluto medio MAE, la raíz del error cuadrático medio RMSE, el coeficiente de determinación R^2 y el error porcentual absoluto medio MAPE, sin incorporar indicadores propios de clasificación cuando la variable de salida era continua.

La propuesta analítica integró depuración de bases, análisis exploratorio, visualización interactiva y modelamiento predictivo; su propósito no consistió en sustituir el criterio profesional de las instituciones responsables de la seguridad pública, sino en proporcionar un recurso complementario para examinar tendencias, establecer prioridades territoriales y respaldar la discusión de políticas públicas con información verificable; por tanto, el estudio aporta insumos para la gestión de la seguridad ciudadana mediante evidencia empírica, procedimientos metodológicos trazables y representaciones visuales comprensibles.

1.1.1 Definición del proyecto

En el Ecuador se registró un aumento de hechos vinculados con la inseguridad ciudadana, entre ellos robos, homicidios y distintos actos de violencia; frente a esta situación, el Sistema Integrado de Seguridad ECU 911, el Ministerio del Interior y la Fiscalía General del Estado intervienen en la atención de emergencias, la investigación de delitos y la protección interna del país.

Los registros generados por estas instituciones contienen datos sobre fechas, horarios, tipos de incidentes y ubicaciones geográficas; sin embargo, esta información no siempre se ha utilizado de forma integrada para reconocer zonas y momentos con mayor nivel de riesgo, lo cual reduce su utilidad dentro de la planificación institucional; un análisis adecuado de los comportamientos espaciales y temporales podría mejorar las acciones preventivas desarrolladas por las entidades públicas y la sociedad ecuatoriana (Torres y Gutiérrez, 2021).

Uno de los principales problemas aparece cuando no se identifican con claridad los lugares y horarios donde se concentran los incidentes; ante esa situación, los recursos destinados a seguridad, investigación y atención de emergencias suelen distribuirse después de que ocurren los

hechos, sin una planificación previa que considere las áreas con mayor exposición; además, la falta de modelos predictivos reduce la capacidad de anticipar posibles eventos según la zona geográfica y la franja horaria (Díaz y Hernández, 2019); por ello, las autoridades locales pueden encontrar dificultades al decidir qué provincias requieren mayor atención y en qué momentos podría aumentar la demanda de servicios de emergencia.

El proyecto propone desarrollar un sistema analítico que reúna, depure, represente y modele información relacionada con la seguridad ciudadana en el Ecuador; el análisis se concentra principalmente en Guayas, Pichincha y Manabí, aunque se utilizan registros procedentes de varias fuentes institucionales y bases de datos abiertas; esta información se procesa para elaborar indicadores, comparar comportamientos y presentar los resultados mediante un dashboard interactivo; además, se busca establecer qué provincias registran mayor nivel de riesgo y cuáles son los horarios con mayor frecuencia de incidentes mediante herramientas de análisis geográfico y temporal (Gómez y Ramírez, 2023).

El proyecto surge porque los datos disponibles se encuentran distribuidos en archivos y estructuras diferentes; aunque contienen información útil, no siempre están organizados para responder preguntas relacionadas con la concentración territorial, los cambios registrados a lo largo del tiempo o las variaciones entre provincias; esta dispersión dificulta la comparación de resultados, la estimación de posibles escenarios y la orientación de recursos hacia las zonas con mayores niveles de inseguridad; por esta razón, se combinan procedimientos de exploración de datos, visualización gráfica y modelamiento predictivo.

La pregunta de investigación planteada es la siguiente: ¿cómo contribuye un modelo de integración de datos institucionales, apoyado en el análisis espacial, temporal y predictivo, a la

identificación de patrones de riesgo en delitos de alto impacto y eventos relacionados con la seguridad ciudadana en las provincias de Guayas, Pichincha y Manabí durante el periodo 2022-2025?

Para responder esta pregunta, el proyecto incluye la organización y depuración de las bases de datos, la elaboración de indicadores descriptivos, la representación territorial y temporal de los incidentes y el entrenamiento de modelos predictivos de regresión; los resultados se presentan de forma agrupada, sin realizar interpretaciones sobre personas o casos individuales; así, el sistema busca facilitar la revisión de tendencias, apoyar la selección preliminar de territorios con mayor riesgo y ofrecer una base técnica para posteriores ajustes institucionales y metodológicos.

Los objetivos específicos se presentan en el apartado correspondiente de este capítulo; su formulación considera que los registros observados comprenden el periodo 2022-2025, mientras que los valores estimados para 2026 corresponden únicamente a proyecciones exploratorias; asimismo, las métricas de evaluación se mantienen relacionadas con modelos de regresión, de acuerdo con el tipo de predicción desarrollada.

1.2 Justificación e importancia del trabajo de investigación

La criminalidad urbana es un problema social que afecta la seguridad ciudadana, la organización de los territorios y la calidad de vida de la población; además, los delitos no se distribuyen de la misma manera en todos los lugares ni ocurren con igual frecuencia durante cada periodo, por ello, su estudio requiere métodos capaces de reconocer concentraciones, variaciones y relaciones entre las zonas analizadas (Anselin, 1995).

Los hechos delictivos y los eventos relacionados con la seguridad presentan comportamientos diferentes según el territorio y el momento en que ocurren; en Ecuador, algunos delitos de alto impacto tienden a concentrarse en determinadas provincias, ciudades o sectores, lo cual hace conveniente trabajar con herramientas que ayuden a describir cambios, ubicar áreas con mayores registros y respaldar decisiones con datos; el análisis espacial y temporal sirve, por tanto, para ordenar la información y convertirla en resultados más fáciles de interpretar dentro de la gestión pública.

El aporte académico del estudio se encuentra en la relación entre ciencia de datos, análisis geográfico y seguridad ciudadana; diversos trabajos han mostrado que los delitos pueden agruparse en zonas concretas y mantener variaciones relacionadas con días, meses o franjas horarias, de manera que su identificación requiere procedimientos estadísticos y espaciales específicos (Anselin, 1995; Chainey y Ratcliffe, 2005); igualmente, los algoritmos de agrupamiento y los modelos predictivos ayudan a explorar estructuras presentes en grandes cantidades de registros, aunque sus resultados dependen de la calidad de los datos, la validación aplicada y la prudencia utilizada durante la interpretación (Ester et al., 1996).

La utilidad práctica del proyecto se relaciona con la integración de fuentes institucionales y bases abiertas que, al revisarse por separado, muestran solo una parte del problema; al reunir registros del ECU 911, el Ministerio del Interior y la Fiscalía General del Estado, se obtiene una visión más ordenada sobre robos, homicidios, armas incautadas, personas detenidas, desapariciones, denuncias de delitos y variables sociales; esta integración no reemplaza los sistemas oficiales de análisis criminal, aunque puede complementarlos mediante indicadores, gráficos, mapas y modelos de carácter exploratorio.

Guayas, Pichincha y Manabí fueron seleccionadas por su importancia poblacional, territorial e institucional, así como por la presencia que mantienen dentro de los registros de delitos de alto impacto; sin embargo, el componente predictivo emplea un panel territorial más amplio, elaborado con las unidades incluidas en las bases procesadas, lo cual expone a los modelos a una mayor variedad de comportamientos; esta decisión puede favorecer la estabilidad del entrenamiento, mientras que la explicación de los resultados se mantiene concentrada en las tres provincias escogidas.

La propuesta metodológica busca superar una revisión basada únicamente en frecuencias y porcentajes; para ello, integra análisis exploratorio, visualizaciones interactivas y modelos predictivos destinados a estimar posibles escenarios, con especial atención al año 2026; estas estimaciones deben interpretarse como aproximaciones sujetas a la cobertura, consistencia y calidad de los registros disponibles, sin asumir que representan predicciones exactas sobre la ocurrencia futura de delitos.

El valor social del estudio se relaciona con la creación de una herramienta que facilita la comparación de tendencias y diferencias entre provincias; sus resultados pueden apoyar conversaciones técnicas sobre planificación preventiva, distribución de recursos y priorización territorial; además, el trabajo ofrece una base reproducible y visualmente comprensible para observar la evolución de distintos indicadores de seguridad ciudadana, ordenar registros dispersos y promover un uso responsable de los datos durante el análisis de la inseguridad.

1.3 Alcance

El proyecto se centra en el análisis cuantitativo de delitos de alto impacto y otros eventos relacionados con la seguridad ciudadana en el Ecuador; aunque se utilizan datos de distintas zonas del país, la interpretación se concentra principalmente en Guayas, Pichincha y Manabí. La información proviene de registros institucionales y bases abiertas, las cuales fueron integradas, depuradas y organizadas para fines académicos; el análisis se realiza a nivel provincial y temporal, sin estudiar casos individuales ni emitir valoraciones sobre personas específicas.

La delimitación temporal se divide en dos partes; por un lado, se revisan los datos observados entre 2022 y 2025, siempre que las bases presenten información completa, coherente y útil para el análisis. Por otro lado, los resultados correspondientes a 2026 se manejan como proyecciones exploratorias obtenidas mediante modelos estadísticos y algoritmos de aprendizaje automático; esto quiere decir que no representan datos reales completos de ese año, sino estimaciones que pueden variar según la calidad de la información utilizada.

Respecto a la delimitación geográfica, Guayas, Pichincha y Manabí se mantienen como las provincias principales del estudio; su selección responde a su peso poblacional, territorial e institucional, así como a la presencia de delitos de alto impacto. Sin embargo, para entrenar y validar los modelos se puede trabajar con un grupo más amplio de provincias incluidas en las bases procesadas; esta decisión ayuda a incorporar una mayor variedad de datos, aunque la explicación de los resultados continúa enfocada en las tres provincias seleccionadas.

La investigación es cuantitativa, aplicada, no experimental y longitudinal; trabaja con datos numéricos, indicadores y registros obtenidos de varias fuentes, mientras su carácter aplicado se relaciona con la creación de una solución analítica para estudiar la seguridad ciudadana. No se

manipulan variables ni se interviene directamente sobre los hechos examinados; además, se observan los cambios registrados durante varios años. El trabajo también es exploratorio, descriptivo y predictivo, ya que primero organiza y revisa bases con estructuras diferentes, luego compara tendencias y concentraciones, y después estima posibles comportamientos futuros.

La parte técnica incluye la recolección, limpieza, integración y preparación de los datos; también se elaboran indicadores, gráficos, mapas y un dashboard interactivo para presentar la información de forma comprensible. Según la estructura de cada base, se pueden aplicar modelos como ARIMA, XGBoost y TFT, junto con técnicas de agrupamiento o segmentación; para medir su rendimiento se utilizan indicadores propios de regresión, entre ellos MAE, RMSE, R^2 y MAPE, evitando métricas de clasificación cuando la variable analizada es numérica y continua.

El proyecto también contempla una aplicación web destinada a mostrar resultados, tendencias, comparaciones provinciales y proyecciones; esta herramienta no funciona como un sistema oficial de alerta temprana ni toma decisiones por cuenta propia. Su utilidad se limita al apoyo exploratorio y académico, ya que facilita la revisión de los datos y la comparación entre territorios; tampoco trabaja en tiempo real, pues su actualización depende de la disponibilidad de nuevas fuentes y de los procesos definidos para cargar la información.

El estudio reconoce varias limitaciones que deben considerarse durante la interpretación; los resultados dependen de la calidad y consistencia de las bases originales, mientras algunos modelos pueden trabajar con escalas temporales o territoriales diferentes. Por esta razón, las comparaciones deben realizarse con cuidado; además, las proyecciones no explican causas ni aseguran que los eventos futuros ocurran exactamente como fueron estimados. Estas condiciones

ayudan a establecer hasta dónde pueden utilizarse las conclusiones y recomendaciones del proyecto.

1.4 Objetivos

1.4.1 Objetivo general

Analizar los patrones espacio-temporales de los delitos de alto impacto registrados por el ECU 911, el Ministerio del Interior y la Fiscalía General del Estado, en las provincias de Guayas, Pichincha y Manabí, mediante técnicas de análisis geoespacial y modelos predictivos de aprendizaje automático utilizando información histórica del período comprendido entre enero de 2022 a diciembre de 2025.

1.4.2 Objetivos específicos

- Integrar y depurar las bases de datos de incidentes proporcionadas por el ECU 911, el Ministerio del Interior y la Fiscalía General del Estado, garantizando la calidad, consistencia temporal y coherencia geográfica de la información.
- Aplicar técnicas de preprocesamiento y análisis exploratorio de datos, incluyendo limpieza, normalización y estructuración, para preparar la información para el análisis de patrones espaciales y temporales.
- Diseñar e implementar técnicas de análisis geoespacial y modelos de aprendizaje automático para la detección de patrones espacio-temporales, incorporando variables contextuales socioeconómicas relevantes.

- Evaluar el desempeño de los modelos geoespaciales y predictivos mediante métricas de evaluación de regresión, como MAE (Error Absoluto Medio), RMSE (Raíz del Error Cuadrático Medio), MAPE (Error Porcentual Absoluto Medio) y R^2 (Coeficiente de Determinación), con el fin de formular lineamientos estratégicos basados en evidencia que contribuyan a optimizar la asignación de recursos y fortalecer las acciones preventivas y operativas en la gestión de la seguridad ciudadana.

CAPÍTULO 2: ESTADO DEL ARTE Y MARCO TEÓRICO

El presente capítulo tiene como finalidad analizar los antecedentes científicos relacionados con la predicción de la violencia y los delitos mediante técnicas de Ciencia de Datos e Inteligencia Artificial. Se revisaron estudios sobre criminalidad en América Latina, aplicaciones de algoritmos de aprendizaje automático en la predicción del delito y los avances recientes en modelos basados en Transformers para series temporales. La revisión de estos trabajos permite identificar las principales tendencias metodológicas, evaluar sus resultados y establecer las bases conceptuales y técnicas que sustentan el desarrollo del modelo propuesto para la predicción del delito y criminalidad en el Ecuador.

2.1 Estado del Arte

2.1.1 Estudios sobre predicción de criminalidad mediante Machine Learning

El aprendizaje automático se ha utilizado para reconocer relaciones complejas vinculadas con la ocurrencia de delitos; estos algoritmos procesan registros históricos, comparan múltiples variables y detectan comportamientos que podrían pasar inadvertidos en análisis convencionales. Perry et al. (2013) desarrollaron modelos orientados a anticipar zonas urbanas con mayor exposición delictiva; según los resultados obtenidos, este tipo de análisis podía apoyar la organización de los recursos policiales y la planificación de medidas preventivas.

La incorporación de Machine Learning al estudio del crimen tomó mayor fuerza durante la última década; uno de los antecedentes más conocidos corresponde a los procesos puntuales autoexcitados, utilizados para estimar la probabilidad de nuevos eventos a partir de delitos ocurridos previamente. Mohler et al. (2011) encontraron que ciertos hechos criminales mostraban

comportamientos de repetición y cercanía espacial parecidos a otros fenómenos de propagación; esta característica ayudó a reconocer sectores donde podían formarse nuevas concentraciones delictivas.

También se aplicaron técnicas de minería de datos para examinar la criminalidad urbana; los resultados mostraron que los algoritmos podían encontrar asociaciones difíciles de identificar mediante procedimientos estadísticos tradicionales, especialmente cuando se trabajaba con grandes cantidades de registros y variables relacionadas entre sí (Yu et al., 2011). Aun así, la utilidad de estas herramientas dependía de la preparación de las bases y de la forma en que se interpretaban las estimaciones obtenidas.

Con el avance del aprendizaje profundo se incorporaron datos demográficos, ambientales y procedentes de redes sociales; Kang y Kang (2017) integraron estas fuentes dentro de un mismo modelo y obtuvieron un mejor rendimiento frente a propuestas que utilizaban información más limitada. La combinación de datos de distinta naturaleza amplió la capacidad para reconocer variaciones territoriales; sin embargo, los resultados continuaban sujetos a la cobertura, consistencia y calidad de los registros empleados.

Otros trabajos recurrieron a Random Forest, máquinas de soporte vectorial, redes neuronales y modelos de grafos para estimar delitos, analizar reincidencia y localizar áreas con mayores concentraciones; en varios casos, estas técnicas mostraron resultados superiores a los alcanzados mediante métodos tradicionales. La inclusión de variables espaciales, temporales y socioeconómicas también estuvo relacionada con una mayor precisión predictiva, aunque ninguna estimación debía asumirse como una representación exacta de los hechos futuros.

Para estudiar la ocurrencia de homicidios, Ribeiro et al. (2022) aplicaron un modelo Random Forest en la ciudad de Belém, estado de Pará, Brasil; la precisión reportada fue cercana al 76%, resultado que mostró la posible utilidad del aprendizaje automático como apoyo para examinar sectores con mayores registros y organizar medidas relacionadas con la seguridad pública. Este porcentaje, no obstante, debía interpretarse según las características de la base, la unidad territorial utilizada y el procedimiento de validación aplicado.

La comparación entre varios algoritmos también mostró diferencias en su rendimiento; Random Forest y XGBoost alcanzaron mejores resultados que ciertos modelos estadísticos convencionales, tanto por su precisión como por su capacidad para trabajar con datos distintos a los empleados durante el entrenamiento (Wang et al., 2017). Estos antecedentes respaldan el uso de métodos computacionales dentro del análisis delictivo; pese a ello, las predicciones requieren acompañamiento técnico y una interpretación prudente antes de ser consideradas dentro de decisiones institucionales.

2.1.2 Aplicación de Deep Learning en el análisis criminal

El Deep Learning se ha consolidado como una de las principales técnicas de inteligencia artificial para el análisis y la predicción de fenómenos delictivos y de violencia. Su capacidad para procesar grandes volúmenes de datos y modelar relaciones complejas entre variables espaciales y temporales permite identificar patrones criminales que difícilmente pueden ser detectados mediante métodos estadísticos tradicionales. Diversos estudios han demostrado que las arquitecturas de aprendizaje profundo mejoran la precisión de las predicciones delictivas, contribuyendo al diseño de estrategias de prevención y seguridad pública (Stalidis et al., 2021; Mao et al., 2025).

El avance del Deep Learning ha permitido desarrollar modelos capaces de capturar relaciones complejas entre variables espaciales y temporales. La implementación de arquitecturas CNN-LSTM para la predicción de delitos urbanos, obteniendo mejoras significativas respecto a modelos ARIMA tradicionales.

Los resultados demostraron que las redes neuronales permiten modelar patrones no lineales presentes en grandes volúmenes de datos delictivos, especialmente cuando se integran variables geográficas y socioeconómicas.

Recientemente, las arquitecturas basadas en Transformers han comenzado a utilizarse en tareas de predicción temporal. Lim et al. (2021) desarrollaron el modelo Temporal Fusion Transformer (TFT), diseñado para trabajar con series temporales multivariantes y variables exógenas.

Los resultados mostraron mejoras significativas en precisión e interpretabilidad respecto a modelos tradicionales como ARIMA y LSTM, especialmente en escenarios donde intervienen múltiples factores explicativos.

2.1.3 Estudios relacionados con criminalidad en América Latina y el Ecuador

La CEPAL (2024) señala que América Latina y el Caribe mantienen altos niveles de desigualdad en la distribución del ingreso y en el acceso a oportunidades; estas condiciones pueden relacionarse con distintas expresiones de violencia y criminalidad. Frente a esta realidad, varias investigaciones han recurrido a la ciencia de datos para estudiar problemas vinculados con la seguridad ciudadana; aun así, una parte importante de estos trabajos se concentra en análisis descriptivos o modelos estadísticos convencionales, mientras la aplicación de arquitecturas

avanzadas de inteligencia artificial para la predicción espacial y temporal del delito todavía resulta limitada.

La incorporación del aprendizaje automático abrió nuevas posibilidades para estimar tendencias delictivas a partir de grandes volúmenes de información. Ordoñez et al. (2020), por ejemplo, desarrollaron un modelo basado en Random Forest con más de dos millones de registros históricos de Colombia; los resultados mostraron que este tipo de algoritmo podía servir como apoyo para analizar la evolución de los homicidios y respaldar decisiones dentro de las instituciones encargadas de la seguridad pública.

En Brasil también se han utilizado técnicas de agrupamiento para reconocer concentraciones territoriales de homicidios; las tasas más altas se localizaron principalmente en municipios próximos a grandes ciudades, lo cual muestra que la violencia no se distribuye de forma uniforme y que su comportamiento podría estar relacionado con condiciones propias de cada territorio (Sousa et al., 2018). Este tipo de análisis ayuda a comprender que la cercanía con grandes centros urbanos puede influir en la formación de zonas con mayores niveles de incidencia delictiva.

En Ecuador, la predicción del delito se ha trabajado principalmente mediante métodos estadísticos y algoritmos de aprendizaje automático; buena parte de las investigaciones se ha concentrado en homicidios y en la identificación de variables asociadas con su ocurrencia. Chávez y Abril (2023) aplicaron una regresión logística en el Distrito Metropolitano de Quito para estimar la probabilidad de homicidios; entre las variables incorporadas, el lugar del hecho, especialmente la vía pública, mostró una relación relevante con el comportamiento del modelo.

Otro estudio realizado en la región interandina combinó regresión logística y redes neuronales artificiales para analizar la incidencia delictiva; Gallegos y Bolaños (2026) encontraron

que la integración de ambos métodos mejoraba la capacidad de estimación frente al uso aislado de una sola técnica. Esta combinación mostró que los modelos híbridos pueden adaptarse mejor a fenómenos sociales donde intervienen varias condiciones al mismo tiempo.

Por su parte, Zapata y Vivas (2024) utilizaron árboles de decisión y Random Forest para estimar homicidios intencionales en Ecuador; los algoritmos identificaron relaciones relevantes dentro de los registros analizados y alcanzaron mejores niveles de precisión que algunos procedimientos convencionales. Sin embargo, el rendimiento de estos modelos dependió de la calidad de los datos, la selección de variables y el proceso de validación aplicado.

La inteligencia artificial también se ha utilizado para estudiar otras formas de violencia. Alarcón (2024) combinó aprendizaje automático y procesamiento de lenguaje natural para analizar la violencia de género en Ecuador; la propuesta integró información social y digital, ampliando el estudio más allá de los delitos contra la vida. Este trabajo mostró que las fuentes textuales y los datos obtenidos de espacios digitales también pueden aportar información útil para reconocer comportamientos relacionados con la violencia.

Rosero (2025) desarrolló un modelo predictivo de criminalidad para Ecuador mediante la integración de técnicas estadísticas y aprendizaje profundo; el análisis incorporó variables externas asociadas con el comportamiento delictivo, lo cual produjo estimaciones más amplias que aquellas basadas únicamente en registros criminales. Los resultados también mostraron que la combinación de Deep Learning con métodos estadísticos podía ofrecer una representación más completa de las variaciones observadas.

Pese a estos avances, todavía existen pocos estudios ecuatorianos que utilicen arquitecturas como Temporal Fusion Transformer; también se observa una integración reducida de variables

delictivas, geográficas y socioeconómicas dentro de un mismo proceso de predicción. Esta ausencia limita la comparación entre modelos tradicionales, algoritmos de aprendizaje automático y arquitecturas profundas aplicadas a series temporales.

2.2 Marco Teórico

2.2.1 Criminalidad y Seguridad Ciudadana

La criminalidad constituye uno de los fenómenos sociales más complejos debido a su naturaleza dinámica, multicausal y territorial. Diversos estudios sostienen que los eventos delictivos no pueden ser explicados únicamente por factores individuales, sino que responden a la interacción de variables sociales, económicas, institucionales, demográficas y espaciales que influyen en la ocurrencia de conductas delictivas (Palomeque, 2001).

Este estudio aborda la violencia más allá de las conductas individuales y patológicas, concibiéndola desde un marco teórico que la define como un fenómeno estructural y multidimensional. Un tipo particular de relación social en la que intervienen, al menos, dos clases de actores que, como forma de resolver el conflicto de sus intereses diferentes, hacen o intentan hacer daño, en términos físicos o psicológicos (Guerrero, 1997; Guzmán, 1994).

La seguridad ciudadana requiere herramientas analíticas capaces de mostrar no solo dónde se concentran los delitos, sino también en qué momentos ocurren y qué factores podrían estar relacionados con sus variaciones; el aumento de información generada por organismos públicos ha favorecido el uso de ciencia de datos e inteligencia artificial para reconocer comportamientos delictivos, comparar tendencias y construir modelos predictivos que sirvan como apoyo para la toma de decisiones.

En Ecuador, el aumento de la violencia registrado a partir de finales de 2020, junto con los cambios observados en las dinámicas delictivas, ha impulsado la búsqueda de métodos más precisos para estudiar este fenómeno; por ello, resulta conveniente trabajar con herramientas que ayuden a ubicar zonas de mayor riesgo, reconocer variaciones temporales y examinar posibles tendencias asociadas con la criminalidad.

2.2.2 Ciencia de Datos aplicada al análisis criminal

La ciencia de datos reúne aportes de la estadística, las matemáticas, la informática y la inteligencia artificial; su propósito consiste en analizar grandes cantidades de información para obtener resultados útiles y comprensibles. Provost y Fawcett (2013) explican que esta disciplina combina métodos, procesos y herramientas orientados a descubrir patrones dentro de los datos; aplicada a la seguridad ciudadana, puede convertir registros históricos en insumos para organizar recursos, examinar tendencias y apoyar medidas preventivas.

En el análisis criminal, estas técnicas ayudan a reconocer concentraciones geográficas, cambios temporales, comportamientos atípicos y relaciones entre variables que no siempre son visibles a simple vista; también pueden utilizarse para estimar posibles variaciones futuras en hechos como homicidios, robos y desapariciones. Su utilidad, sin embargo, depende de la calidad de las bases, la forma en que se preparan los datos y la prudencia empleada al interpretar los resultados.

El aprendizaje automático forma parte de la inteligencia artificial y trabaja con algoritmos capaces de aprender a partir de registros históricos; en lugar de seguir únicamente reglas programadas de forma directa, los modelos identifican regularidades y las utilizan para generar estimaciones sobre datos nuevos. Este proceso se apoya en la teoría del aprendizaje estadístico y

busca que el modelo mantenga un rendimiento aceptable fuera de la muestra utilizada durante el entrenamiento.

Dentro de esta área suelen distinguirse tres tipos de aprendizaje; el supervisado utiliza datos previamente etiquetados para relacionar variables de entrada con una respuesta conocida, el no supervisado busca agrupaciones o estructuras sin categorías definidas y el aprendizaje por refuerzo se basa en recompensas obtenidas después de ejecutar determinadas acciones. En los estudios sobre criminalidad se emplean con mayor frecuencia los dos primeros, ya que facilitan tanto la predicción como la identificación de agrupamientos territoriales.

Las investigaciones orientadas a estimar delitos suelen recurrir a regresión logística, árboles de decisión, Random Forest y XGBoost; estos algoritmos procesan variables temporales, geográficas y socioeconómicas para reconocer relaciones asociadas con la ocurrencia de determinados eventos. La selección de un modelo no debería basarse únicamente en el nivel de precisión alcanzado; también conviene considerar su capacidad de generalización, la facilidad para interpretar los resultados y la consistencia de la información utilizada.

El Deep Learning amplía estas posibilidades mediante redes neuronales con varias capas de procesamiento; dichas arquitecturas pueden representar relaciones no lineales y trabajar con grandes volúmenes de datos. LeCun et al. (2015) señalan que el aprendizaje profundo extrae representaciones jerárquicas de la información, lo cual puede ser útil cuando los registros contienen estructuras complejas o relaciones difíciles de modelar con procedimientos más simples.

En seguridad ciudadana, las redes neuronales se han utilizado para combinar fuentes distintas y analizar dependencias entre variables; entre las arquitecturas más conocidas se encuentran las redes neuronales artificiales, las redes convolucionales, las redes recurrentes, los

modelos Long Short-Term Memory y los Transformers. Algunas investigaciones han reportado un mejor rendimiento frente a ciertos métodos estadísticos tradicionales; aun así, estos resultados deben revisarse según el tipo de dato, la cantidad de observaciones, la validación aplicada y la finalidad del análisis.

2.2.3 Análisis Espacio-Temporal de la Criminalidad

Sherman et al. (1989) plantean que el delito mantiene una relación directa con el espacio y el tiempo; esto quiere decir que los hechos delictivos no aparecen de forma aleatoria, sino que suelen concentrarse en determinados lugares y durante periodos concretos. A partir de esta distribución pueden reconocerse comportamientos recurrentes mediante técnicas analíticas; el análisis espacial ayuda a ubicar zonas con alta concentración criminal, conocidas como hotspots, mientras el análisis temporal sirve para identificar tendencias, estacionalidades y ciclos vinculados con la ocurrencia de delitos.

La combinación de ambas dimensiones ofrece una visión más completa del fenómeno; al revisar al mismo tiempo la ubicación y el momento en que ocurren los eventos, es posible comprender mejor sus variaciones y construir modelos predictivos con mayor detalle. Esta integración resulta especialmente útil cuando se busca anticipar cambios en la incidencia delictiva o comparar comportamientos entre distintos territorios.

Las series temporales también se utilizan para estudiar fenómenos que cambian de manera continua; están formadas por observaciones registradas en distintos momentos y organizadas según una secuencia cronológica. Su análisis ayuda a reconocer aumentos, disminuciones, ciclos y comportamientos que se repiten durante ciertos periodos.

Aplicadas a la seguridad ciudadana, las series temporales facilitan la identificación de etapas con mayor presencia delictiva; además, pueden emplearse para estimar posibles comportamientos futuros a partir de los registros históricos disponibles. Sin embargo, estas proyecciones deben interpretarse con prudencia, ya que su precisión depende de la calidad de los datos, la estabilidad de los patrones y el modelo utilizado.

2.2.4 Modelos ARIMA, ExpSmooth, XGBoost y Temporal Fusion Transformer

Los modelos de pronóstico de series temporales pueden clasificarse en enfoques estadísticos tradicionales y enfoques de aprendizaje automático. Entre los primeros destacan ARIMA y suavizamiento exponencial (*exponential smoothing*), mientras que *XGBoost* pertenece a los métodos de aprendizaje automático basados en árboles de decisión potenciados (*boosted trees*) (Box et al., 2015; Chen y Guestrin, 2016; Hyndman y Athanasopoulos, 2021).

AutoRegressive Integrated Moving Average (ARIMA) es uno de los modelos más utilizados para el análisis de series temporales debido a su capacidad para modelar tendencias y componentes estacionales. Sin embargo, presenta limitaciones cuando existen relaciones altamente no lineales o múltiples variables explicativas.

Exponential Smoothing (ExpSmooth), estos modelos de suavizado exponencial asignan mayor importancia a las observaciones recientes, permitiendo adaptarse rápidamente a cambios en la dinámica de los datos. Aunque son eficientes para modelar tendencias simples, presentan dificultades para capturar relaciones complejas entre múltiples variables.

Extreme Gradient Boosting (XGBoost) es un algoritmo basado en árboles de decisión que ha demostrado un desempeño sobresaliente en problemas de clasificación y regresión. Entre sus

principales ventajas destacan: alta precisión predictiva, capacidad para modelar relaciones no lineales, manejo eficiente de grandes volúmenes de datos, tratamiento automático de valores faltantes, incorporación de mecanismos de regularización para reducir el sobreajuste.

Estas características convierten a XGBoost en una alternativa robusta para la predicción de fenómenos delictivos, especialmente cuando se combinan variables temporales, espaciales y socioeconómicas.

El Temporal Fusion Transformer (TFT) representa una de las arquitecturas más avanzadas para el análisis de series temporales multivariantes. A diferencia de los modelos tradicionales, TFT combina mecanismos de atención, redes neuronales profundas y componentes recurrentes para modelar relaciones complejas entre variables temporales y contextuales. Sus principales ventajas incluyen: integración simultánea de múltiples variables, incorporación de variables exógenas, predicción multihorizonte (*multi-horizon forecasting*), interpretabilidad mediante mecanismos de atención, capacidad para modelar relaciones no lineales complejas.

En esta investigación, TFT constituye el modelo principal debido a su capacidad para integrar variables socioeconómicas, espaciales y temporales relacionados con la criminalidad en las provincias de Guayas, Pichincha y Manabí.

2.2.5 Calidad de Datos y Preprocesamiento, IA en la seguridad ciudadana

La calidad de los datos representa un factor determinante para el desempeño de cualquier modelo de inteligencia artificial. Por esta razón, antes de la construcción de los modelos predictivos se aplican procesos de limpieza, normalización, integración y transformación de datos.

Entre las técnicas utilizadas destacan: tratamiento de valores nulos, eliminación de duplicados, corrección de inconsistencias, estandarización de variables, normalización de datos, ingeniería de características, imputación mediante técnicas estadísticas y algoritmos de aprendizaje automático. Estos procedimientos permiten reducir sesgos, mejorar la calidad de la información y aumentar la capacidad de generalización de los modelos predictivos.

La incorporación de inteligencia artificial en el análisis criminal amplió las posibilidades de estudio frente a los métodos apoyados únicamente en estadística descriptiva; estas técnicas ayudan a procesar grandes volúmenes de información y a reconocer relaciones que podrían pasar desapercibidas en revisiones más básicas.

Al combinar análisis geoespacial, aprendizaje automático, Deep Learning y modelos de series temporales, se pueden construir sistemas orientados a identificar comportamientos complejos, estimar tendencias y generar información útil para la toma de decisiones; aun así, los resultados deben interpretarse con cautela, pues dependen de la calidad de los datos, la selección de variables y el proceso de validación aplicado.

CAPÍTULO 3: DESARROLLO DEL PROYECTO

El desarrollo práctico del proyecto siguió una metodología de ciencia de datos orientada a recopilar, integrar, depurar, transformar, explorar, visualizar y modelar información criminológica procedente de distintas fuentes; a partir de este proceso se construyó una plataforma analítica capaz de organizar grandes volúmenes de datos sobre delitos y seguridad ciudadana. La propuesta convirtió registros dispersos en información más ordenada y comprensible, útil como apoyo para decisiones sustentadas en evidencia.

La solución reunió datos sobre homicidios, robos, denuncias de delitos, armas incautadas, personas desaparecidas y detenidos; como la información provenía del ECU 911, el Ministerio del Interior y la Fiscalía General del Estado, fue necesario aplicar procesos de normalización, homologación y estandarización. Con ello se buscó mantener coherencia entre variables temporales, geográficas, sociodemográficas y categóricas; luego, toda la información se organizó dentro de una estructura común para su análisis estadístico y computacional.

El proyecto se diseñó como una solución modular destinada a la integración de datos, el análisis exploratorio, la visualización interactiva y el modelamiento predictivo de la seguridad ciudadana; la plataforma se organizó en varios componentes que automatizan tareas como la limpieza de registros, la estandarización de variables, la elaboración de indicadores socioeconómicos, la generación de gráficos y mapas, y la ejecución de modelos de predicción.

La arquitectura modular separó cada proceso analítico para evitar que las funciones dependieran unas de otras de manera innecesaria; esta organización facilita el mantenimiento del sistema, la reutilización de componentes y la incorporación de nuevas fuentes, indicadores o modelos sin alterar el funcionamiento general de la plataforma. También brinda mejores

condiciones para ampliar la solución conforme se disponga de nuevos datos o se requieran ajustes técnicos.

Como resultado, se obtuvo un entorno de análisis criminológico orientado a reconocer patrones espaciales y temporales, revisar tendencias históricas, comparar indicadores territoriales y examinar características sociodemográficas; además, la plataforma genera visualizaciones dinámicas que facilitan el seguimiento de eventos relacionados con la seguridad ciudadana y la comparación entre distintas unidades geográficas.

La integración de estas capacidades ofrece una visión más amplia del comportamiento delictivo; al combinar análisis estadístico, aprendizaje automático, series temporales e inteligencia artificial, la plataforma sirve como base para estimar tendencias y respaldar procesos de planificación en seguridad pública. Sus resultados no sustituyen el criterio institucional, aunque aportan información organizada para apoyar la revisión técnica y la toma de decisiones.

Tabla 1

Arquitectura del Sistema

Componente	Responsabilidad	Artefactos principales
Configuración	Define rutas, fuentes, periodo, variables objetivo y carpetas de salida.	config.py
Ingeniería de datos	Lee CSV, detecta codificaciones, estandariza columnas, fechas, provincias y tipos numéricos.	utils/cleaning.py, utils/data_loader.py
Analítica visual	Aplica filtros, calcula KPIs y genera gráficos, tablas y descargas.	utils/filters.py, metrics.py, charts.py, ui.py
Aplicación	Presenta las páginas temáticas y la navegación del dashboard.	app.py y pages/

Modelamiento	Integra variables sociales, entrena modelos y genera predicciones y métricas.	scripts/ y procesados_ml/
--------------	---	---------------------------

Nota. Modelos de Análisis Predictivo.

3.1 Recopilación de datos

La recopilación de datos constituyó una de las etapas fundamentales del proyecto, debido a que la calidad y representatividad de la información influyen directamente en la precisión de los análisis y en la capacidad de los modelos para identificar patrones asociados a la criminalidad. Para este estudio se utilizaron bases de datos oficiales provenientes del ECU 911, Ministerio del Interior y Fiscalía General del Estado, las cuales fueron almacenadas y procesadas en formato CSV para facilitar su integración y análisis.

Las fuentes seleccionadas reúnen información histórica sobre distintos hechos relacionados con la actividad delictiva y la seguridad ciudadana en el Ecuador; entre las categorías revisadas se encuentran homicidios, robos, denuncias de delitos, armas incautadas, personas desaparecidas y detenidos. La variedad de estos registros ayudó a examinar la criminalidad a partir de varias dimensiones; se consideraron datos temporales, geográficos, demográficos y categóricos que ampliaron la comprensión del fenómeno.

También se incorporaron bases complementarias con tasas provinciales de incidencia para algunas categorías; en esta información no solo se revisó la cantidad total de eventos, sino también indicadores relativos que facilitaron la comparación entre provincias con tamaños poblacionales y características demográficas diferentes. Esta forma de análisis ofreció una visión más equilibrada

del comportamiento delictivo en los territorios estudiados; además, evitó que la interpretación dependiera únicamente de valores absolutos.

La integración de todas las fuentes dio lugar a una base consolidada que sirvió como insumo para las etapas de limpieza, transformación, análisis exploratorio, visualización y modelamiento predictivo; con ello se obtuvo una visión más amplia de la criminalidad y fue posible reconocer tendencias, concentraciones geográficas y variaciones temporales relevantes para la investigación.

Figura 1

Arquitectura funcional de la solución analítica.



Nota. Modelos de Análisis Predictivo.

3.2 Selección de los conjuntos de datos

Las bases de datos se seleccionaron según su utilidad para estudiar la seguridad ciudadana; se incluyeron variables de tiempo, ubicación, características de las personas y tipo de delito. Entre ellas constan año, mes, provincia, cantón, sexo, edad, nacionalidad, tipo de delito, motivación y arma utilizada; esta información ayudó a comparar territorios y reconocer tendencias.

Al inicio se trabajó con homicidios, robos, denuncias de delitos, armas incautadas, personas desaparecidas y detenidos. Como esta última base registra kilogramos incautados y no cantidad de hechos, se analizó por separado; todas las fuentes integradas sirvieron para la exploración, visualización y predicción de los datos.

Figura 2

Flujo de procesamiento y análisis de datos.



Nota. Modelos de Análisis Predictivo.

3.3 Análisis exploratorio de datos

El Análisis Exploratorio de Datos se utilizó para conocer la estructura y calidad de las bases antes de elaborar visualizaciones y modelos predictivos; durante esta etapa se revisaron tipos de variables, valores nulos, duplicados, cantidades de registros y estadísticas descriptivas. También se corrigieron formatos temporales, numéricos y geográficos, con el fin de trabajar con información más ordenada y consistente.

Para comprender el comportamiento de los datos se elaboraron histogramas, gráficos de barras y líneas, tablas dinámicas, mapas de calor y representaciones geográficas; estas herramientas ayudaron a reconocer cambios temporales, concentraciones territoriales y características sociodemográficas.

En homicidios se analizaron variables como sexo, edad, nacionalidad, nivel educativo, tipo de arma, posible motivación, lugar y fecha del hecho; además, se revisaron zonas con mayor concentración y horarios con más registros. Las denuncias se estudiaron según provincia, mes, tipo de delito y condición de consumado o tentativa, mientras que las armas incautadas se revisaron por categoría, calibre, territorio y delito asociado.

Para las personas desaparecidas se consideraron edad, sexo, nacionalidad, provincia, estado del caso y tiempo transcurrido hasta su localización; en el caso de los detenidos, se analizaron características demográficas, presuntas infracciones, lugares de detención, armas relacionadas y distribución territorial.

Este análisis ayudó a reconocer tendencias, concentraciones geográficas, perfiles frecuentes y relaciones entre variables; los resultados sirvieron para construir indicadores, diseñar el dashboard y preparar los datos utilizados posteriormente en los modelos de inteligencia artificial.

3.4 Preprocesamiento de los datos

La etapa de preprocesamiento se realizó para ordenar y preparar la información antes del análisis, la visualización y el modelamiento; como las bases provenían de distintas fuentes, fue necesario revisar valores nulos, registros duplicados, errores de formato y datos incompletos.

También se corrigieron problemas de codificación, diferencias en nombres de provincias, cantones y categorías delictivas, además de variaciones en fechas, horas y valores numéricos; las coordenadas geográficas fueron revisadas y ajustadas para su uso en mapas y análisis territoriales.

Luego, las fuentes se organizaron bajo una estructura común que reunió variables como fecha, ubicación, cantidad de eventos, tipo de delito y características sociodemográficas; con ello se logró consolidar la información en una sola base y facilitar las comparaciones entre categorías.

Este proceso dejó los datos en mejores condiciones para construir indicadores, generar visualizaciones y aplicar modelos predictivos; además, redujo errores que podían afectar la interpretación de los resultados.

3.5 Análisis de modelo y visualización

Como parte de la fase aplicada se desarrolló una plataforma interactiva para explorar y analizar información relacionada con la seguridad ciudadana; su diseño combina inteligencia de negocios y analítica visual, con el fin de convertir grandes volúmenes de datos en resultados más claros y fáciles de revisar.

La plataforma reúne varias fuentes criminológicas dentro de un solo entorno; además, incorpora filtros por categoría, periodo, ubicación y tipo de evento, lo cual facilita consultas específicas y comparaciones entre distintos territorios o momentos.

También se incluyeron indicadores como total de eventos, número de registros, variaciones anuales, tasas provinciales y cambios frente a periodos anteriores; estos datos ofrecen una visión rápida del comportamiento de cada fenómeno y ayudan a reconocer aumentos o disminuciones.

Para presentar la información se utilizaron gráficos anuales, distribuciones mensuales, rankings por provincias y cantones, mapas temáticos, mapas de calor y representaciones basadas

en coordenadas; junto con ello, se analizaron variables como sexo, edad, nacionalidad y nivel educativo.

La plataforma no se limita a mostrar datos históricos; también organiza la información necesaria para aplicar series temporales, algoritmos de aprendizaje automático y otras técnicas de inteligencia artificial orientadas a estimar tendencias y reconocer patrones delictivos.

3.6 Evolución y versiones del desarrollo

El desarrollo de la solución siguió un proceso iterativo e incremental; cada versión incorporó ajustes técnicos y nuevas funciones según las necesidades identificadas durante las pruebas. La primera versión se centró en integrar las bases y mostrar gráficos básicos sobre tendencias, indicadores generales y distribución territorial de los eventos.

Más adelante, la plataforma se organizó mediante módulos especializados por categoría criminológica; se añadieron análisis territoriales, tasas provinciales, validación de datos geográficos, indicadores comparativos y mapas de intensidad. Estos cambios ampliaron las opciones de análisis y facilitaron la exploración de la información.

También se incorporaron funciones para descargar datos filtrados, agilizar consultas y revisar con mayor precisión las coordenadas geográficas; así, el dashboard pasó de mostrar resultados descriptivos a funcionar como un entorno integrado para analizar información criminológica y preparar los datos utilizados en los modelos predictivos.

3.7 Modelos predictivos implementados

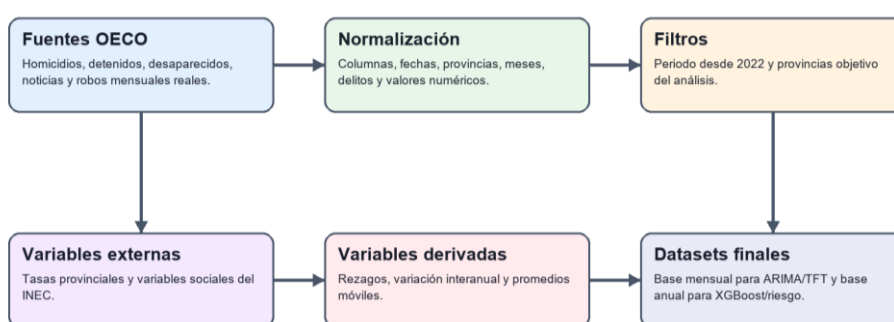
Esta sección presenta la configuración técnica de los modelos aplicados en el proyecto; la descripción se elaboró a partir del código del dashboard y de los scripts de entrenamiento, con el fin de mantener correspondencia entre la metodología y la solución desarrollada.

3.7.1 Preparación de datos para el modelamiento

Los modelos predictivos utilizaron los archivos generados por el módulo de integración; allí se reunieron datos sobre robos, homicidios, denuncias de delitos, armas incautadas, personas detenidas, personas desaparecidas y variables sociales, con registros disponibles a partir de 2022. Durante el procesamiento se ajustaron nombres de columnas, fechas, meses, provincias, cantones, categorías delictivas y valores numéricos; además, la base de robos se incorporó con información mensual real.

Figura 3

Flujo de integración y construcción de datasets.



Nota. El flujo representa el proceso implementado en scripts/00_integracion_social_crimen.py.

Tabla 2*Archivos generados para el modelamiento predictivo*

Archivo generado	Unidad de análisis	Uso principal
oeco_normalizado_eventos.csv	Evento normalizado	Base integrada de eventos después de limpieza, normalización territorial y filtrado temporal.
criminalidad_mensual_con_sociales.csv	Provincia-año-mes	Insumo mensual para ARIMA/SARIMAX y Temporal Fusion Transformer.
dataset_ml_provincia_anio.csv	Provincia-año	Insumo anual para XGBoost y para el índice de riesgo provincial.
tasas_provincia_anio.csv	Provincia-año	Tasas provinciales incorporadas al análisis anual.
variables_sociales_provincia_anio.csv	Provincia-año	Variables sociales agregadas y enlazadas con la base criminal.
diccionario_dataset_ml.csv	Diccionario de variables	Resumen de columnas, tipos de dato y valores nulos del dataset final.

Nota. Los archivos se generan en la carpeta procesados_ml del proyecto.

3.7.2 Variables objetivo y variables derivadas

El modelamiento consideró como variables objetivo los conteos agregados de delitos o eventos disponibles en la base final. Para fortalecer la señal temporal de los modelos, el proceso de integración generó rezagos de uno, dos y tres años, variación absoluta interanual, variación porcentual interanual y promedios móviles de dos y tres años.

Tabla 3

Variables objetivo y transformaciones utilizadas

Elemento	Configuración aplicada
Variables objetivo	total_homicidios, total_detenidos, total_desaparecidos, total_noticias y total_robos.
Periodo mínimo	Registros desde 2022.
Horizonte predictivo	Año 2026.
Variables de rezago	lag_1, lag_2 y lag_3 por provincia y variable objetivo.
Variación temporal	Variación absoluta interanual y variación porcentual interanual.
Promedios móviles	Ventanas de dos y tres años.
Predicciones negativas	Recorte inferior en cero, debido a que los eventos son conteos.

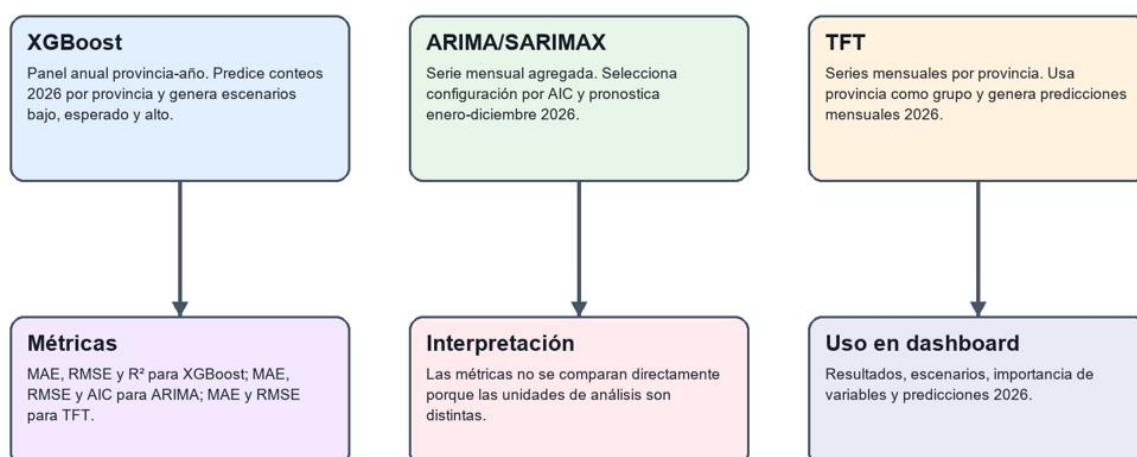
Nota. El año 2026 corresponde a escenarios predictivos exploratorios y no a datos observados.

3.7.3 Comparación metodológica de modelos

Los modelos utilizados responden a unidades de análisis distintas. Por esta razón, sus métricas no deben interpretarse como una competencia directa entre algoritmos, sino como resultados válidos dentro de la configuración específica de cada modelo.

Figura 4

Unidad de análisis y salida de cada modelo



Nota. La comparación es metodológica; no implica declarar un modelo ganador general.

Tabla 4*Resumen comparativo de configuración de modelos*

Modelo	Unidad de análisis	Archivo de entrada	Salida principal	Métricas
XGBoost	Provincia-año	dataset_ml_provincia_anio.csv	Predicción anual 2026 por provincia y escenarios.	MAE, RMSE y R^2
ARIMA/SARIMAX	Serie mensual agregada	criminalidad_mensual_con_sociales.csv	Pronóstico mensual agregado para 2026.	MAE, RMSE y AIC
Temporal Fusion Transformer	Serie mensual por provincia	criminalidad_mensual_con_sociales.csv	Predicción mensual 2026 por provincia.	MAE y RMSE
Índice de riesgo provincial	Provincia-año	Dataset anual y predicción XGBoost	Ranking, nivel de riesgo y clúster territorial.	Índice 0-100 y KMeans

Nota. Las métricas no son directamente comparables porque los modelos trabajan con escalas temporales y territoriales distintas.

3.7.4 Configuración de XGBoost

XGBoost se implementó como un modelo supervisado de regresión para predecir conteos anuales a nivel provincia-año. El modelo utiliza variables criminales agregadas, tasas provinciales, variables sociales y variables derivadas de rezagos, variaciones y promedios móviles.

Tabla 5*Configuración técnica de XGBoost*

Componente	Configuración
Algoritmo	XGBRegressor
Tipo de problema	Regresión de conteos agregados
Objetivo del algoritmo	reg:squarederror
Estimadores	500
Profundidad máxima	4
Tasa de aprendizaje	0.03
Submuestreo de filas	0.9
Submuestreo de columnas	0.9
Semilla aleatoria	42
Procesamiento numérico	Imputación con mediana
Procesamiento categórico	Imputación con moda y OrdinalEncoder
Validación	Validación temporal escalonada
Escenarios 2026	Bajo: 90%; esperado: 100%; alto: 110% de la predicción central

Nota. El modelo evita utilizar como predictoras la variable objetivo-evaluada y las demás variables objetivo-disponibles.

3.7.5 Configuración de ARIMA/SARIMAX

ARIMA/SARIMAX se utilizó como línea base estadística para series temporales mensuales agregadas. Para cada variable objetivo, la información se agrupa por año y mes, sumando los conteos mensuales.

Tabla 6*Configuración técnica de ARIMA/SARIMAX*

Componente	Configuración
Librería	statsmodels.tsa.statespace.SARIMAX
Unidad de análisis	Serie mensual agregada
Frecuencia temporal	Mensual, inicio de mes (MS)
Candidatos no estacionales	(1,1,1), (1,1,0), (0,1,1), (1,0,1)
Candidatos estacionales	(0,0,0,0), (1,0,1,12), (1,1,1,12)
Criterio de selección	Menor AIC válido
Optimizador	lbfgs
Iteraciones máximas	100
Validación	Últimos meses de la serie como tramo de prueba
Métricas	MAE, RMSE y AIC
Condición mínima	Al menos 24 meses de serie

Nota. Si ningún candidato converge, el procedimiento utiliza ARIMA(1,1,1) sin componente estacional como configuración de respaldo.

3.7.6 Configuración de Temporal Fusion Transformer

Temporal Fusion Transformer se utilizó como modelo de aprendizaje profundo para series mensuales por provincia. Este enfoque permite modelar múltiples series temporales en paralelo, usando la provincia como identificador de grupo.

Tabla 7*Configuración técnica del Temporal Fusion Transformer*

Componente	Configuración
Librería	pytorch-forecasting
Modelo	TemporalFusionTransformer
Unidad de análisis	Provincia-mes
Identificador de grupo	Provincia
Variable temporal	time_idx
Categorías estáticas	Provincia
Categorías conocidas en el tiempo	Mes (mes_str)
Variables reales conocidas	time_idx, año y mes
Normalización	GroupNormalizer por provincia
Longitud máxima de codificador	Hasta 24 meses
Horizonte de predicción	12 meses
Función de pérdida	QuantileLoss
Optimizador	Adam
Tasa de aprendizaje	0.03
Tamaño oculto	16
Cabezas de atención	2
Dropout	0.1
Épocas máximas	20
Parada temprana	Paciencia de 5 épocas
Métricas	MAE y RMSE

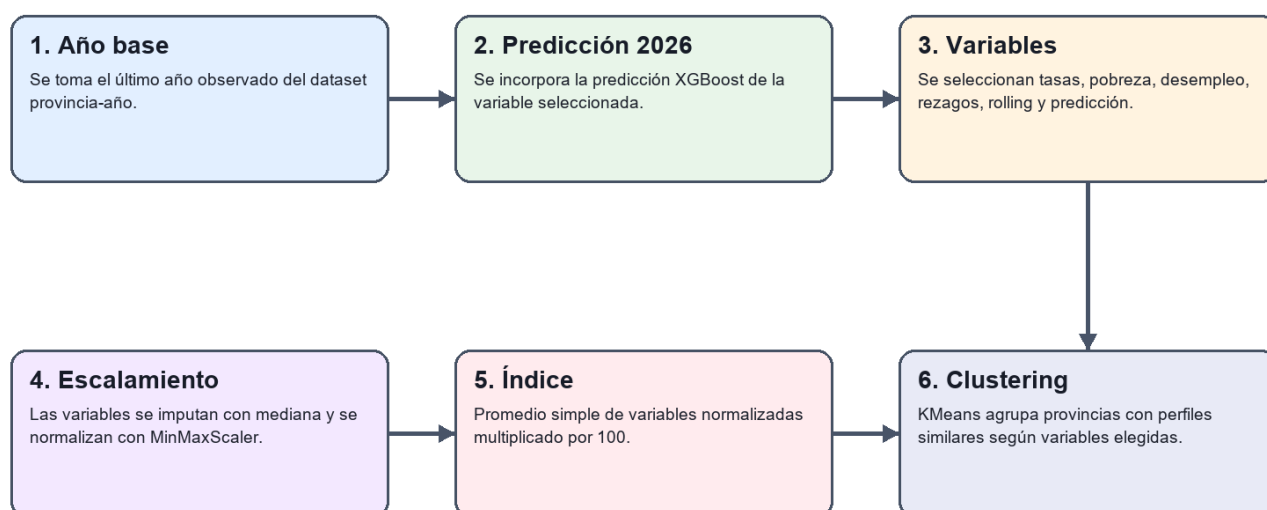
Nota. El modelo se interpreta como exploratorio debido al tamaño limitado de la serie histórica disponible.

3.7.7 Índice de riesgo provincial y clustering

El dashboard incorpora un índice de riesgo provincial configurable. Este índice toma como base el último año observado en `dataset_ml_provincia_anio.csv` y, si existe el archivo de predicciones XGBoost, incorpora la predicción 2026 de la variable seleccionada por el usuario.

Figura 5

Cálculo de índice de riesgo provincial



Nota. El índice es exploratorio y no constituye una clasificación oficial de peligrosidad.

Tabla 8

Configuración del índice de riesgo

Componente	Configuración
Base de entrada	<code>dataset_ml_provincia_anio.csv</code>
Predicción incorporada	<code>xgboost_predictions_2026.csv</code> , si está disponible
Año base	Último año observado del dataset

Variables candidatas	Predicción 2026, tasas, pobreza, desempleo, Gini, hacinamiento, servicios, rezagos, rolling y variaciones
Imputación	Mediana de cada variable
Escalamiento	MinMaxScaler
Fórmula del índice	Promedio simple de variables normalizadas multiplicado por 100
Clasificación	Bajo, medio y alto
Clustering	KMeans
Número de clústeres	Configurable entre 2 y 5
Semilla	42
Inicializaciones	n_init=10

Nota. El índice sintetiza información histórica y predictiva, pero no establece causalidad ni reemplaza el análisis contextual de cada provincia.

3.7.8 Consideraciones para la interpretación

Los resultados de los modelos deben interpretarse dentro de su propia unidad de análisis. XGBoost trabaja con panel anual provincia-año; ARIMA/SARIMAX trabaja con series mensuales agregadas; y TFT trabaja con series mensuales por provincia. Por ello, no se debe concluir que un modelo es superior a otro únicamente a partir de MAE o RMSE sin homologar previamente el horizonte, la escala territorial, el conjunto de prueba y la métrica.

Las predicciones para 2026 deben entenderse como escenarios exploratorios contruidos con datos históricos y variables disponibles; no consideran hechos futuros aún desconocidos, cambios normativos, nuevas medidas institucionales, variaciones en el registro de denuncias ni modificaciones inesperadas en la dinámica delictiva. Por esta razón, sus resultados sirven como

apoyo para el análisis, el seguimiento y la planificación; no deben asumirse como decisiones automáticas ni como pronósticos exactos.

CAPÍTULO 4: ANÁLISIS DE RESULTADOS

Este capítulo presenta los resultados obtenidos durante la integración, el análisis y el modelamiento de datos relacionados con la criminalidad y la seguridad ciudadana en el Ecuador; la información fue construida a partir de registros del ECU 911, la Fiscalía General del Estado, el Ministerio del Interior y bases complementarias del INEC, reunidas dentro de la plataforma analítica desarrollada para el estudio.

El análisis se concentra principalmente en Guayas, Pichincha y Manabí por su peso dentro de los registros nacionales y por los niveles de incidencia observados en las categorías estudiadas; para examinar estos territorios se combinaron técnicas exploratorias, geoespaciales, temporales y predictivas, con las cuales se revisaron tendencias históricas, concentraciones territoriales y posibles escenarios futuros.

También se comparó el rendimiento de modelos estadísticos, algoritmos de aprendizaje automático y técnicas de inteligencia artificial; su selección respondió a la capacidad de trabajar con series temporales, incorporar variables externas y reconocer relaciones entre factores sociales, territoriales y delictivos.

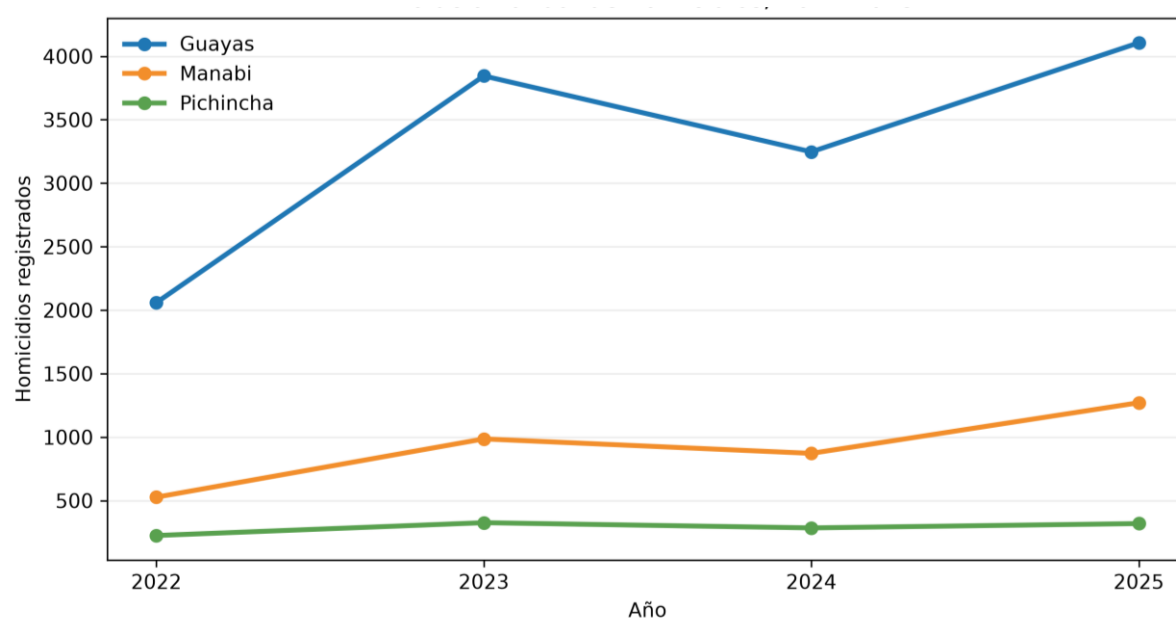
Los resultados se organizan en cuatro partes; primero se revisan la cobertura, calidad e integración de las bases, luego se presentan los patrones temporales y espaciales identificados, después se compara el desempeño de los modelos mediante distintas métricas y, al cierre, se muestran las proyecciones para 2026 junto con su interpretación.

Durante la discusión se diferencia la información observada de las estimaciones generadas por los modelos; esta separación evita presentar las proyecciones como datos reales y ayuda a

mantener una interpretación acorde con el alcance del estudio. También se consideran las limitaciones encontradas durante la validación, ya que estas influyen en la forma de comprender los resultados y formular las conclusiones.

Figura 6

Evolución anual de homicidios en Guayas, Manabí y Pichincha 2022 – 2025



Nota. Modelos de Análisis Predictivo

4.1 Aplicación de la metodología propuesta

4.1.1. Población y Muestreo

La investigación utilizó un muestreo no probabilístico por conveniencia; los registros se seleccionaron según su disponibilidad, acceso y relación con los objetivos planteados. Para ello, se revisaron las bases oficiales del ECU 911, la Fiscalía General del Estado y el Ministerio del

Interior; luego se conservaron los datos correspondientes al delito y al periodo analizado, mientras que los registros incompletos, inconsistentes o ajenos a los criterios definidos fueron excluidos.

La elección de este tipo de muestreo se justifica porque el propósito de la investigación fue desarrollar y evaluar modelos predictivos de criminalidad mediante técnicas de aprendizaje automático y análisis de series temporales, más que realizar inferencias estadísticas sobre toda la población de delitos. El uso de registros disponibles y depurados permitió contar con información confiable y homogénea para identificar patrones espacio-temporales, entrenar los modelos predictivos y evaluar su desempeño de manera consistente.

La información utilizada provino de plataformas oficiales de datos abiertos; las principales fuentes fueron el ECU 911, el Ministerio del Interior, la Fiscalía General del Estado y el Instituto Nacional de Estadística y Censos. Estas instituciones aportaron registros sobre denuncias de delitos, violencia, seguridad ciudadana y condiciones socioeconómicas relacionadas con los territorios estudiados.

La población estuvo formada por registros históricos de hechos delictivos y eventos de seguridad reportados a nivel nacional; sin embargo, solo se incluyeron bases con información suficiente sobre tipo de evento, fecha, ubicación y variables descriptivas. Los datos incompletos, inconsistentes o sin trazabilidad fueron descartados durante la depuración.

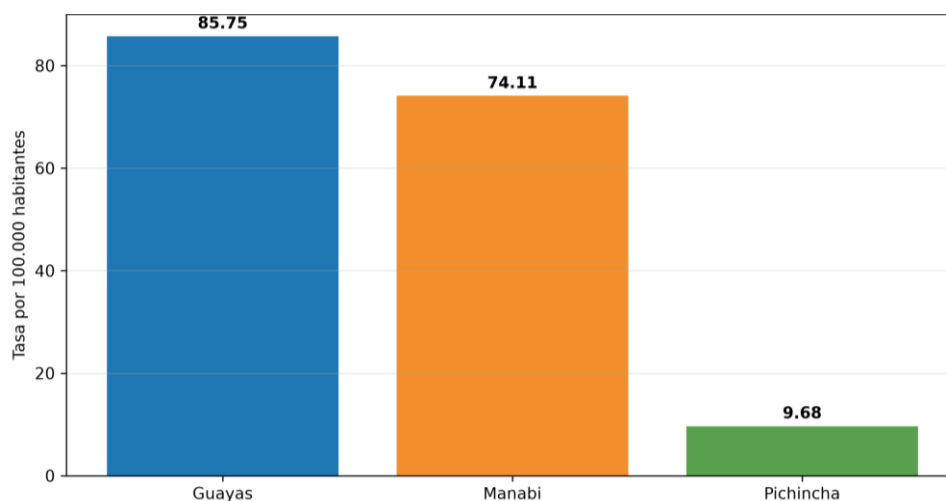
El análisis se concentró en Guayas, Pichincha y Manabí por su incidencia dentro de las categorías estudiadas y por la cantidad de información disponible; esta selección facilitó la comparación territorial y el entrenamiento de los modelos. Aun así, algunas etapas utilizaron información de otras provincias para ampliar la variabilidad de los datos.

Los registros del ECU 911 sirvieron para revisar la evolución temporal de los eventos; no obstante, se detectó la ausencia de información correspondiente a febrero y diciembre de 2023, situación considerada durante la validación. La Fiscalía aportó datos sobre denuncias, fechas y ubicación de los hechos, mientras que el Ministerio del Interior añadió variables como sexo, edad y nacionalidad.

También se incorporaron indicadores del INEC relacionados con población, pobreza multidimensional, educación y empleo; estas variables ayudaron a comparar el comportamiento delictivo con ciertas condiciones sociales y económicas. La integración de todas las fuentes produjo una base multidimensional utilizada para explorar relaciones entre criminalidad, territorio y variables socioeconómicas, así como para mejorar el entrenamiento de los modelos predictivos.

Figura 7

Tasa de homicidios en Guayas, Manabí y Pichincha 2025



Nota. Modelos de Análisis Predictivo.

4.2 Cobertura de datos y unidad de análisis

El conjunto integrado utilizado para el modelamiento anual contiene 104 observaciones provincia-año correspondientes a 26 unidades territoriales durante el periodo 2022-2025. Para responder al objetivo de la investigación, el análisis sustantivo se concentra en Guayas, Pichincha y Manabí, que reúnen 12 observaciones provincia-año en dicho periodo. Las proyecciones de 2026 se interpretan como estimaciones de escenario y no como registros observados.

Las fuentes incorporadas fueron homicidios, robos, denuncias de delitos, armas incautadas, personas desaparecidas y detenidos. Para el componente predictivo se emplearon los agregados de homicidios, robos, denuncias de delitos, armas, desapariciones, detenidos, además de tasas provinciales y variables derivadas de rezagos, variación interanual y promedios móviles.

4.3 Resultados del análisis descriptivo espacio-temporal

4.3.1 Tendencia de homicidios

Los homicidios presentan la señal temporal más crítica dentro de las tres provincias priorizadas. Guayas concentró los mayores niveles absolutos durante todo el periodo y pasó de 2.059 homicidios en 2022 a 4.106 en 2025. Manabí mostró un incremento de 527 a 1.271 eventos en el mismo periodo, mientras que Pichincha se mantuvo en un nivel absoluto menor, aunque aumentó de 225 a 319 casos. El descenso observado en 2024 no se consolidó como una tendencia sostenida, pues las tres provincias registraron incrementos en 2025. Según Carrion (2024) definió el periodo del 2022 al 2024 como la consolidación de Ecuador como el principal *hub* logístico de la cocaína en el Cono Sur. La duplicación de los homicidios en Guayas es un indicador sintomático

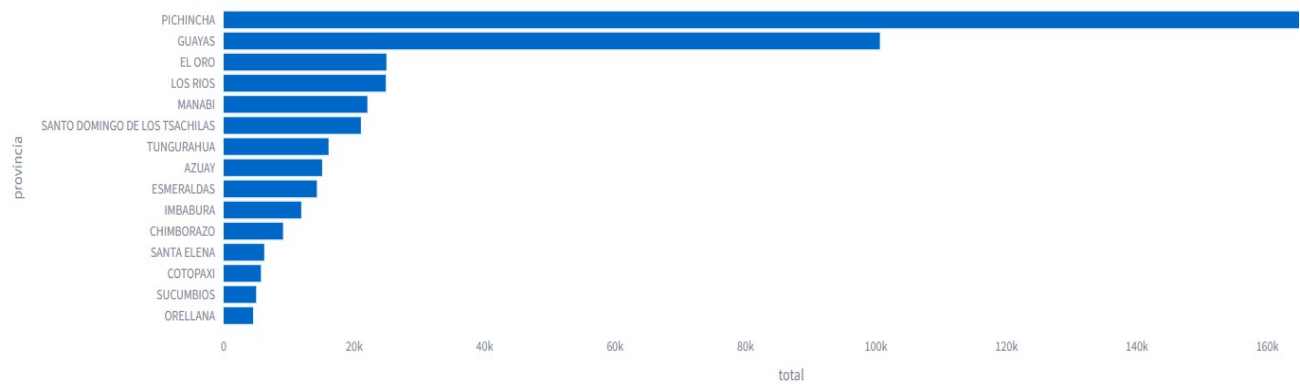
Guayas	2.059	3.845	3.247	4.106	+26,5 %
Manabí	527	986	872	1.271	+45,8 %
Pichincha	225	326	285	319	+11,9 %

Nota. Modelos de Análisis Predictivo

4.3.2 Tendencia de Robos

La distribución de los robos muestra una concentración mucho mayor en Pichincha y Guayas; Pichincha registra cerca de 160 mil casos, mientras que Guayas alcanza aproximadamente 100 mil. La diferencia frente a las demás provincias es amplia, lo que podría relacionarse con la elevada actividad urbana, comercial y poblacional de ambos territorios. El Instituto de Altos Estudios Nacionales (IAEN, 2025) señala que su concentración económica también puede atraer redes de delincuencia común; esta situación amplía la distancia respecto a otras zonas del país.

Figura 9
Robos por provincia



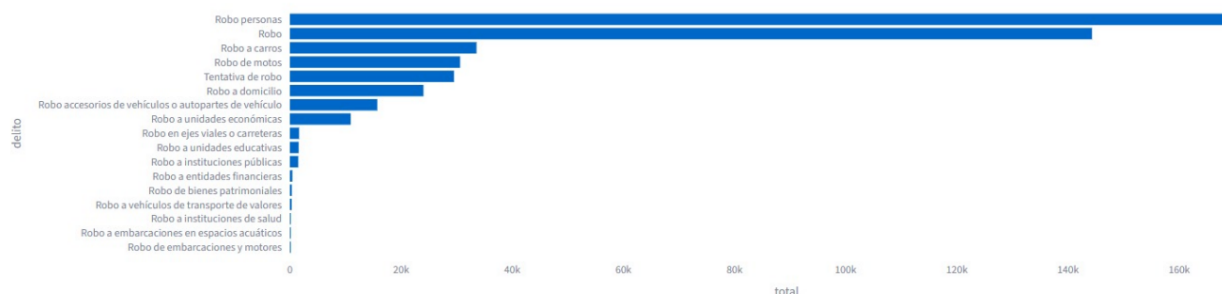
Nota. Modelos de Análisis Predictivo.

Los resultados muestran que el robo a personas fue la modalidad delictiva con mayor incidencia durante el período de estudio, alcanzando aproximadamente 170 mil registros. En segundo lugar, se ubicó el robo a domicilios, con alrededor de 145 mil casos. Las modalidades de robo de vehículos, robo de motocicletas y tentativa de robo registraron frecuencias inferiores; sin embargo, mantienen una participación significativa dentro del total de eventos delictivos analizados.

Por otro lado, modalidades como el robo a domicilio, el robo de accesorios de vehículos o autopartes, el robo a unidades económicas y el robo en ejes viales o carreteras presentan una incidencia intermedia.

Figura 10

Tipos de robos más frecuentes



Nota. Modelos de Análisis Predictivo

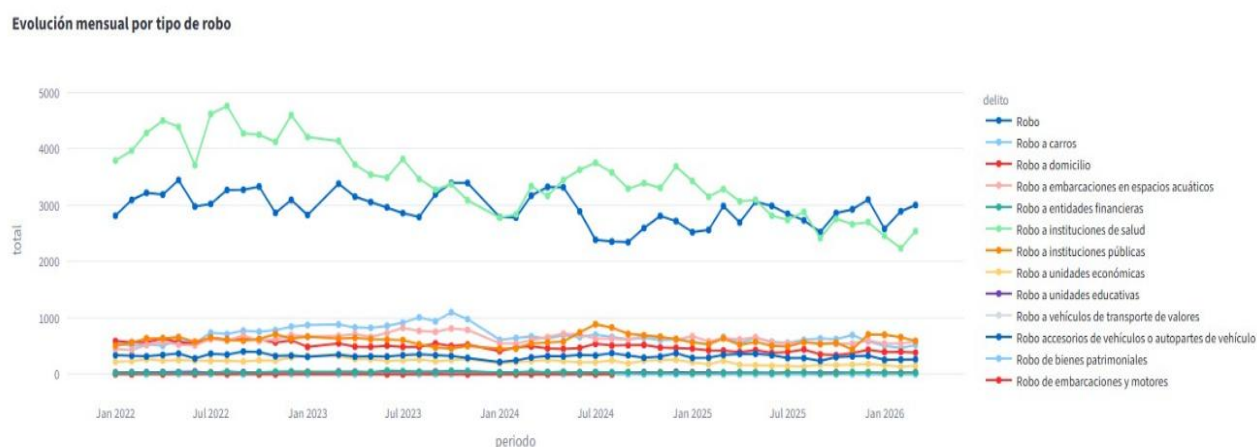
El análisis mensual entre enero de 2022 hasta diciembre de 2025 muestra que el robo y el robo a personas concentraron la mayor cantidad de casos durante casi todo el periodo; ambas categorías se mantuvieron por encima del resto, aunque presentaron variaciones de un mes a otro. El robo a personas alcanzó sus valores más altos al inicio de la serie, con más de 4.500 casos

mensuales, mientras que el robo general se ubicó cerca de los 3.000; hacia los años más recientes se observa una disminución gradual en las dos modalidades.

Las demás categorías registraron cantidades mucho menores y un comportamiento más estable; entre ellas se encuentran el robo a carros, domicilios, autopartes y unidades económicas, con cambios moderados a lo largo del tiempo. En cambio, los robos a instituciones públicas, entidades financieras, centros de salud, bienes patrimoniales y embarcaciones aparecieron con frecuencias bajas; por tanto, la mayor parte de los casos se concentró en delitos dirigidos contra las personas y sus bienes.

Figura 11

Análisis mensual por tipo de robo



Nota. Modelos de Análisis Predictivo.

La dispersión con líneas de tendencia correspondiente a la pestaña "Relaciones", donde se analiza la relación entre la variable año (eje X) y total_robos (eje Y) para las provincias de Guayas, Manabí y Pichincha durante el período 2022–2025.

Se observa una tendencia decreciente en el número de robos en las tres provincias, aunque con diferentes niveles de intensidad. Pichincha presenta los valores más altos a lo largo de todo el período, alcanzando aproximadamente entre 30.000 y 45.000 robos anuales, con una disminución gradual reflejada por la pendiente negativa de la línea de tendencia. Guayas registra valores intermedios, mostrando una reducción más pronunciada, al pasar de alrededor de 32.000 robos en 2022 a menos de 16.188 en 2025. Por su parte, Manabí mantiene los niveles más bajos, con una disminución moderada desde aproximadamente 7.000 hasta cerca de 4.000 robos en el mismo período.

Las líneas de tendencia confirman una correlación negativa entre el tiempo y el número de robos, lo que sugiere una reducción progresiva de este delito en las provincias analizadas. No obstante, se observan fluctuaciones interanuales, especialmente en Pichincha, donde los valores aumentan en 2024 hasta el 2025.

4.3.3 Lectura integrada de las fuentes

Guayas reúne los mayores registros acumulados de homicidios, denuncias de delitos y armas; por esta razón, concentra buena parte del análisis territorial. Pichincha presenta el mayor número de desapariciones entre las tres provincias, mientras que Manabí muestra un aumento de homicidios y una cantidad importante de armas registradas.

Estos datos deben revisarse de forma conjunta, pero sin asumir que una variable causa directamente a otra; su utilidad está en reconocer señales que coinciden en un mismo territorio y orientar análisis más detallados en las provincias con mayor concentración de eventos.

Tabla 10*Totales acumulados del panel integrado, 2022 – 2025*

Provincia	Homicidios	Robos	Detenidos	Desaparecidos	Denuncias de delitos	Armas
Guayas	13.257	100.000	45.274	6.431	74.332	13.408
Manabí	3.656	30.000	15.543	1.519	14.671	4.075
Pichincha	1.155	160.000	37.543	7.371	27.843	2.920

Nota. Modelos de Análisis Predictivo

Los resultados muestran diferencias claras entre las provincias analizadas; Pichincha registra la mayor cantidad de robos, con 160.000 casos, y de personas desaparecidas, con 7.371. Guayas concentra los valores más altos de homicidios, detenidos, denuncias de delitos y armas incautadas; Manabí presenta cifras menores en la mayoría de las categorías, aunque mantiene registros importantes de homicidios y robos.

Al incorporar variables sociales y económicas, se observa que la dinámica delictiva podría relacionarse con la pobreza multidimensional, el empleo, el desempleo, las dificultades de acceso a una educación de calidad, las condiciones de vivienda y el gasto social; estas situaciones afectan con mayor fuerza a zonas urbanas densamente pobladas, como varias áreas de Guayas y Pichincha.

Los resultados indican que la criminalidad no depende únicamente de las acciones de seguridad; también podría estar vinculada con condiciones sociales y económicas. Por esta razón, incluir variables externas en los modelos ayuda a comprender mejor las variaciones delictivas y a generar estimaciones más completas para la planificación de políticas públicas de prevención.

4.4 Evaluación de los modelos predictivos

Se aplicaron tres modelos según el tipo de información disponible; XGBoost se utilizó con datos provinciales anuales, ARIMA y SARIMAX con series mensuales agregadas, mientras que TFT trabajó con series mensuales por provincia. Las métricas deben revisarse dentro de la estructura de cada modelo; por ello, sus errores no pueden compararse como si todos evaluaran el mismo problema, ya que cambian el nivel de agregación, el horizonte de predicción y la forma de validación.

Tabla 11

XGBoost promedios de validaciones 2023, 2024 y 2025

Variable objetivo	MAE promedio	RMSE promedio	R ² promedio	Interpretación
Homicidios	94,44	279,29	0,84	Desempeño explicativo aceptable en validaciones anuales.
Robos	3.900	14.200		Estos resultados indican una menor precisión predictiva y evidencian dificultades para capturar la alta variabilidad y complejidad temporal de esta serie.
Desaparecidos	72,37	246,79	0,66	Capacidad predictiva moderada; requiere seguimiento.
Noticias	620,75	1.445,76	0,84	Buen ajuste relativo, condicionado por la escala del conteo.
Armas	53,92	104,04	0,98	Mayor estabilidad entre los objetivos evaluados.
Detenidos	830,69	1.594,01	No interpretable	Serie 2025 incompleta; se excluye de conclusiones.

Nota. Modelos de Análisis Predictivo

Para el objetivo de homicidios, XGBoost alcanzó un R^2 promedio de 0,84. Las variables de mayor importancia fueron el promedio móvil de tres años de homicidios (0,507), el promedio móvil de denuncias de delitos (0,276) y el rezago de un año de homicidios (0,119). Este patrón indica que la dinámica reciente de los homicidios aporta la mayor señal predictiva, mientras que las variables asociadas a denuncias de delitos y armas complementan la explicación del comportamiento anual.

Tabla 12

TFT y ARIMA en homicidios, desaparecidos y Armas

Modelo	Unidad de análisis	Homicidios MAE	Robos MAE	Desaparecidos MAE	Armas MAE	Observación
TFT	Provincia- mes	9,24	20	5,40	8,60	48 meses y 26 series provinciales.
ARIMA/SARIMAX	Serie mensual agregada	193,83	1.700	46,65	152,76	Útil como referencia temporal agregada.

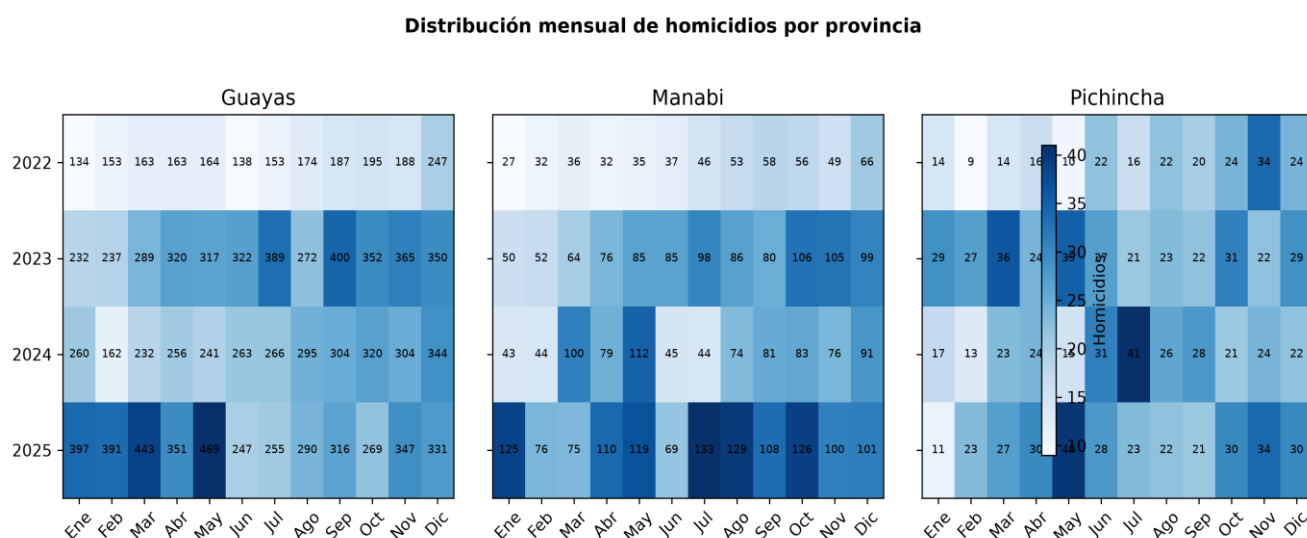
Nota. Modelos de Análisis Predictivo

El TFT reportó errores medios absolutos menores en los objetivos de homicidios, desapariciones y armas dentro de su configuración provincial-mensual. No obstante, esta ventaja debe presentarse como evidencia preliminar y no como una superioridad concluyente frente a ARIMA/SARIMAX, porque las métricas provienen de esquemas de agregación distintos. Para una

comparación definitiva se recomienda evaluar todos los modelos con el mismo conjunto de prueba, el mismo horizonte temporal y una métrica normalizada, por ejemplo, MAPE o MAE escalado.

Figura 12

Distribución mensual de homicidios por provincia, 2022 – 2025



Nota. Modelos de Análisis Predictivo.

4.5 Proyección para 2026

Las proyecciones de XGBoost para 2026 se presentan como escenarios de planificación; en homicidios, Guayas, Manabí y Pichincha alcanzarían una estimación conjunta de 5.673,74 casos. El escenario bajo se ubica en 5.106,37 y el alto en 6.241,11; Guayas aporta la mayor parte del total estimado, seguida de Manabí y Pichincha.

Tabla 13*XGBoost Predicciones 2026*

Provincia	Predicción central	Escenario bajo	Escenario alto	Cambio vs. 2025
Guayas	4.006,10	3.605,49	4.406,71	-2,4 %
Manabí	1.339,94	1.205,95	1.473,93	+5,4 %
Pichincha	327,70	294,93	360,47	+2,7 %

Nota. Modelos de Análisis Predictivo.

La proyección indica que Guayas podría mantener un nivel alto de homicidios durante 2026, aunque el escenario central muestra una leve reducción frente a 2025; Manabí presenta una posible tendencia al alza que requiere seguimiento, mientras que Pichincha conserva valores menores en comparación con las otras dos provincias. Estas estimaciones no deben asumirse como resultados exactos; reflejan el comportamiento esperado según los datos históricos incluidos en el modelo y podrían cambiar por nuevas medidas institucionales, variaciones en las denuncias, reformas normativas o hechos no registrados durante el entrenamiento.

4.6 Análisis de riesgo provincial

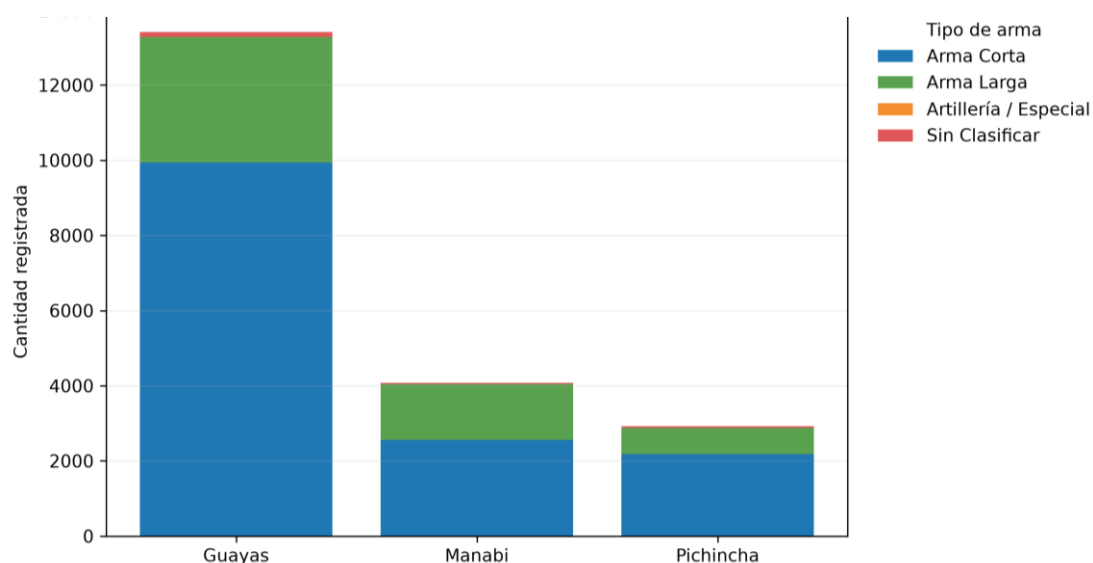
La plataforma incorpora un índice de riesgo provincial construido con variables delictivas, tasas, rezagos, promedios móviles y predicciones; este índice organiza las provincias según comportamientos similares y forma grupos mediante técnicas de clustering. Su función es apoyar la priorización del análisis, no asignar etiquetas definitivas a los territorios ni a su población.

Guayas, Manabí y Pichincha se analizan como un grupo de contraste; Guayas concentra el mayor número de homicidios y denuncias de delitos asociadas con hechos delictivos, Manabí

muestra un aumento reciente de homicidios y Pichincha registra cifras menores en esta categoría, aunque reúne una cantidad elevada de desapariciones. Para interpretar el índice también deben considerarse las tasas poblacionales, la calidad de las bases y las características propias de cada provincia.

Figura 13

Composición de armas incautadas por provincia, 2022 – 2025



Nota. Modelos de Análisis Predictivo.

4.7 Limitaciones y consideraciones para la interpretación

El periodo utilizado para entrenar los modelos comprende de 2022 a 2025; los valores de 2026 corresponden a proyecciones y no a registros observados. Por ello, el título y los objetivos deben diferenciar con claridad el periodo analizado y el horizonte estimado.

La base de personas detenidas presenta una cobertura incompleta para 2025; esta condición reduce la solidez de cualquier comparación o proyección relacionada con dicha categoría. Además, los modelos trabajaron con unidades de análisis distintas; por esta razón, no sería adecuado escoger un único modelo ganador comparando solamente los valores de MAE o RMSE.

Los modelos desarrollados son de regresión y estiman cantidades de eventos; en este caso, métricas como precisión, recall, F1-score y ROC-AUC no corresponden al tipo de problema analizado. Para evaluar su rendimiento se utilizaron MAE, RMSE, MAPE, R^2 y validación temporal escalonada; esta decisión mantiene coherencia entre los datos, la metodología y las conclusiones presentadas.

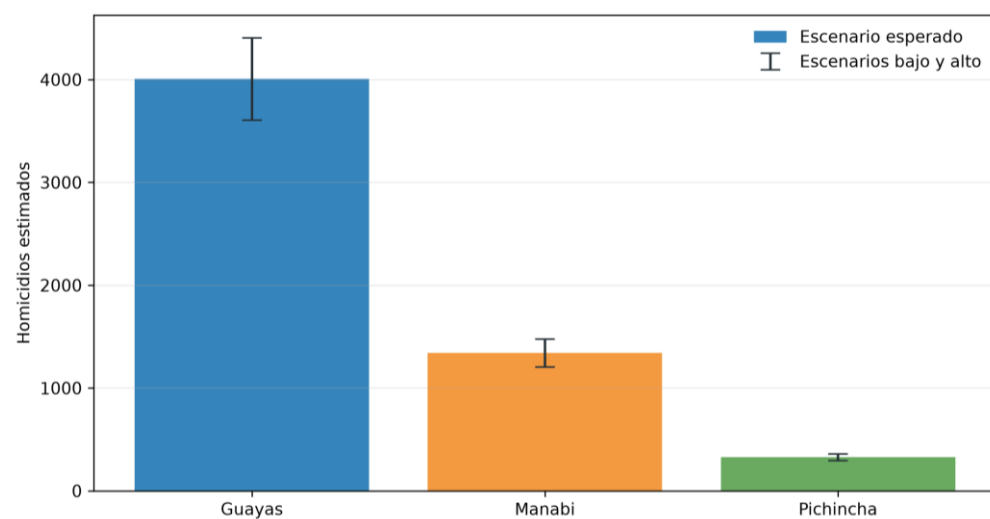
4.8 Síntesis de resultados

Los resultados muestran diferencias territoriales y temporales en los eventos de seguridad ciudadana; Guayas mantiene el mayor volumen de homicidios, Manabí registra un aumento durante 2025 y en la proyección de 2026, mientras que Pichincha presenta cifras menores de homicidios, aunque acumula una cantidad elevada de desapariciones.

XGBoost obtuvo resultados útiles para homicidios, denuncias de delitos y armas dentro del panel anual; por su parte, TFT mostró un comportamiento favorable en las series mensuales por provincia. La plataforma reúne estos resultados en un espacio de consulta, seguimiento y comparación territorial; aun así, su interpretación debe considerar la calidad, cobertura y disponibilidad de las bases utilizadas.

Figura 14

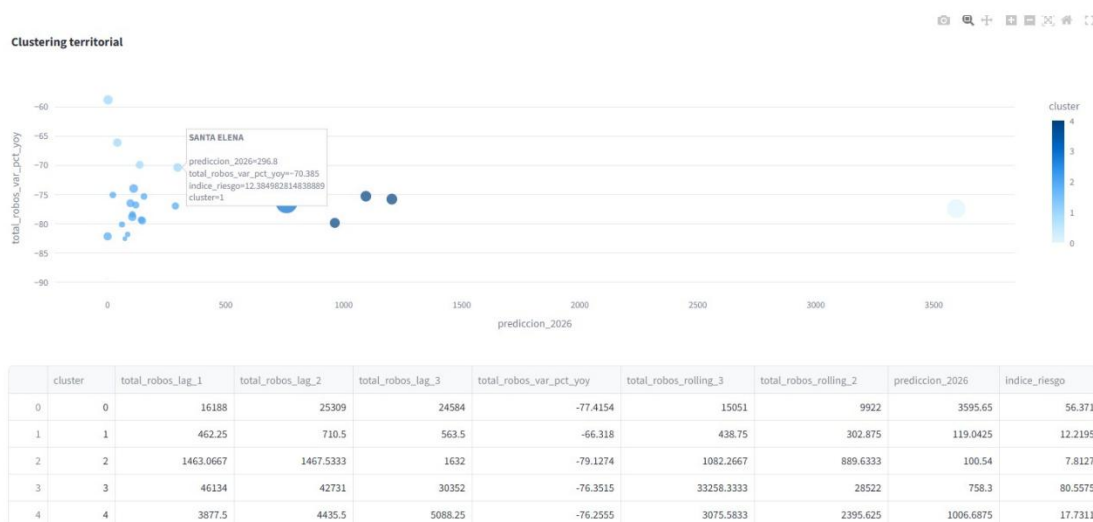
Predicción de homicidios para 2026 con escenarios XGBoost



Nota. Modelos de Análisis Predictivo.

El análisis de clustering agrupó las provincias del Ecuador según sus registros históricos y las proyecciones de robos para 2026; se identificaron cinco grupos con comportamientos y niveles de riesgo diferentes. El clúster de mayor riesgo reúne las provincias con más robos acumulados y estimaciones futuras más altas, mientras que los grupos de menor riesgo incluyen territorios con menos casos y una posible tendencia a la baja.

Esta clasificación ayuda a comparar provincias y reconocer cuáles requieren mayor atención; además, puede servir como apoyo para organizar medidas preventivas, distribuir recursos y orientar políticas de seguridad basadas en datos.

Figura 15*Clustering Territorial*

Nota. Modelos de Análisis Predictivo.

La categoría de robos presentó los errores más altos entre las variables analizadas; esto refleja una mayor variación en la serie y cierta dificultad para reconocer un comportamiento estable. XGBoost obtuvo un MAE cercano a 3.900 y un RMSE aproximado de 14.200, mientras que ARIMA/SARIMAX alcanzó un MAE de alrededor de 1.700 y un RMSE próximo a 2.000.

Dentro de esta serie, ARIMA/SARIMAX mostró un mejor ajuste que XGBoost; aun así, los errores obtenidos indican que los robos siguen siendo difíciles de modelar. Su comportamiento podría cambiar por factores sociales, económicos, territoriales y por variaciones en el registro de denuncias; por ello, las predicciones deben interpretarse con cautela.

CAPÍTULO 5: CONCLUSIONES Y RECOMENDACIONES

5.1 Conclusiones

- La integración y depuración de las bases del ECU 911, la Fiscalía General del Estado y el Ministerio del Interior ayudó a construir un repositorio unificado; se corrigieron nombres de variables, códigos, fechas y unidades de medida, lo que facilitó la comparación entre fuentes y la incorporación de tasas provinciales e indicadores derivados.
- El preprocesamiento ayudó a detectar registros incompletos, duplicados y diferencias entre las estructuras de las bases; luego de la limpieza y estandarización, la información quedó preparada para estudiar variaciones espaciales y temporales. Aun así, la cobertura de los datos condicionó los resultados; esta situación se observó con mayor claridad en los registros de personas detenidas durante 2025.
- El análisis territorial mostró comportamientos distintos entre las provincias; Guayas concentró los valores más altos de homicidios, denuncias de delitos y armas incautadas, Manabí registró el mayor aumento relativo de homicidios entre 2024 y 2025, mientras que Pichincha acumuló una cantidad elevada de personas desaparecidas. Estas diferencias indican que las medidas de seguridad deberían adaptarse a las características de cada territorio; además, los valores absolutos deben revisarse junto con las tasas provinciales.
- La incorporación de variables sociales y económicas amplió la interpretación del comportamiento delictivo; su integración ayudó a examinar posibles relaciones entre criminalidad, condiciones territoriales y características de la población, sin asumir vínculos causales directos.
- Los modelos de aprendizaje automático generaron estimaciones sobre la evolución de varios delitos; XGBoost obtuvo resultados favorables en homicidios, denuncias de delitos y armas incautadas dentro del análisis anual por provincia. Para homicidios alcanzó un R^2 promedio de 0,84; además, los rezagos y promedios móviles aportaron información útil durante el entrenamiento.
- Temporal Fusion Transformer mostró resultados favorables en la predicción mensual por

provincia, mientras que ARIMA/SARIMAX funcionó como referencia para las series mensuales agregadas; en robos, ARIMA/SARIMAX registró errores menores que XGBoost. Sin embargo, no corresponde afirmar que un modelo supera a los demás en todos los casos; cada uno trabaja con escalas, estructuras y objetivos diferentes.

- La evaluación mostró que las métricas deben revisarse junto con la cobertura y representatividad de las bases; aunque algunos modelos alcanzaron un rendimiento adecuado, series históricas más amplias, nuevas variables y una mayor desagregación territorial podrían mejorar las estimaciones posteriores.
- Como producto aplicado, se desarrolló una plataforma de analítica criminológica que reúne preparación de datos, análisis exploratorio, mapas, predicciones y evaluación territorial; el dashboard incorpora filtros, indicadores, rankings y opciones de descarga que facilitan la consulta de resultados por parte de usuarios académicos y técnicos.
- El índice de riesgo provincial reunió registros históricos, tasas territoriales y proyecciones dentro de una sola medida; su uso está orientado a ordenar provincias y reconocer áreas que requieren una revisión más detallada. No representa una clasificación definitiva de peligrosidad; funciona como apoyo para organizar recursos de análisis y orientar medidas preventivas.

5.2 Recomendaciones

- Se recomienda revisar las bases antes de actualizar el dashboard o volver a entrenar los modelos; esta revisión debe considerar periodos disponibles, valores nulos, duplicados, coordenadas, nombres territoriales y continuidad de las series. La base de detenidos requiere mayor atención por la información incompleta de 2025.
- También conviene registrar la fuente, fecha de descarga, variables incluidas y cambios aplicados en cada archivo; este control ayuda a repetir el análisis, revisar versiones y detectar modificaciones en las bases públicas.
- Para comparar los modelos de forma más clara, se debería trabajar con el mismo periodo de prueba, unidad territorial y métricas compatibles; una validación temporal escalonada

serviría para evaluar XGBoost, ARIMA/SARIMAX y TFT bajo condiciones similares. En regresión deben usarse MAE, RMSE, MAPE y R^2 ; las métricas de clasificación solo corresponden si el problema cambia a categorías de riesgo.

- Las predicciones deberían incluir rangos de incertidumbre, pruebas de sensibilidad y una referencia sencilla; por ejemplo, comparar los resultados con el valor del último periodo disponible ayudaría a conocer si los modelos complejos ofrecen una mejora real.
- Los rankings y proyecciones deben utilizarse como apoyo para la planificación, no como decisiones automáticas; cualquier priorización territorial requiere revisión técnica y conocimiento de las condiciones locales. También se debe evitar etiquetar comunidades o realizar interpretaciones sobre personas concretas.
- Como mejora futura, podría añadirse un módulo de densidad geográfica cuando las coordenadas tengan suficiente calidad; el análisis por horas debería aplicarse solo en las fuentes con horarios completos, mientras que el tiempo de localización de personas desaparecidas requiere fechas consistentes.
- Otra mejora sería automatizar la actualización de las bases y guardar un registro de cada proceso de integración y entrenamiento; las pruebas de calidad, el control de versiones y el almacenamiento de parámetros ayudarían a mantener la reproducibilidad del sistema.
- Estudios posteriores podrían incorporar información de población, movilidad, condiciones urbanas, infraestructura y variables económicas con mayor frecuencia; también sería conveniente evaluar modelos espaciales o jerárquicos que consideren la relación entre provincias y cantones.
- Cuando se disponga de datos observados de 2026, será posible comparar las proyecciones con los valores reales y ajustar los modelos; esta revisión servirá para conocer su rendimiento fuera del periodo de entrenamiento.
- El proyecto muestra que se puede construir una plataforma de análisis criminológico a partir de fuentes públicas distintas, visualizaciones y modelos predictivos; aunque los problemas de cobertura y calidad siguen presentes, la solución los hace visibles y ofrece una base reproducible para el seguimiento territorial y la planificación basada en datos.

REFERENCIAS BIBLIOGRÁFICAS

- Alarcón Arévalo, A. A. (2024). *Modelo de machine learning para la predicción de violencia de género en Ecuador* [Tesis de maestría, Universidad Tecnológica ECOTEC].
- Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, 27(2), 93–115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., y Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley.
- Carrión, F. (2025). Gobernanza criminal y mutaciones de la violencia urbana en el Ecuador. FLACSO Ecuador.
- Chainey, S., y Ratcliffe, J. H. (2005). *GIS and crime mapping*. Wiley.
- Chávez-Cadena, M. I., y Abril-Donoso, M. E. (2023). Predicción de homicidios desde la regresión logística en el Distrito Metropolitano de Quito, Ecuador. *Revista Arbitrada Interdisciplinaria Koinonía*, 8(16), 238–255. <https://doi.org/10.35381/r.k.v8i16.2551>
- Chen, T., y Guestrin, C. (2016). XGBoost: A scalable tree boosting system. En *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939785>
- Comisión Económica para América Latina y el Caribe. (2024). *The portal of inequalities in Latin America*. <https://www.cepal.org/en/publications/69191-portal-inequalities-latin-america>
- Díaz, C., & Hernández, R. (2019). Modelos predictivos y análisis espacio-temporal aplicados a la gestión de emergencias urbanas. *Revista Iberoamericana de Análisis de Datos*, 5(1), 33–48.
- Ester, M., Kriegel, H.-P., Sander, J., y Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. En E. Simoudis, J. Han, y U. Fayyad (Eds.), *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (pp. 226–231). AAAI Press.

- Gómez, M., & Ramírez, S. (2023). Patrones espaciales y temporales en la gestión de incidentes urbanos en áreas metropolitanas. *Revista Latinoamericana de Seguridad Ciudadana*, 11(2), 77–92
- Guerrero, R. (1997). Epidemiología de la violencia: El caso de Cali, Colombia. En L. Ratinoff (Ed.), *Hacia un desarrollo integrado: Ética, violencia y seguridad ciudadana* (pp. 41–64). Banco Interamericano de Desarrollo.
- Guzmán, Á. (Ed.). (1994). *Ciudad y violencias en América Latina*. Programa de Gestión Urbana. Instituto de Altos Estudios Nacionales [IAEN]. (2025). Informe de coyuntura: Seguridad ciudadana y asimetrías delictivas en el territorio ecuatoriano. Dirección de Investigación
- Kang, H.-W., y Kang, H.-B. (2017). Prediction of crime occurrence from multi-modal data using deep learning. *PLOS ONE*, 12(4), e0176244.
<https://doi.org/10.1371/journal.pone.0176244>
- LeCun, Y., Bengio, Y., y Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
<https://doi.org/10.1038/nature14539>
- Lim, B., Arık, S. Ö., Loeff, N., y Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4), 1748–1764. <https://doi.org/10.1016/j.ijforecast.2021.03.012>
- Mao, L., Du, W., Wen, S., Li, Q., Zhang, T., y Zhong, W. (2025). Crime forecasting: A spatio-temporal analysis with deep learning models. *Journal of Computational Methods in Sciences and Engineering*. Advance online publication.
<https://doi.org/10.1177/14727978251337993>
- Medina Ariza, J. J. (2010). *Políticas y estrategias de prevención del delito y seguridad ciudadana*. Edisofer.
- Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., y Tita, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493), 100–108. <https://doi.org/10.1198/jasa.2011.ap09546>

- Ordóñez-Eraso, H. A., Pardo-Calvache, C. J., y Cobos-Lozada, C. A. (2020). Detección de tendencias de homicidios en Colombia usando machine learning. *Revista Facultad de Ingeniería*, 29(54), e11740. <https://doi.org/10.19053/01211129.v29.n54.2020.11740>
- Palomeque Vallejo, E. (2001). *Diagnóstico sobre seguridad ciudadana en Ecuador: Un paso hacia la definición de políticas públicas*. FLACSO Ecuador.
- Perry, W. L., McInnis, B., Price, C. C., Smith, S. C., y Hollywood, J. S. (2013). *Predictive policing: The role of crime forecasting in law enforcement operations*. RAND Corporation. <https://doi.org/10.7249/RR233>
- Provost, F., y Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media.
- Ribeiro, J., Meneses, L., Costa, D., Miranda, W., y Alves, R. (2022). Prediction of homicides in urban centers: A machine learning approach. En K. Arai (Ed.), *Intelligent systems and applications: Proceedings of the 2021 Intelligent Systems Conference (IntelliSys)*, volume 3 (Vol. 295, pp. 344–361). Springer. https://doi.org/10.1007/978-3-030-82199-9_22
- Rosero Aguas, M. Á. (2025). Modelo predictivo de criminalidad en Ecuador mediante la comparación de Temporal Fusion Transformer (TFT), XGBoost y ARIMA [Tesis de maestría]. Universidad Alfonso X el Sabio.
- Sherman, L. W., Gartin, P. R., y Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology*, 27(1), 27–56. <https://doi.org/10.1111/j.1745-9125.1989.tb00862.x>
- Sousa, S. B. da S., Del-Fiaco, R. de C., y Berton, L. (2018). *Cluster analysis of homicide rates in the Brazilian state of Goiás from 2002 to 2014* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.1811.06366>
- Stalidis, P., Semertzidis, T., y Daras, P. (2021). Examining deep learning architectures for crime classification and prediction. *Forecasting*, 3(4), 741–762. <https://doi.org/10.3390/forecast3040046>

- Torres, L., & Gutiérrez, S. (2021). Análisis espacio-temporal para la identificación de zonas de riesgo en contextos urbanos. *Revista de Estudios en Seguridad y Sociedad*, 10(1), 23–38.
- Wang, T., Rudin, C., Wagner, D., y Sevieri, R. (2017). Learning to detect patterns of crime. En H. Blockeel, K. Kersting, S. Nijssen, y F. Železný (Eds.), *Machine learning and knowledge discovery in databases* (Lecture Notes in Computer Science, Vol. 8190, pp. 515–530). Springer. https://doi.org/10.1007/978-3-642-40994-3_33
- Yu, C.-H., Ward, M. W., Morabito, M., y Ding, W. (2011). Crime forecasting using data mining techniques. En *2011 IEEE 11th International Conference on Data Mining Workshops* (pp. 779–786). IEEE. <https://doi.org/10.1109/ICDMW.2011.56>

APÉNDICES

Apéndice A

Enlace al aplicativo

Aplicativo:

https://mailinternacionaledu-my.sharepoint.com/:f:/g/personal/mamayorgama_uide_edu_ec/IgAprjspLnONT7eLdx1yusfsAZzNhxoSptjX2IjsIf6E5AM?e=jIPfJu

Apéndice B

Documentación técnica

Documentación:

<https://mailinternacionaledu-my.sharepoint.com/my?id=%2Fpersonal%2Fimvizcainoim%5Fuide%5Fedu%5Fec%2FDocuments%2FTesisPrimeraEntregayviewid=e88d9132%2D2ffd%2D4e4f%2Db99c%2Daca502151750>