

ORIGINAL ARTICLE

Spatiotemporal analysis of homicides in Ecuador: Application of DBSCAN and Random Forest for hotspot detection***Análisis espaciotemporal de homicidios en Ecuador: aplicación de DBSCAN y Random Forest para la detección de zonas calientes***José Alberto León Alarcón¹  ¹Universidad Técnica de Manabí, Ecuador.

How to cite: León Alarcón, J.A. (2026). Spatiotemporal analysis of homicides in Ecuador: Application of DBSCAN and Random Forest for hotspot detection. *Revista San Gregorio*, 1(65), 1-10. <http://dx.doi.org/10.36097/rsan.v1i65.3472>

Received: 06-03-2025

Accepted: 08-10-2025

Published: 31-03-2026

ABSTRACT

The objective of this research was to develop a geospatial and predictive analysis model to identify high-incidence homicide areas in Ecuador for strategic planning in public security. The geospatial distribution of homicides in the country during 2024 was analyzed using the official database provided by the Ministry of the Interior, with a particular focus on the provinces of Guayas and Manabí, which report the highest rates of violence. A combined methodology was implemented, integrating the unsupervised clustering algorithm DBSCAN to detect spatial homicide clusters and the supervised prediction model Random Forest to anticipate the likelihood of new cases belonging to the identified hot zones. The analysis revealed significant crime clusters in urban and peri-urban areas, as well as the presence of scattered homicides classified as noise, suggesting complex or atypical patterns. The predictive model achieved high levels of accuracy (F1-score of 0.9987), highlighting the importance of geospatial and demographic variables in classification. The results are consistent with national and international research and confirm that spatiotemporal analysis using machine learning techniques provides valuable tools for targeting security interventions.

Keywords: Homicides; Geospatial analysis; Machine learning; Criminality; Ecuador; Public security.

RESUMEN

El objetivo de esta investigación fue desarrollar un modelo de análisis geoespacial y predictivo que permita identificar zonas de alta incidencia de homicidios en Ecuador para la planificación estratégica en materia de seguridad pública. Se analizó la distribución geoespacial de los homicidios en el país durante el año 2024, utilizando como fuente la base oficial del Ministerio del Interior, con un enfoque particular en las provincias de Guayas y Manabí, por concentrar los mayores índices de violencia. Se implementó una metodología combinada que integra el algoritmo de clustering no supervisado DBSCAN, para la detección de clústeres espaciales de homicidios, y el modelo de predicción supervisado Random Forest, para anticipar la pertenencia de nuevos casos a las zonas calientes identificadas. El análisis evidenció la presencia de agrupamientos delictivos significativos en zonas urbanas y periurbanas, así como la existencia de homicidios dispersos clasificados como ruido, lo que sugiere patrones complejos o atípicos. El modelo predictivo alcanzó altos niveles de precisión (F1-score de 0.9987), resaltando la relevancia de las variables geoespaciales y demográficas en la clasificación. Los resultados son consistentes con investigaciones nacionales e internacionales y confirman que el análisis espaciotemporal mediante técnicas de aprendizaje automático ofrece herramientas valiosas para focalizar intervenciones de seguridad.

Palabras clave: Homicidios; Análisis geoespacial; Aprendizaje automático; Criminalidad; Ecuador; Seguridad pública.



INTRODUCCIÓN

Los homicidios representan uno de los problemas más críticos para la seguridad pública en América Latina, y Ecuador no es ajeno a esta realidad. En los últimos años, Ecuador ha experimentado una creciente preocupación por la seguridad pública debido al aumento en la tasa de homicidios. En 2024, la tasa de homicidios en Ecuador disminuyó en 7,42 muertes por cada 100.000 habitantes en comparación con el año anterior. De acuerdo con los datos oficiales de la Policía Nacional, 6.964 homicidios se registraron en todo el país durante 2024, lo que equivale a una tasa de 38,76 homicidios por cada 100.000 habitantes. En promedio, una persona fue asesinada cada una hora y 15 minutos en Ecuador durante 2024 (González, 2024).

A pesar de esta reducción reciente, la violencia sigue siendo uno de los desafíos estructurales más graves y persistentes en Ecuador. Este fenómeno se ha vuelto tan habitual que la mayoría de los ecuatorianos los perciben como la principal amenaza para la convivencia ciudadana. La sociedad se ha adaptado a su presencia, al punto de que los tiroteos en las calles ya no sorprenden y, poco después de ocurrir, la rutina diaria continúa con aparente normalidad (Ojeda Segovia, 2023).

Para comprender mejor esta problemática, es importante analizar la forma en que la violencia se manifiesta. Como señala Carrión Mena (2023), la violencia abarca diversas formas y contextos, incluyendo motivos de carácter interpersonal, económico, social y político. Los homicidios, por ejemplo, ocurren por múltiples motivos, ya sea como resultado de conflictos interpersonales (violencia común), por encargo en el contexto de mercados ilegales (sicariato), por razones de discriminación de género (violencia de género) o con fines políticos (magnicidio). En todos estos casos, los asesinatos se caracterizan por su planificación y ejecución deliberada, adaptándose a diferentes contextos de violencia.

Las causas detrás del aumento de la violencia en Ecuador también son variadas y complejas. El crimen organizado se destaca como un factor clave en este escenario. Actividades ilícitas como el narcotráfico, la trata de personas, el tráfico de armas y el tráfico de órganos generan enormes beneficios económicos, que a su vez fomentan la corrupción en diferentes niveles del poder público. Entre estas actividades ilegales, el narcotráfico ocupa un lugar predominante, siendo una fuente principal de ingresos por la producción, distribución y venta de sustancias controladas. Esta dinámica alimenta el conflicto entre cárteles, quienes compiten por el control de territorios estratégicos para el tráfico y microtráfico de estas sustancias (Leonel et al., 2022).

Otra de las causas directamente vinculadas al aumento de la violencia en Ecuador es el desempleo. De acuerdo con Estévez-Calapaqui (2021), un aumento porcentual en el desempleo incide directamente en el incremento de los delitos contra la propiedad privada. Por ello, ratifica que la generación de empleo, especialmente para jóvenes desempleados, es fundamental para enfrentar la violencia.

En este contexto, Fernando et al. (2024) ofrecen un análisis matizado de las dinámicas territoriales de la violencia en Ecuador. Mientras que algunas subzonas han experimentado una reducción en la tasa de homicidios intencionales, otras han visto un incremento. Esta variación podría estar vinculada a factores socioeconómicos, como el desempleo y la pobreza, así como a la efectividad de las intervenciones en políticas de seguridad a nivel local. Aunque la tendencia nacional apunta a una disminución de homicidios, es crucial enfocar los esfuerzos en las áreas con mayores aumentos para abordar los factores que impulsan esta violencia.

Con el fin de optimizar estos esfuerzos, es fundamental identificar las denominadas “zonas calientes” de homicidios. A pesar de que los métodos tradicionales de análisis presentan limitaciones, el uso de técnicas de análisis de datos y aprendizaje automático ha demostrado ser una herramienta poderosa para mejorar la precisión y la efectividad de estas estrategias.

Este estudio propone una perspectiva innovadora que combina técnicas de *clustering* no supervisado, específicamente el algoritmo DBSCAN (Density-Based Spatial Clustering of Applications with Noise), que permite detectar clústeres de forma automática, con modelos de predicción supervisados como Random Forest. Esto facilita la identificación de zonas de alto riesgo.

Inicialmente, se exploraron otros algoritmos de *clustering*, como *K-means* o *K-nearest neighbors* (k-NN), dado que también son métodos no supervisados. Sin embargo, DBSCAN es particularmente útil en este caso debido a su capacidad para detectar patrones en espacios geográficos con densidades variables, sin requerir una cantidad fija de clústeres, como lo harían K-means o k-NN. Además, DBSCAN distingue entre puntos de ruido y clústeres densos, lo que resulta crucial para identificar áreas de alta criminalidad frente a zonas de baja incidencia.

El objetivo de esta investigación es desarrollar un modelo de análisis geoespacial y predictivo que permita identificar zonas de alta incidencia de homicidios en Ecuador para la planificación estratégica en materia de seguridad pública. Se busca contribuir al entendimiento de los factores subyacentes que influyen en la distribución espacial de los homicidios en el país.

METODOLOGÍA

En el presente estudio se utilizó como fuente de información la base de datos ofrecida por el Ministerio del Interior del Ecuador, la cual contiene un registro detallado de los homicidios cometidos en el territorio ecuatoriano durante el año 2024. Esta base de datos incluye información clave que permite un análisis exhaustivo desde diversas perspectivas, integrando variables geográficas, demográficas, temporales y circunstanciales. Este conjunto de datos reporta inicialmente 6987 homicidios en todo el territorio nacional durante el año 2024 (Subsecretaría de Estudios y Estadística de la Seguridad, 2025).

Entre las variables más relevantes para el análisis se encontraron:

- **Geográficas:** Identificación de las zonas donde ocurrieron los eventos mediante datos de provincia, cantón, distrito, circuito, subcircuito y coordenadas geográficas específicas (latitud y longitud).
- **Demográficas:** Características de las víctimas como edad, género, estado civil, nivel de instrucción y nacionalidad.
- **Circunstanciales:** Detalles del evento, como tipo de arma utilizada, lugar del hecho, presunta motivación, y probable causa.
- **Temporales:** Fecha exacta del evento, lo que permite analizar patrones estacionales o relacionados con días específicos del año.

A pesar de que la base de datos utilizada en este estudio abarca registros de homicidios ocurridos en todo el territorio nacional del Ecuador, se optó por focalizar el análisis en un subconjunto de provincias específicas como Manabí y Guayas. Esta decisión se fundamenta en el hecho de que estas provincias, históricamente, han reportado las tasas más altas de homicidios y representan los principales puntos críticos de violencia en el país (El Diario, 2024). Estas zonas han sido identificadas previamente como áreas de preocupación prioritaria en informes de seguridad pública y, además, poseen características socioeconómicas, geográficas y demográficas que podrían influir significativamente en los patrones de criminalidad.

Como se puede observar en la Figura 1, que muestra el número total de homicidios registrados en cada provincia durante el año 2024, estas dos provincias destacan significativamente en comparación con el resto del país. Guayas presenta el mayor número de incidentes, seguida por Manabí, lo que refuerza su clasificación como zonas prioritarias para la investigación.

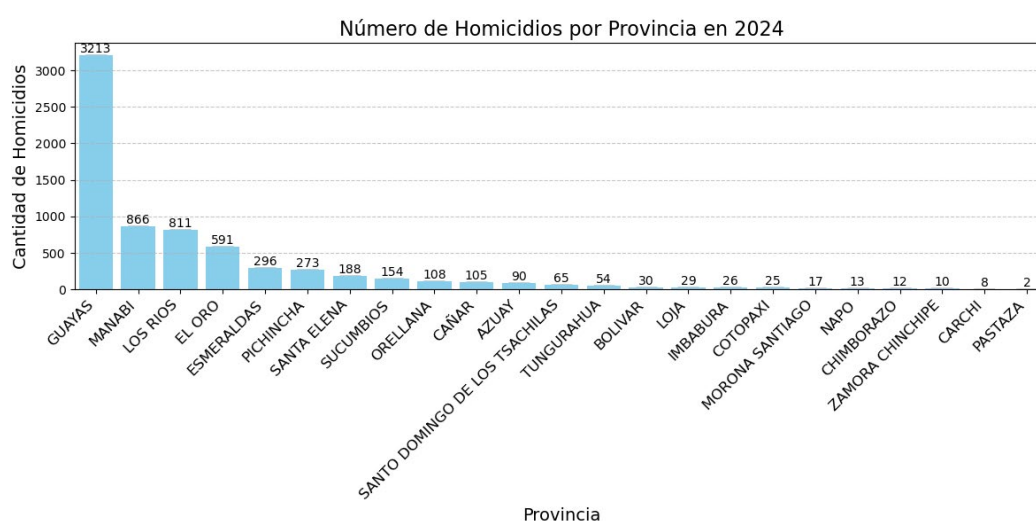


Figura 1. Número de homicidios por provincias 2024.

Nota. Fuente. Conjunto de datos abiertos del Ministerio del Interior.

La concentración del análisis en estas provincias permitió profundizar en las zonas de mayor incidencia, priorizando aquellas áreas donde los resultados pueden generar un impacto inmediato y significativo. Esto no significa que las demás provincias sean menos relevantes en el análisis criminalístico, sino que se busca enfocar los esfuerzos en zonas críticas para maximizar la utilidad de los resultados. La selección de estas provincias para el análisis de datos resultó en un total de 4079 homicidios registrados.

Implementación del Algoritmo DBSCAN

Para la aplicación del algoritmo DBSCAN, se utilizaron datos georreferenciados de homicidios en Ecuador, los cuales incluyeron coordenadas de latitud y longitud. Estas coordenadas se transformaron a radianes para facilitar el cálculo de distancias geodésicas utilizando la fórmula del “haversine”.

Inicialmente, se exploró la posibilidad de incluir variables adicionales, como características socioeconómicas y demográficas, en el proceso de *clustering*. Sin embargo, tras un análisis preliminar, se observó que estas

variables introducían ruido y dificultaban la identificación de patrones espaciales claros. Por lo tanto, se decidió utilizar únicamente las coordenadas geográficas como variables de entrada para dicho algoritmo.

Para el entrenamiento del algoritmo DBSCAN el parámetro eps (epsilon) el cual define el radio de búsqueda para agrupar puntos. Dado que el objetivo era identificar zonas calientes dentro de un umbral de 5 km, se calculó el valor de eps en función de la distancia geodésica. La fórmula del “*haversine*” se utilizó para calcular la distancia entre dos puntos en la superficie de la Tierra, considerando el radio terrestre de 6,371 km. Por lo tanto, el valor de eps seleccionado fue 0.000784 radianes, equivalente a una distancia aproximada de 5 km en la superficie terrestre.

Asimismo, como muestra la Tabla 1, el parámetro min_samples se estableció en 20, lo que significa que se requieren al menos 20 puntos dentro del radio de 5 km para formar un clúster. Los puntos que no cumplieron con este criterio se clasificaron como ruido (outliers).

Tabla 1. Hiperparámetros de DBSCAN.

Hiperparámetros	Valor	Justificación
epsilon	0.000784	Radio de búsqueda para agrupar puntos
min_samples	20	Se requieren al menos 20 puntos dentro del radio de 5 km para formar un clúster

Las etiquetas de clúster asignadas por DBSCAN se agregaron como una nueva variable al conjunto de datos original. Esta variable, denominada zonas calientes, representaba la pertenencia de cada observación a un grupo específico (o al ruido, en caso de los puntos no asignados a ningún clúster). Los clústeres identificados como zonas calientes se codificaron como valores positivos, mientras que el ruido se codificó como -1.

Implementación del Algoritmo Random Forest tras DBSCAN

Luego de la aplicación del algoritmo DBSCAN para la identificación de zonas de alta incidencia de homicidios, se procedió a construir un modelo de clasificación basado en Random Forest con la finalidad de predecir la ocurrencia de homicidios en función de variables contextuales y temporales. Se utilizó este algoritmo por su capacidad para manejar grandes volúmenes de datos, su resistencia al sobreajuste y su habilidad para capturar relaciones no lineales entre variables. En comparación con redes neuronales artificiales (ANN), que requieren grandes cantidades de datos y un proceso de entrenamiento más complejo, Random Forest permite interpretabilidad y resultados confiables con menor ajuste de parámetros. Su desempeño en problemas de clasificación con conjuntos de datos diversos lo hace una opción adecuada para este estudio. A continuación, se describen los pasos seguidos en la implementación del modelo.

Dado que ya se tenían las variables geográficas en radianes, se procedió a utilizar la técnica codificación por etiqueta para transformar variables categóricas en valores numéricos mapeándolos uno a uno. Esta técnica fue aplicada a características como el tipo de muerte, tipo de arma, lo que permitió que el modelo de Random Forest pudiera interpretar estos datos de manera eficiente.

Las dimensiones escogidas incluyeron coordenadas geográficas (coordenada x, coordenada y en radianes), características sociodemográficas (edad) y detalles contextuales del evento (tipo de muerte, arma utilizada y tipo de arma). Además, las etiquetas de clústeres generadas por DBSCAN se utilizaron como variable objetivo, permitiendo al modelo predecir la pertenencia a zonas calientes de homicidios. Dependiendo de la naturaleza de la variable se incluyó codificación por etiqueta para variables categóricas y normalización para variables numéricas.

Luego de haber realizado este proceso, el conjunto de datos se dividió en dos subconjuntos disjuntos: un conjunto de entrenamiento y un conjunto de prueba. Esta división fue fundamental en el proceso de modelado predictivo, ya que permitió entrenar el modelo con una porción de los datos y evaluar su desempeño con datos no vistos, simulando así su comportamiento en un escenario real.

El conjunto de datos se dividió en una proporción de **80%-20%**, siguiendo prácticas comunes en la literatura de aprendizaje automático. Fue realizada mediante la función `train_test_split` de la librería Scikit-learn (2025) que permite realizar una división estratificada del conjunto de datos a través del parámetro `stratify`. Este muestreo estratificado dividió el conjunto de datos de manera que cada subconjunto mantiene la misma distribución de clases que el conjunto original evitando un posible sesgo. Además, la proporción se seleccionó para equilibrar la necesidad de disponer de suficientes datos para entrenar el modelo (80%) mientras se reserva una cantidad significativa de datos para evaluar su generalización (20%).

La división se realizó de manera aleatoria para garantizar que ambos subconjuntos fueran representativos de la distribución general de los datos. Este enfoque minimiza el riesgo de sesgos en la evaluación del modelo, ya que asegura que tanto el conjunto de entrenamiento como el de prueba contengan una mezcla equilibrada de las diferentes características presentes en los datos originales.

Seguido a esto se entrenó el modelo con los hiperparámetros que se muestran en la Tabla 2:

Tabla 2. Hiperparámetros de Random Forest.

Hiperparámetros	Valor	Justificación
n_estimators	100	Número de árboles de decisión que se incluirán en el ensemble de Random Forest.
random_state	42	Semilla aleatoria para garantizar la reproducibilidad de los resultados.

El desarrollo de este estudio se realizó en un laboratorio Python, utilizando una computadora personal con un procesador Intel Core i7-13650HX, 16 GB de memoria RAM y un SSD NVMe de 1TB. El sistema operativo utilizado fue Windows 11, 64 bits. El código se implementó en Jupyter Notebook, ejecutado en el navegador web Brave, y se gestionó mediante Anaconda Distribution. Las librerías de Python v3.12.3 utilizadas incluyeron Pandas y NumPy para la manipulación de datos, Scikit-learn para la implementación de DBSCAN y Random Forest, y Folium para la visualización de mapas interactivos.

RESULTADOS

En primera instancia se realizó un análisis de patrones temporales asociados a los homicidios, donde se generaron nuevas dimensiones temporales derivadas de la variable original *fecha_homicidio*. El análisis de estas dimensiones temporales se presenta a continuación utilizando gráficos descriptivos que destacan la distribución de homicidios según el día de la semana y la tendencia mensual a lo largo del año.

En la Figura 2, se presenta el número de homicidios registrados por día de la semana, evidenciando que ciertos días como los días sábado, domingo y lunes muestran una mayor incidencia que otros, lo que puede estar relacionado a múltiples factores, uno de ellos puede ser la presencia de días festivos emitidos por el Gobierno Nacional y que en ciertas ocasiones son trasladados a los fines de semana. Lo cual se corroboró calendario de feriados nacionales 2024 emitidos por el Ministerio de Turismo del Ecuador (Ministerio del Turismo, 2022).

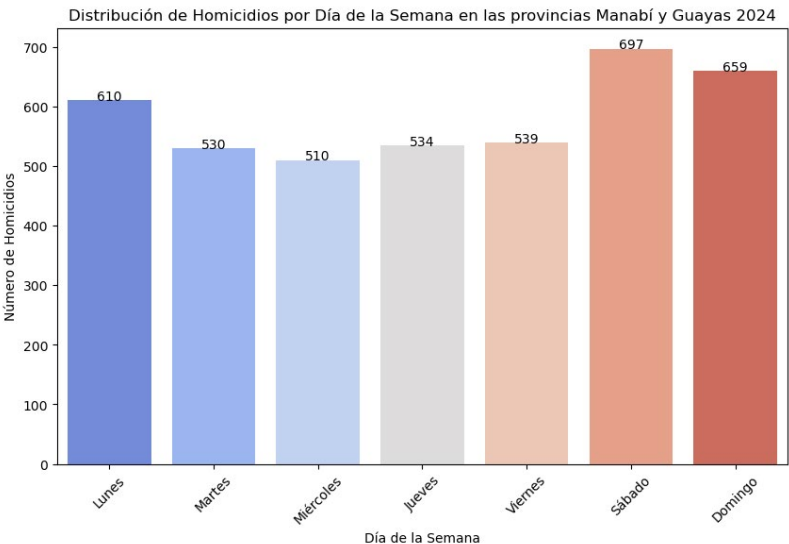


Figura 2. Distribución de homicidios por días de la semana.
Nota. Fuente. Conjunto de datos abiertos del Ministerio del Interior.

La Figura 3 muestra la tendencia mensual de los homicidios durante 2024, evidenciando variaciones con picos específicos a lo largo del año. En febrero se observó un descenso, mientras que diciembre el valor más alto, con 423 homicidios en las provincias analizadas. Estas variaciones están asociadas a factores coyunturales y a las medidas de seguridad implementadas por el Gobierno Nacional en determinados periodos. Fue el período donde se declaró el conflicto armado interno y se catalogaron como grupos terroristas a 22 organizaciones delictivas, dando paso a las fuerzas armadas a combatir dichos grupos terroristas (Ministerio de Defensa Nacional, 2024).

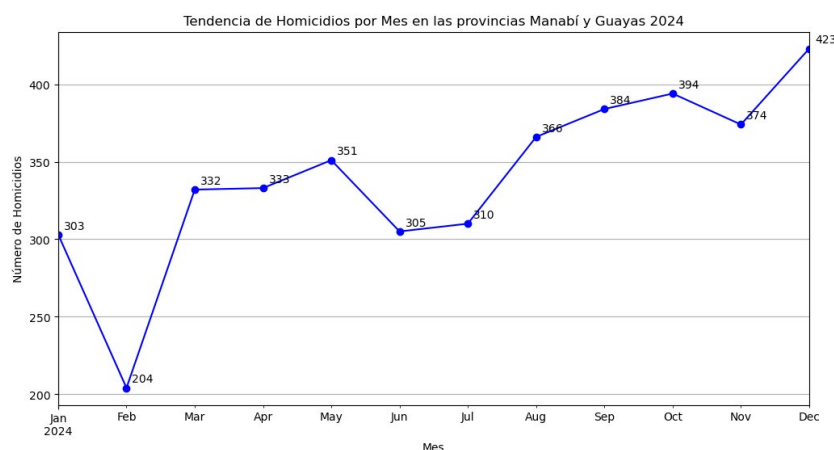


Figura 3. Tendencia de homicidios por mes.

Nota. Fuente. Conjunto de datos abiertos del Ministerio del Interior.

Resultados de la implementación de algoritmo DBSCAN

El algoritmo DBSCAN se aplicó a los datos georreferenciados de homicidios en Ecuador, utilizando un valor de $\text{eps} = 0.000784$ radianes (equivalente a un radio de 5 km) y $\text{min_samples} = 20$, donde el algoritmo identificó un total de 8 clústeres significativos. Estos clústeres representan las zonas con mayor incidencia de homicidios y permiten visualizar los puntos críticos en el territorio analizado.

En la tabla 3, los valores de la primera columna, denominada “Clúster”, representan las etiquetas de agrupación generadas por el modelo. En este caso, el modelo generó las siguientes etiquetas: $C = [1, 0, 3, 2, 4, 7]$. La segunda columna, denominada “Número de homicidios”, indica la densidad de homicidios dentro de cada clúster. Cabe destacar que los clústeres 0 y 3 tienen la misma densidad. Sin embargo, aunque ambos clústeres presenten la misma cantidad de homicidios, esto no implica que correspondan a la misma ubicación ni que compartan el mismo patrón.

Tabla 3. Clústeres generados por DBSCAN.

Clúster	Número de homicidios
1	2842
0	324
3	324
2	157
4	89
7	70

Para una mejor interpretación de los resultados, se generaron mapas interactivos utilizando la biblioteca Folium, donde cada clúster se representó con un color distintivo. Esto permitió identificar visualmente la distribución geográfica de los homicidios y sus concentraciones más críticas.

En el mapa de la Figura 4 se presenta la distribución de los clústeres identificados en la provincia de Guayas. El análisis revela la presencia de un clúster de gran tamaño, representado por el color rojo, que concentra una proporción significativa de homicidios en esta región. Dicho clúster abarca múltiples cantones, destacándose Guayaquil, Durán, Naranjal y otros sectores aledaños, lo que sugiere una alta densidad de eventos delictivos en estas zonas.

Además del clúster principal, se identificaron otros agrupamientos de menor tamaño dentro de la provincia de Guayas, representados en el mapa con los colores verde y naranja. Estos clústeres incluyen localidades como General Villamil (Playas), Simón Bolívar, Naranjito y El Triunfo, entre otros. La presencia de estos grupos indica que la concentración de homicidios no se limita únicamente al área metropolitana de Guayaquil, sino que también se extiende a cantones periféricos con una menor densidad poblacional.

Por otro lado, el análisis también ha permitido detectar la presencia de datos considerados como ruido, los cuales han sido clasificados en el clúster -1 y están representados en el mapa con el color gris. Estos puntos corresponden a homicidios que no cumplen con los criterios de densidad para formar un clúster significativo según los parámetros establecidos en el modelo DBSCAN. La distribución dispersa de estos eventos sugiere que

pueden tratarse de casos aislados o atípicos en comparación con las zonas identificadas como de alta incidencia delictiva.

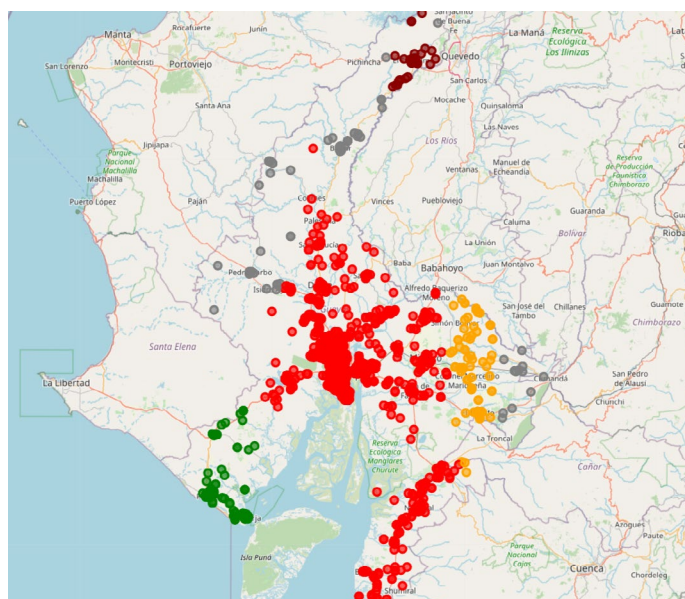


Figura 4. Clústeres en la provincia Guayas.

De manera similar, en la Figura 5 se presentan los clústeres identificados en la provincia de Manabí. El análisis espacial delimitó agrupamientos significativos en diversas ciudades, entre las cuales destacan Manta, Portoviejo, Bahía de Caráquez, Pedernales, Junín y Tosagua. Estos clústeres indican zonas con una mayor concentración de homicidios, lo que sugiere patrones geospaciales relevantes en la dinámica criminal de la provincia. Cabe destacar que, al igual que en la provincia de Guayas, en Manabí también se evidencia la presencia de datos clasificados como ruido, los cuales corresponden a homicidios que no lograron agruparse dentro de un clúster específico debido a su dispersión geográfica.

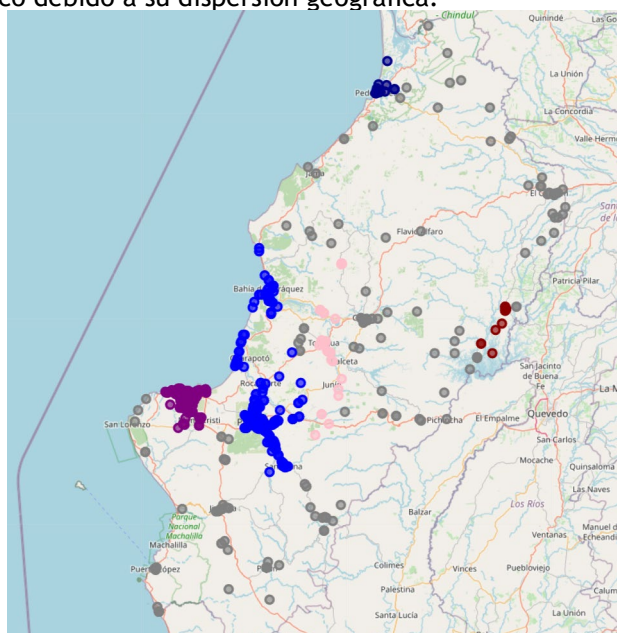


Figura 5. Clústeres en la provincia Manabí.

Resultados de la implementación de algoritmo Random Forest tras DBSCAN

Una vez identificados los clústeres de homicidios a través del algoritmo DBSCAN, se procedió a la implementación de un modelo Random Forest con la finalidad de predecir la pertenencia de nuevos casos a los clústeres generados. Esta etapa permitió evaluar la importancia de las variables en la clasificación de zonas con alta incidencia delictiva, así como estimar la probabilidad de que futuros eventos criminales ocurran en determinadas áreas geográficas.

El modelo Random Forest fue seleccionado debido a su capacidad para manejar datos no lineales y su robustez frente a la variabilidad de los datos. A partir de las características extraídas de la base de datos, se entrenó el modelo utilizando variables geoespaciales y demográficas, lo que permitió establecer patrones asociados a la ocurrencia de homicidios en las diferentes provincias analizadas.

El modelo mostró un desempeño excepcional en la clasificación de zonas calientes de homicidios, con una exactitud de 0.9987, una precisión de 0.9988, una sensibilidad de 0.9987 y un F1-Score de 0.9987. Estos resultados indican que el modelo es altamente confiable y efectivo para identificar zonas calientes, minimizando los errores en las predicciones positivas y manteniendo un equilibrio entre la precisión y la sensibilidad. Aunque el desempeño es notable, es importante considerar posibles limitaciones, como el riesgo de sobreajuste y la necesidad de validar el modelo con datos externos.

Además de las métricas antes mencionadas también se obtuvo una matriz de confusión. Los valores de la matriz de confusión fueron obtenidos evaluando el modelo de Random Forest con los datos del conjunto de prueba. Las predicciones del modelo fueron comparadas con las etiquetas reales para calcular la matriz de confusión.

La matriz de confusión de la Figura 6, fue obtenida para el modelo predictivo basado en Random Forest y reveló un desempeño excepcional, con 761 objetos predichos correctamente y 1 objeto predicho de manera incorrecta. El bajo número de falsos positivos y falsos negativos sugiere que el modelo puede ser utilizado con confianza para identificar zonas calientes de homicidios y asignar recursos de manera eficiente.

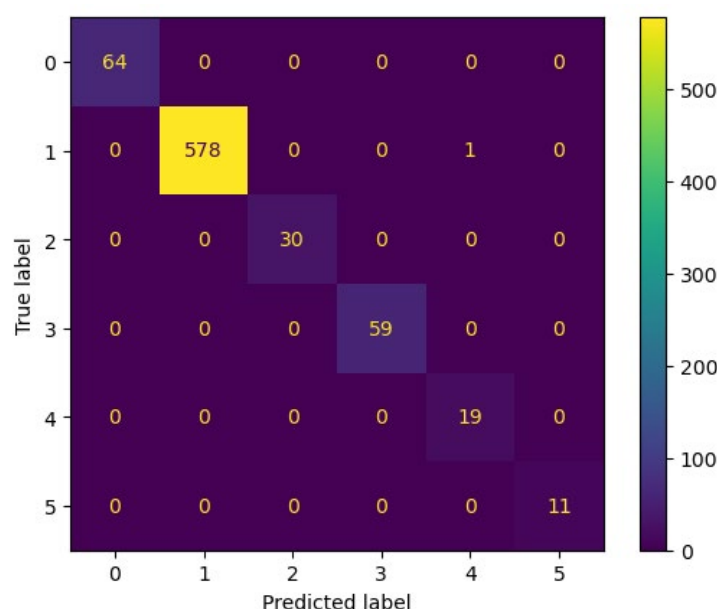


Figura 6. Resultado de la matriz de confusión.

DISCUSIÓN

Los resultados obtenidos a partir del análisis con DBSCAN y la modelización predictiva con Random Forest permiten una visión más integral de los patrones geoespaciales y demográficos de los homicidios en Ecuador. En particular, se identificaron clústeres de alta incidencia en las provincias de Guayas y Manabí, caracterizadas por elevados índices delictivos. Estos hallazgos coinciden con lo señalado por Barrionuevo-Santamaría et al. (2025), quienes sostienen que las zonas urbanas y periurbanas presentan mayores niveles de violencia en comparación con las áreas rurales, debido a la alta densidad poblacional y a la concentración de actividades delictivas en estos espacios.

El modelo de Random Forest, por su parte, mostró que la ubicación geográfica y las variables temporales son predictores clave en la ocurrencia de homicidios. La capacidad del modelo para clasificar correctamente los eventos sugiere que los patrones detectados no son meramente aleatorios, sino que responden a factores estructurales del entorno.

Los resultados obtenidos en este estudio son consistentes con investigaciones internacionales que han aplicado el algoritmo DBSCAN y modelos de aprendizaje automático para el análisis delictivo. Por ejemplo, Ingilevich e Ivanov (2018) emplearon DBSCAN para identificar patrones espaciales de criminalidad y, complementariamente, utilizaron métodos de predicción como la regresión lineal y la regresión logística. A pesar de las diferencias

metodológicas, sus resultados también evidenciaron una alta efectividad en la identificación de zonas con elevada incidencia de homicidios, lo que refuerza la validez del enfoque adoptado en el presente estudio.

De manera similar, Fontalvo Herrera et al. (2023) proponen un modelo implementado en Colombia durante el período 2018-2022, que combina técnicas de *clustering* para identificar zonas con mayor impacto delictivo y redes neuronales para predecir la ocurrencia de delitos. Su enfoque obtuvo un nivel de precisión del 97,7 % en tareas de predicción y clasificación, lo que demuestra la efectividad del uso combinado de modelos no supervisados y supervisados en el análisis espacial de la criminalidad.

Otras investigaciones han explorado la relación entre condiciones socioeconómicas y la mortalidad por homicidio juvenil masculino. En el estudio de Mejía Escamilla et al. (2020), realizado en municipios del Estado de México y delegaciones de la Ciudad de México con más de 100.000 habitantes, se evidenció que los intentos de correlacionar directamente los indicadores socioeconómicos con las tasas de homicidio resultaron inconsistentes. Ante ello, se optó por un análisis de *clústeres*, el cual reveló que las tasas más elevadas de homicidio juvenil se concentraban en áreas con peores indicadores socioeconómicos. Sin embargo, también se identificó un *clúster* con condiciones relativamente mejores, pero caracterizado por alta desigualdad económica y bajos niveles de escolaridad entre los jóvenes, donde se observó el mayor crecimiento en homicidios.

Finalmente, los resultados de este estudio pueden contribuir al diseño de estrategias basadas en evidencia para la prevención del crimen. La identificación de zonas de alta incidencia permite la focalización de recursos y patrullajes, mientras que el análisis temporal puede guiar la asignación de personal de seguridad en períodos críticos. Asimismo, la integración de modelos predictivos en centros de monitoreo y sistemas de seguridad ciudadana podría mejorar la capacidad de respuesta de las autoridades ante posibles eventos delictivos. Esto resalta la importancia de fortalecer las infraestructuras tecnológicas en la lucha contra el crimen.

No obstante, este estudio presenta varias limitaciones que deben ser consideradas. En primer lugar, el análisis se focalizó únicamente en dos provincias: Guayas y Manabí; si bien permitió una mayor profundidad en contextos de alta incidencia delictiva, limita la generalización de los hallazgos al resto del país. En segundo lugar, la exclusión de variables socioeconómicas y contextuales en el modelo de *clustering*, por haber generado ruido en las pruebas preliminares, podría haber limitado la detección de patrones más complejos o multifactoriales. Finalmente, aunque el modelo Random Forest mostró un desempeño excepcional, no se realizó una validación cruzada con datos externos, lo que deja abierta la posibilidad de sobreajuste y limita la robustez del modelo en escenarios diferentes.

CONCLUSIONES

El presente estudio logró identificar, mediante el algoritmo DBSCAN, agrupaciones significativas de homicidios en zonas específicas de las provincias de Guayas y Manabí, destacándose ciudades como Guayaquil, Durán, Naranjal, Manta, Portoviejo y Bahía de Caráquez. Estos resultados refuerzan evidencias previas sobre la concentración de violencia en sectores urbanos y periurbanos con alta densidad poblacional. Asimismo, la detección de un número considerable de puntos clasificados como ruido indica la existencia de homicidios dispersos, no asociados a patrones espaciales definidos, lo que podría deberse a eventos aislados o a limitaciones en la granularidad de los datos disponibles.

Por su parte, el modelo Random Forest demostró una alta eficacia en la clasificación de homicidios dentro de los *clústeres* identificados, resaltando la importancia de variables geoespaciales y demográficas en la predicción de zonas de alta incidencia. Esta perspectiva combinada de análisis no supervisado y supervisado no solo permite caracterizar territorialmente el fenómeno delictivo, sino que también ofrece un marco técnico aplicable a sistemas de seguridad ciudadana y planificación estratégica. No obstante, futuras investigaciones deberían incorporar validaciones externas y variables contextuales adicionales, así como ampliar el análisis a otras regiones del país para fortalecer la generalización de los resultados.

REFERENCIAS

- Barriónuevo-Santamaria, L. E., Basantes-Rivera, N. R., Barahona-Bolaños, J. C., & Quespás-Valencia, V. R. (2025). Análisis de tendencias y factores determinantes de la violencia en el Ecuador, año 2020-2024. *Dominio de las Ciencias*, 11(1), 106-123. <https://doi.org/10.23857/dc.v11i1.4199>
- Carrión Mena, F. (2023). El territorio unificado del crimen en el Ecuador. *Ecuador Debate*, 119, 75-105. <http://hdl.handle.net/10469/19923>
- El Diario. (2024, diciembre 18). *El 83 % de la violencia de Ecuador se concentra en Manabí, Guayas, Los Ríos y El Oro*. <https://diariocorreo.com.ec/107500/nacional/el-83--de-la-violencia-de-ecuador-se-concentra-en-manabi-guayas-los-rios-y-el-oro#:~:text=Seg%C3%BAn%20los%20datos%20presentados%2C%20el,han%20focalizado%20los%20mayores%20esfuerzos>
- Estévez-Calapaqui, N. (2021). Relación entre desempleo y delincuencia. *Saberes*, 1, 38-47. <https://orcid.org/0000-0002-2311-2158>

- Fernando, R., Moncayo, V., & Pazos Espinoza, F. E. (2024). Análisis de índices de violencia criminal en Ecuador, 2023-2024. *Revista de Investigación en Seguridad Ciudadana y Orden Público*, 6-11. <https://innovacionsaber.isupol.edu.ec/index.php/innovacion/article/view/274>
- Fontalvo Herrera, T. J., Vega Hernández, M. A., & Mejía Zambrano, F. (2023). Clustering method and artificial intelligence to classify and project violent crime in Colombia. *Revista Científica General José María Córdova*, 21(42), 551-572. <https://doi.org/10.21830/19006586.1117>
- González, M. A. (2024, diciembre 30). *Las cifras de la violencia: En 2024, Ecuador atravesó una mejoría que no es suficiente*. <https://www.primicias.ec/seguridad/cifras-violencia-inseguridad-robos-ecuador-2024-86371/>
- Ingilevich, V., & Ivanov, S. (2018). Crime rate prediction in the urban environment using social factors. *Procedia Computer Science*, 136, 472-478. <https://doi.org/10.1016/j.procs.2018.08.261>
- Leonel, A., Molina, S., Valeria, N., Ochoa, V., Fernando, C., Salazar, B., & Santillán Ojeda, S. J. (2022). Drogas, tráfico y crimen organizado como detonante de actos violentos en las cárceles del Ecuador. *Universidad y Sociedad*, 478-486. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2218-36202022000300478&lng=es&nrm=iso
- Mejía Escamilla, Ma. G., Spijker, J., & Manetta, A. (2020). Distribución espacial del homicidio juvenil masculino y desigualdad territorial en Ciudad de México y Estado de México (2000-2002 y 2010-2012). *Revista Latinoamericana de Población*, 14(27), 185-227. <https://doi.org/10.31406/relap2020.v14.i12.n27.6>
- Ministerio de Defensa Nacional. (2024, enero 12). *Síntesis noticiosa*. <https://www.defensa.gob.ec/wp-content/uploads/downloads/2024/01/12-SINTESIS-NOTICIOSA-VIERNES-12-DE-ENERO-DE-2024.pdf>
- Ministerio del Turismo. (2022). *Calendarios de Feriados Nacionales Ecuador 2023-2025*. <https://www.turismo.gob.ec/wp-content/uploads/2023/12/CALENDARIO-FERIADOS-2023-2025-06-12-2022-.pdf>
- Ojeda Segovia, L. (2023). Violencia y delincuencia en el Ecuador: principales problemas, mitos y desafíos. *Ecuador Debate*, 119, 107-130. <http://hdl.handle.net/10469/19924>
- Scikit-learn. (2025, enero). *train_test_split – scikit-learn 1.6.1 documentation*. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html
- Subsecretaría de Estudios y Estadística de la Seguridad. (2025). *Homicidios Intencionales*. <https://www.datosabiertos.gob.ec/dataset/homicidios-intencionales>

Conflicts of Interest:

The author declares no conflict of interest.

Author Contributions:

The author was responsible for all aspects of the study, including conceptualization, methodology, analysis, and writing.

Disclaimer/Publisher's Note:

The statements, opinions, and data contained in all publications are solely those of the authors and individual contributors and not of Revista San Gregorio or the editors. Revista San Gregorio and/or the editors disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.