

Escuela Superior Politécnica del Litoral

Facultad de Ciencias Naturales y Matemáticas

Desarrollo de un Modelo Bayesiano para la Predicción de Incidencias Delictivas en la Zona 8 del Ecuador Basado en la Teoría de los Lugares de Riesgo

Proyecto Integrador

Previo a la obtención del Título de:
Ingeniero en Estadística

Presentado por:

Geovanny Gutiérrez

Fabrizio Quinto

Guayaquil - Ecuador

Año: 2024

Agradecimientos

Me gustaría expresar mi agradecimiento a todas esas personas que han sido fundamentales en la realización de este proyecto.

A mi familia, que me ha apoyado desde un inicio y han depositado su fe incondicional en mí. A mi madre querida, Rita, que ha luchado a diario sin descanso para que nunca me falte nada y me ha enseñado que el amor verdadero existe, esa persona que cada día de mi vida me ha hecho sentir amado y protegido. A mi padre Luis que con sangre, sudor y lágrimas me ha podido llevar a delante a pesar de los contratiempos y me ha mostrado su apoyo incondicional incluso en los desacuerdos. A mi hermano Gustavo, mi ñaño, que ha representado una fuente de inspiración de gran valor en mi vida, un ser deslumbrante de determinación, esfuerzo y dedicación que me ha enseñado que el camino difícil nunca debe ser esquivado sino afrontado con valentía.

A mis amigos que han estado a mi lado en las buenas, en las malas y en las peores. A Tamara que ha sido un pilar emocional y fuente de consejos que me han enseñado a madurar como persona y a tomar decisiones encaminadas al éxito y bienestar personal para mi desarrollo. A Anita que se ha vuelto en esa consejera y compañera de anécdotas que me ha acompañado y escuchado en mis momentos más vulnerables y que me ha sacado más de una sonrisa. A Cristhian que desde la secundaria se ha vuelto como mi hermano, un personaje de corazón noble que ha sido también una especie de guía espiritual para mí y ha permitido poder acercarme a Dios y que además nos compartió sus recursos tecnológicos para la compleción del componente computacional de este proyecto. A Fabricio que ha demostrado, a más de un excelente amigo, un compañero de proyecto excepcional, un apoyo fundamental en la construcción de este trabajo, una pieza indispensable en el proceso investigativo y de ejecución. A todos mis colegas de carrera: Miguel, Laura, Gerleen, Abraham, Pablo, Nicolás, Camilita, Odalys, Hellen, Ana Gabriela, quienes a diario me han hecho sentir parte de una familia de hermanitos y me han colmado de sus "¡tu puedes!" hasta que se volvió realidad este sueño que parecía tan lejano.

A mis maravillosos profesores que han sabido compartir conmigo sus conocimientos, orientación y consejos. A mi tutora por excelencia, Andrea, que nos supo guiar desde las etapas más tempranas del proyecto y que además, nos ha sembrado el interés por la

Estadística desde sus clases en las que refleja su transparente y notable amor a la ciencia. A Katherine que ha tomado el rol improvisado de una especie de segunda tutora, aportando siempre con sus amplios conocimientos y con su gran apoyo a quien me atrevo a decir que se ha convertido en una gran amiga e inspiración. A Jessica que ha demostrado ser una coordinadora de carrera extraordinaria y que, así mismo, ha logrado que nos acerquemos más a la estadística y la divulgación científica. A Christian que nunca dudó de mis capacidades y depositó su fe en mi potencial. A Puri, que desde las prácticas me enseñó el abanico de posibilidades del mundo de la investigación y me ha demostrado la relevancia de la ciencia en la resolución de problemas.

Finalmente, pero no menos importante, al Padre Celestial que ha permitido que todo esta sea posible, que ha obrado en mí desde el momento en que nací para que yo pueda llegar a este punto de mi vida en el que pueda usar mi ingenio para aportar mi granito de arena en problemáticas de impacto social.

Geovanny Gutiérrez

Agradecimientos

Quiero expresar mi agradecimiento a todas las personas que han estado a mi lado durante este camino.

A mi familia, que ha sido mi pilar fundamental en todo momento. A mi madre, Lucy, por su amor incondicional y por ser una fuente de inspiración; siempre he tenido presente toda la entrega que me ha permitido cumplir con mis objetivos. Gracias por apoyarme en cada paso. A mi padre, Vidal, por su constante respaldo desde el inicio y por la confianza que siempre ha depositado en mí. A mi abuela América, por cada uno de sus consejos y su comprensión. A mis tíos Alfredo y Mariuxi, por su apoyo incondicional a lo largo de este camino; y a mi hermano ,Erick, espero que sea una fuente de motivación para seguir explorando y descubriendo nuevas cosas.

A mis amigos, Miguel, Hellen, Jair, Odalys, Gerleen, Gabriela y Abigail, por su ayuda en momentos complicados y por brindarme apoyo tanto académico como emocional. Además, a Laura, cuya amistad ha significado mucho para mí; siempre estaré agradecido de contar con ella. A Geovanny, por su colaboración durante toda esta etapa. Gracias por tu amistad y dedicación.

A Ángel, mi compañero de muchas aventuras y mi apoyo constante. Gracias por tus palabras de aliento cuando más lo necesitaba, y por recordarme la importancia de soñar en grande.

A mis profesores, quienes no solo me han transmitido conocimientos, sino que también me han inspirado a superarme cada día. En especial, quiero agradecer a nuestra tutora, Andrea García, por su paciencia, dedicación y el tiempo que ha invertido en este proceso. Su guía ha sido fundamental para la realización de este proyecto.

Fabricio Quinto

Declaración Expresa

“Los derechos de titularidad y explotación, nos corresponde conforme al reglamento de propiedad intelectual de la institución; Gutiérrez Galarza Geovanny André y Quinto Sánchez Vidal Quinto damos nuestro consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual.”



Gutiérrez Galarza Geovanny André



Quinto Sánchez Fabricio Vidal

Evaluadores

Ph.D. García Bustos Sandra
Lorena
Profesor de Materia

Ph.D. García Angulo Andrea
Cristina
Tutor de proyecto

Resumen

En este proyecto se construye un modelo estadístico con un enfoque bayesiano utilizando la metodología INLA (Integrated Nested Laplace Approximation) para estimar parámetros que capturen la relación espacial y temporal de incidentes de delitos como tráfico de drogas, robos, homicidios y extorsiones. El modelo considera un índice de riesgo calculado mediante la técnica de RTM (Risk Terrain Modeling), que incorpora la influencia de factores ambientales como establecimientos generadores, atractores o detractores de delitos. Los resultados muestran cómo diferentes variables explicativas, como la proximidad a estaciones de policía, bares y centros comerciales, impactan en las tasas de delitos, demostrando efectos multiplicativos significativos. Para el análisis predictivo, se identifican las distribuciones estadísticas más adecuadas para cada tipo de delito, destacando la distribución de Poisson con ceros inflados para homicidios y la Binomial Negativa para tráfico de drogas, lo que permite visualizar mapas de riesgo y realizar predicciones específicas por zonas.

Abstract

In this project, a statistical model is constructed with a Bayesian approach using the INLA (Integrated Nested Laplace Approximation) methodology to estimate parameters that capture the spatial and temporal relationships of crime incidents such as drug trafficking, robberies, homicides, and extortion. The model considers a risk index calculated using the RTM (Risk Terrain Modeling) technique, which incorporates the influence of environmental factors such as establishments that generate, attract, or deter crime. The results show how different explanatory variables, such as proximity to police stations, bars, and shopping centers, impact crime rates, demonstrating significant multiplicative effects. For the predictive analysis, the most suitable statistical distributions for each type of crime are identified, highlighting the Poisson distribution with inflated zeros for homicides and the Negative Binomial for drug trafficking, which allows for visualizing risk maps and making specific predictions by area.

Índice

1. Introducción	10
1.1. Descripción del Problema	10
1.2. Justificación del Problema	11
1.3. Objetivos	11
1.3.1. Objetivo General	11
1.3.2. Objetivos Específicos	11
2. Marco Teórico	12
2.1. Revisión de Literatura	12
2.2. Risk Terrain Model	13
2.2.1. Modelo de Regresión de Poisson Penalizada	13
2.2.2. Regression Lasso	14
2.2.3. Elastic-net	15
2.3. Modelos generalizados mixtos	16
2.4. Modelos espacio-temporales	16
2.4.1. Integrated Nested Laplace Approximation	17
2.5. Criterios de Información	17
2.5.1. Bayesian Information Criterion (BIC)	17
2.5.2. Deviance Information Criterion (DIC)	18
2.5.3. Watanabe Akaike Information Criterion (WAIC)	19
3. Metodología	20
3.1. Flujo de Trabajo	20
3.2. Delimitación de la Zona de Estudio	20
3.3. Cruadrículas	22
3.4. Recolección de Datos	22
3.4.1. Indidentes	22
3.4.2. Factores de Riesgo	23
3.4.3. Variables	24
3.5. Entrenamiento del Modelo de Riesgo	24
4. Resultados y análisis	25
4.1. Estadística Descriptiva	25
4.2. Modelo de Riesgo	26
4.2.1. Valores de BIC para cada modelo	26
4.2.2. Drogas	27

4.2.3.	Robos	28
4.2.4.	Homicidios	29
4.2.5.	Extorsión	30
4.3.	Modelo Predictivo	31
4.3.1.	Drogas	32
4.3.2.	Robos	32
4.3.3.	Homicidios	33
4.3.4.	Extorsiones	33
5.	Consideraciones finales	34
5.1.	Conclusiones	34
5.2.	Recomendaciones	34

1 Introducción

Durante los últimos 3 años se ha experimentado una creciente ola de delitos cometidos en el Ecuador causando que en 2022 sea catalogado como el décimo país más violento de Latinoamérica y el Caribe (Zevallos et al., 2024). Según el Observatorio Ecuatoriano de Crimen Organizado (OECO, 2023) en su boletín anual de homicidios intencionales en Ecuador para el año 2023, solamente en la Zona 8 (Guayaquil, Durán y Samborondón) suceden el 35.06 % de los homicidios a nivel nacional, siendo esta una distorsión estadística a nivel nacional, por lo cual representa un caso aparte de estudio. A estos se suman otros delitos como extorsión, robo, secuestros, etc. Debido a esto, a inicios del año 2024 se declaró al país en un estado de “conflicto armado interno” (Zevallos et al., 2024). Entre las causas que podrían explicar este aumento significativo de delitos está el tráfico de drogas, contrabando, entre otras. Además, pueden existir factores que influyen en la incidencia de delitos, algunos de estos pueden ser sociales, económicos e incluso ambientales. Estos factores pueden estar relacionados con delitos de forma distinta por periodos de tiempo o por ubicación geográfica. Algunas teorías han intentado explicar estas asociaciones, por ejemplo, la teoría de la desorganización social explica las variaciones en la delincuencia como producto de la desintegración institucional (familia, escuela, iglesia, etc.). Estas instituciones fomentan relaciones organizadas y cooperativas, vinculando a los individuos a su comunidad y reduciendo la probabilidad de comportamientos delictivos. La teoría destaca que los cambios sociales rápidos, como la urbanización e industrialización, debilitan estos lazos y aumentan la criminalidad (Porter et al., 2015). Así también, la teoría de los lugares riesgosos sugiere que existen ciertos tipos de lugares que dada sus características tienen un mayor riesgo en comparación con otros lugares, además considera que el crimen surge en lugares con alta vulnerabilidad debido a múltiples influencias criminales combinadas (Caplan and Kennedy, 2016). En este trabajo se propone implementar un modelado de terreno de riesgos para calcular los niveles de riesgo en la Zona 8. Luego, se construirá un modelo mixto generalizado utilizando datos desde entre el período 2021-2023 para analizar patrones espacio-temporales, utilizando estos niveles de riesgo como covariables para predecir el índice de delincuencia en los próximos meses. Esta investigación busca modelar los delitos violentos más relevantes de la Zona 8 tomando en cuenta factores ambientales que complementen las asociaciones espacio-temporales.

1.1 Descripción del Problema

La ocurrencia de diversos tipos de delitos en la Zona 8 del Ecuador durante los últimos tres años ha ido en aumento, situación que ha generado un ambiente de inseguridad para

los ciudadanos ecuatorianos y ha planteado desafíos para las autoridades encargada de mantener el orden, como lo es la Policía Nacional y las Fuerzas Armadas. La Zona 8 incluye ciudades de alta densidad poblacional como Guayaquil, Durán y Samborondón siendo también un núcleo económico y social del país. Sin embargo, el incremento en la tasa de criminalidad amenaza el bienestar de sus ciudadanos y afecta negativamente la percepción de seguridad, la cohesión social y el desarrollo económico. Diversos factores como la falta de empleo, la desigualdad socioeconómica, el acceso limitado a servicios educativos y de salud, y la presencia de redes de crimen organizado contribuyen a una tendencia creciente de delitos en el país. La expansión de actividades ilícitas como el narcotráfico, robos, extorsiones y la violencia urbana ha generado un ambiente temerario en la población ecuatoriana, impactando tanto la calidad de vida de los residentes como la inversión y el turismo en la región.

1.2 Justificación del Problema

Las autoridades se enfrentan a múltiples desafíos para abordar la problemática de la delincuencia. Las fuerzas del orden a menudo manejan recursos limitados para manejar la complejidad y el volumen de los delitos. Además, la falta de organización en ciertas agencias de seguridad y justicia complican la implementación de políticas efectivas. La percepción de impunidad entre los ciudadanos también influye en la confianza que se tiene en las instituciones y dificulta la cooperación comunitaria, por lo que proveer a las fuerzas del orden de estudios científicos sobre el problema y modelos de predicción de los crímenes aportará a que se pueda focalizar el uso de recursos limitados a sectores de alto riesgo en tiempos apropiados.

1.3 Objetivos

1.3.1. Objetivo General

Construir un modelo estadístico mediante metodología bayesiana para predecir incidencias delictivas dada la ubicación geográfica y temporal en la Zona 8 del Ecuador.

1.3.2. Objetivos Específicos

1. Construir mapas y visualizaciones que permitan entender la naturaleza de los datos y sinteticen la información del conjunto de datos.

2. Implementar un modelo estadístico para la identificación de factores de riesgo en base a la teoría de los lugares de Riesgo.
3. Construir un modelo para crímenes de la Zona 8 y validar su poder predictivo para garantizar su funcionalidad.

2 Marco Teórico

2.1 Revisión de Literatura

Entre los modelos inferenciales para datos espacio-temporales que se han propuesto en la literatura se incluyen procesos de Cox similares a los procesos binomial negativos (Vaillant, 1991), regresión logística con variables de proximidad como la distancia a la parada de tren, carretera o terminal más cercanos, además de variables ambientales como el número de bares, restaurantes, supermercados, entre otros edificios comerciales (Rummen et al., 2017). Modelos de ceros inflados (zero-inflated) acotados superiormente se proponen también para análisis de datos de crímenes, sugiriendo que hay escenarios en los que la distribución de crímenes presenta altos niveles de ausencia en ciertos contextos y que además tienen una cota máxima como lo es la población de la región estudiada (Britt et al., 2017).

En lo que corresponde a Machine Learning se han utilizado técnicas como árboles de decisión, Random Forest (RF), AdaBoost, Support Vector Machine (SVM), K-Nearest Neighbours (KNN) para identificar hotspots de crímenes considerando factores temporales, geográficos y resultados históricos de patrones delictivos, como se realizó en la Universidad Tecnológica Nacional de Buenos Aires (González et al., 2021), en la Universidad Pedagógica y Tecnológica de Colombia para detección de tendencias en homicidios en Colombia (Ordoñez-Eraso et al., 2020), y en la Universidad de California-Irvine del estado de Mississipi implementando Regresión y Árboles de Decisión para patrones de crímenes violentos (McClendon and Meghanathan, 2015). En (Hossain et al., 2020) se consiguieron los mejores resultados con Random Forest y AdaBoost, superando a los árboles de decisiones e incluso a la regresión por KNN para predicción de crímenes. Una aplicación para clasificación en minería de datos es explorada en (Kianmehr and Alhajj, 2006) mediante SVM para identificar hotspots.

En lo que corresponde a modelos temporales, se estudió en China un modelo ARIMA para predecir tendencias de crímenes (Chen et al., 2008), clusterización multivariante de series de tiempo relacionadas a crímenes para encontrar tendencias similares (Chandra

et al., 2008), mientras que en la Universidad de Alabama se exploraron series de tiempo difusas (fuzzy time series) para predicción de crímenes como parte de un conjunto de técnicas que llaman Soft Computing en los que incluyen también Redes Neuronales y Computación Evolucionaria (Shrivastav and Ekata, 2012).

En (Muneer et al., 2021) se presenta un análisis multivariante mediante el método de clusterización Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) para identificar las zonas calientes de crímenes y posteriormente aplicaron un modelo predictivo estacional autorregresivo integrado de medias móviles (SARIMA) en las regiones más densas de crimen para el componente temporal del conjunto de datos. En (Luo et al., 2022) se realizó una caracterización automática de zonas geográficas con la ayuda de imágenes de Street View que permiten encontrar patrones ambientales que complementen la relación espacial de crímenes urbanos. Finalmente, un proyecto similar se realizó en Guayaquil estudiando también patrones espacio-temporales con un enfoque bayesiano para predicción de homicidios, tomando en cuenta factores de escolaridad, empleo y tasa de jóvenes en la población (Bravo, 2024).

2.2 Risk Terrain Model

Los lugares de riesgo son el resultado de la combinación de vulnerabilidad y exposición. Estos lugares están influidos por (1) la vulnerabilidad, o características del entorno que atraen el crimen, (2) la exposición local, que se refiere a crímenes repetidos en un corto período de tiempo, y (3) la exposición global, que indica áreas con alta concentración de delitos Caplan and Kennedy (2016). La Teoría de los Lugares de Riesgo sugiere que el riesgo de crimen se puede calcular analizando estos factores juntos. La vulnerabilidad se puede evaluar con el Modelo de Terreno de Riesgo (RTM), la exposición global con el mapeo de puntos calientes, y la exposición local con el análisis de repeticiones cercanas. Esta teoría integra conceptos de criminología ambiental para comprender la dinámica espacial del crimen y proporciona métodos para identificar y comparar lugares de riesgo.

2.2.1. Modelo de Regresión de Poisson Penalizada

El modelo de regresión de Poisson se utiliza para modelar datos de conteos, los cuáles son números enteros, que ocurren en un intervalo dado (Shengping and Berdine, 2015).

Cuando tenemos datos con alta dimensionalidad, a menudo se utiliza un procedimiento de penalización, el cuál propone agregar un término de penalidad a la función de log-verosimilitud $P_\lambda(\beta)$ para obtener una mejor estimación del error de predicción al evitar

el sobreajuste (Yahya and Muhammad, 2015).

El modelo de regresión de Poisson con número de eventos y_i tiene una distribución con media condicional que depende de las variables x_i según la función:

$$f(y_i) = \frac{e^{-\theta_i} \theta_i^{y_i}}{y_i!}, \quad y_i = 0, 1, \dots; \quad i = 1, 2, \dots, n \quad (1)$$

donde el parámetro $\theta_i = \exp(x_i' \beta)$.

Suponiendo además que las observaciones son independientes, podemos obtener la función de log-verosimilitud dada por:

$$\ell(\beta) = \sum_{i=1}^n (y_i x_i' \beta - \exp(x_i' \beta) - \ln(y_i!)) \quad (2)$$

Por lo tanto, la regresión Poisson penalizada (PRD) se define como:

$$PPR = \ell(\beta) + \lambda P(\beta) \quad (3)$$

donde λ se define como un parámetro de ajuste $\lambda \geq 0$, controlando la fuerza de contracción de las variables explicativas; cuando λ toma un valor mayor, se dará más peso al término de penalización.

Antes de resolver PPR , puede ser útil estandarizar x_i para obtener:

$$\frac{1}{n} \sum_{i=1}^n X_{ij} = 0, \quad \text{y} \quad \sum_{i=1}^n X_{ij}^2 = 1 \quad (4)$$

Para lograr que el intercepto $\beta_0 = 0$.

2.2.2. Regression Lasso

El método de regresión Lasso (Least Absolute Shrinkage and Selection Operator) es una técnica que combina la selección de variables y la regularización para mejorar la precisión de las predicciones y la interpretabilidad del modelo de regresión. Lasso logra esto añadiendo una penalización a la función de verosimilitud que restringe la suma de los valores absolutos de los coeficientes, forzando algunos de ellos a ser exactamente cero. El estimador Lasso es:

$$\widehat{\beta}_L = \arg \min_{\beta} \left(-\ell(\beta) + \lambda \sum_{j=1}^k |\beta_j| \right) \quad (5)$$

donde $\lambda \geq 0$ es el parámetro de ajuste que controla la intensidad de la penalización, asumiendo gran importancia. Para valores grandes de λ , los coeficientes se fuerzan a ser exactamente iguales a cero. De esta manera, se puede reducir la dimensionalidad (Young and Trevor, 2007).

2.2.3. Elastic-net

El método Elastic Net es una técnica de regularización que combina las penalizaciones Lasso (L1) y Ridge (L2) para mejorar la estabilidad y la precisión predictiva de los modelos de regresión en presencia de multicolinealidad y un gran número de variables predictoras. La combinación de estas penalizaciones ayuda a superar las limitaciones individuales de Lasso y Ridge, permitiendo la selección de variables y la reducción de la dimensionalidad de manera más efectiva que cada método por separado.

Según (Hui and Trevor, 2005), Elastic Net ofrece ventajas significativas sobre métodos como Lasso en términos de estabilidad y manejo de variables correlacionadas. El método permite controlar dos parámetros de ajuste, α (peso entre Lasso y Ridge) y λ (intensidad de la penalización), lo cual es importante para optimizar el balance entre sesgo y varianza en el modelo.

El estimador está dado por:

$$\widehat{\beta}_E = \arg \min_{\beta} \left(-\ell(\beta) + \lambda_1 \sum_{j=1}^k |\beta_j| + \lambda_2 \sum_{j=1}^k |\beta_j|^2 \right) \quad (6)$$

El segundo término (penalización Ridge) fomenta que las variables altamente correlacionadas se promedien, mientras que el primer término (penalización LASSO) fomenta una solución dispersa en los coeficientes de estas variables promediadas. El estimador Elastic Net para el modelo de regresión de Poisson depende de dos parámetros de ajuste no negativos λ_1 y λ_2 , y conduce a una solución de regresión de Poisson penalizada (Mwikali et al., 2019).

2.3 Modelos generalizados mixtos

Los modelos generalizados mixtos se enfocan en modelar respuestas de tal forma que se consideren tanto efectos fijos como aleatorios. Es así como se estipuló por primera vez en (Breslow and Clayton, 1993) la formulación matemática de un modelo generalizado mixto.

$$g(\mathbf{y}|\mathbf{b}) = \mathbf{X}\alpha + \mathbf{Z}\mathbf{b} \quad (7)$$

$$\mathbf{b} \sim N^q(\mathbf{0}, \mathbf{C})$$

donde $g(x)$ corresponde a una función de enlace que permite modelar respuestas no gaussianas de la familia exponencial, α representa el vector asociado a las covariables con efectos fijos \mathbf{X} , contrastado con la matriz \mathbf{Z} en la que se incluyen las q covariables con efectos aleatorios en el vector \mathbf{b} que siguen una distribución normal multivariante de dimensión q con vector de medias $\mathbf{0}$ y matriz de varianzas y covarianzas \mathbf{C} .

2.4 Modelos espacio-temporales

Como una extensión de los modelos longitudinales propuestos por (Diggle et al., 2002) en los que se consideran modelos con medidas repetidas a través del tiempo, se proponen también modelos en los que la ubicación se incluye como medida repetida, dando como resultado modelos espacio-temporales. Al ser el espacio y el tiempo unidades de magnitudes y soportes distintos, se consideran casos en los que ambos componentes de la función de covarianzas, que debe ser definida positiva, pueden ser separables y no separables. Las matrices no separables son más sofisticadas pero con mayor coste computacional, como las presentadas en (Cressie and Hsin-Cheng, 1999), mientras que las separables realizan cálculos de manera individual para el componente espacial y temporal, combinándolos en una misma matriz como se propone en (Gneiting et al., 2006) mediante un producto tensorial:

$$\mathbf{D} = \mathbf{D}_t \otimes \mathbf{D}_s$$

Donde \mathbf{D}_s es la matriz de varianzas y covarianzas del componente espacial, la cual es construida a partir de la función Matérn de covarianza espacial isotrópica (Cressie,

1993), mientras que \mathbf{D}_t corresponde al componente temporal y es modelada a través de dinámicas autorregresivas.

2.4.1. Integrated Nested Laplace Approximation

Integrated Nested Laplace Approximation (INLA) es un método para realizar inferencias en modelos estadísticos bayesianos de manera eficiente. INLA fue desarrollado como una alternativa frente a las limitaciones de los métodos tradicionales de simulación, como MCMC (Markov Chain Monte Carlo), que a menudo resultan ineficientes y lentos en problemas de alta dimensión.

Para un modelo bayesiano con parámetros θ y datos y , INLA estima la densidad posterior

$$p(\theta | y) \propto p(y | \theta) p(\theta)$$

descomponiéndola en componentes manejables usando la fórmula de Bayes.

La aproximación de Laplace se aplica para simplificar estas integrales complejas. Específicamente, para una función $g(\theta)$, la aproximación de Laplace es:

$$\int e^{g(\theta)} d\theta \approx e^{g(\hat{\theta})} \sqrt{\frac{(2\pi)^k}{|\det(-H(\hat{\theta}))|}} \quad (8)$$

donde $\hat{\theta}$ es el punto de máxima a posteriori (MAP) y $H(\hat{\theta})$ es la matriz Hessiana de $g(\theta)$ evaluada en $\hat{\theta}$.

INLA es particularmente útil para modelos espaciales y espacio-temporales, donde las dependencias complejas y la gran cantidad de datos hacen que otros métodos sean impracticables. Implementado en el paquete R-INLA, facilita la especificación de modelos y la obtención de resultados rápidos y precisos (Martino and Riebler, 2019).

2.5 Criterios de Información

2.5.1. Bayesian Information Criterion (BIC)

El BIC (Bayesian Information Criterion) es una medida estadística utilizada para seleccionar entre modelos de regresión o modelos de mezcla en función de su ajuste y complejidad. Se utiliza para penalizar los modelos que tienen un número excesivo de parámetros, ayudando a evitar el sobreajuste (*overfitting*).

La fórmula del BIC es igual a:

$$BIC = \ln(n) \cdot k - 2 \cdot \ln(\hat{L})$$

Donde:

- n : Número de observaciones
- k : Número de parámetros
- \hat{L} : Verosimilitud del modelo

El objetivo es elegir el modelo con el BIC más bajo, que indica un buen equilibrio entre el ajuste del modelo y la complejidad. (Schwarz, 1978)

2.5.2. Deviance Information Criterion (DIC)

El DIC (Deviance Information Criterion) es una medida utilizada en modelos estadísticos bayesianos para comparar la calidad de ajuste de diferentes modelos. A diferencia del BIC, que se basa en la verosimilitud máxima, el DIC toma en cuenta la devianza del modelo y el número efectivo de parámetros. El DIC es útil para modelos con incertidumbre en los parámetros y ayuda a balancear el ajuste del modelo con la complejidad, penalizando modelos que se ajustan demasiado bien a los datos de entrenamiento, pero que podrían no generalizar bien. (Spiegelhalter, Best, Carlin, Linde, & Best, 2002)

La expresión para calcular el DIC es:

$$DIC = \bar{D}(\theta) + p_D$$

Definimos la devianza como:

$$D(\theta) = -2\log(p(y | \theta)) + C,$$

donde y son los datos, θ son los parámetros desconocidos del modelo y $p(y | \theta)$ es la función de verosimilitud. C es una constante que se anula al comparar modelos y el valor esperado de la devianza corresponde a $\bar{D}(\theta)$.

Además, p_D es el número efectivo de parámetros, que se calcula como $\bar{D} - D(\hat{\theta})$, con $D(\hat{\theta})$ siendo la devianza evaluada en la media de la distribución *a posteriori* de los parámetros.

2.5.3. Watanabe Akaike Information Criterion (WAIC)

El WAIC es un criterio de información bayesiano utilizado para evaluar la calidad de ajuste de modelos estadísticos. Es una extensión del AIC y el DIC, que toma en cuenta la incertidumbre en los parámetros del modelo y se basa en la evidencia bayesiana. El WAIC es útil para comparar modelos en contextos donde los datos pueden no estar distribuidos de manera independiente y puede manejar modelos con alta complejidad. (Watanabe, 2010)

El WAIC se define como:

$$WAIC = -2(\text{lppd} - p_{WAIC})$$

Donde:

$$p_{WAIC} = \sum_{i=1}^n V_{\theta \sim p(\theta|Y)} [\log p(Y_i | \theta_s)],$$

$$p_{WAIC} = \sum_{i=1}^n V_{s=1}^S \log p(Y_i | \theta_s)$$

donde V es la varianza en log-verosimilitud para la observación i en la muestra de entrenamiento. Además, lppd es la función de verosimilitud evaluada en la media *a posteriori* del ajuste del modelo.

3 Metodología

3.1 Flujo de Trabajo

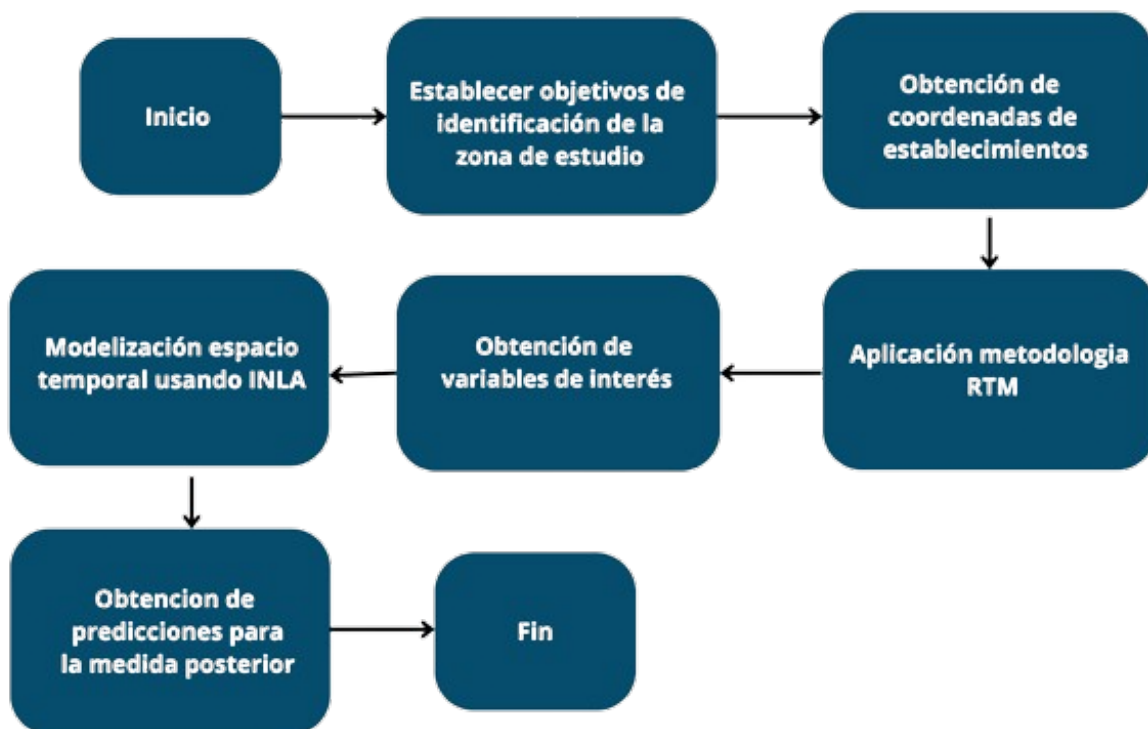


Figura 1: Diagrama de Flujo

3.2 Delimitación de la Zona de Estudio

La zona de estudio se delimitó en los cantones Guayaquil, Durán y Samborondón, que comprenden la Zona 8 de la delimitación administración por Zonas de planificación en el Ecuador. Con la obtención de las coordenadas de los polígonos para cada cantón se realizó una delimitación interna que consiste en la creación de cuadrículas de $500m^2$ lo que proporcionará más exactitud en los resultados obtenidos a la hora de predecir la cantidad de un delito específico dentro del área.

Adicionalmente, se localizaron las zonas más pobladas de los tres cantones y se delimitó el estudio dentro de estas áreas.

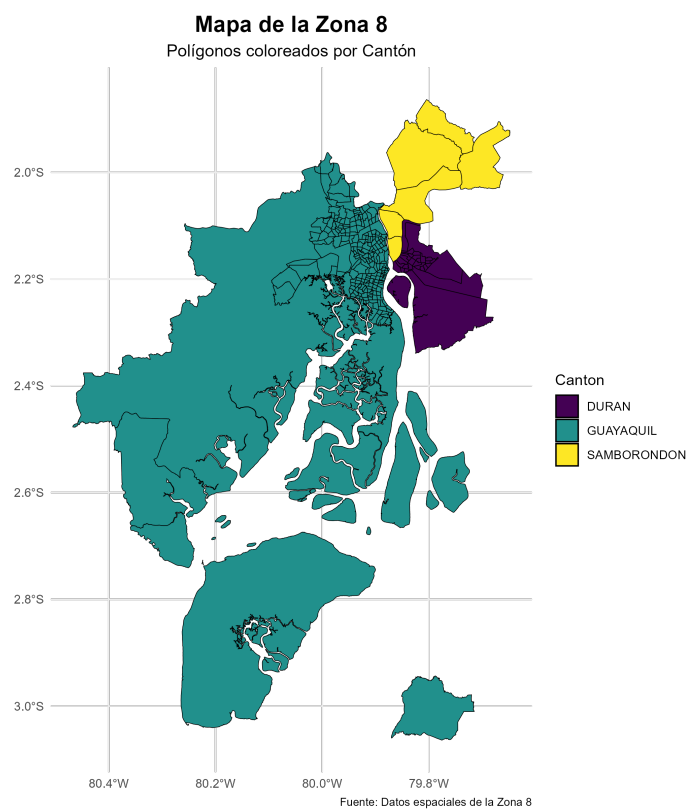


Figura 2: Mapa de la Zona 8

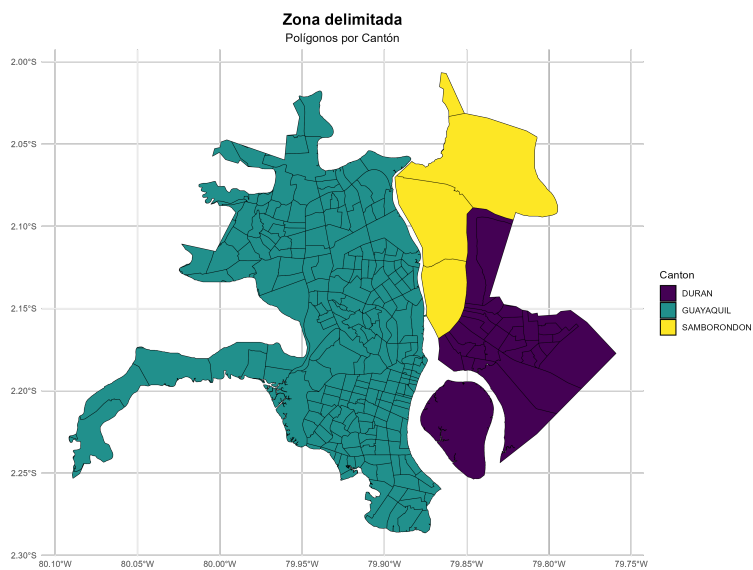


Figura 3: Área urbana de la Zona 8

3.3 Cruadrículas

Inicialmente se dividió el área urbana en cuadrículas de 1000 m^2 , esta será nuestra unidad de análisis. Se obtuvieron 520 cuadrículas. Se contabilizaron los incidentes de delitos y eran asignados a la cuadrícula más cercana en distancia euclidiana. Además, se repartieron los delitos en 135 semanas para estudiar su componente temporal, resultando en $520 \times 135 = 70.720$ cuadrículas espacio-temporales por cada delito. Para lo cual se consideraron 4 tipos de delitos relacionados con: 1. Drogas 2. Robos 3. Homicidios 4. Extorsión

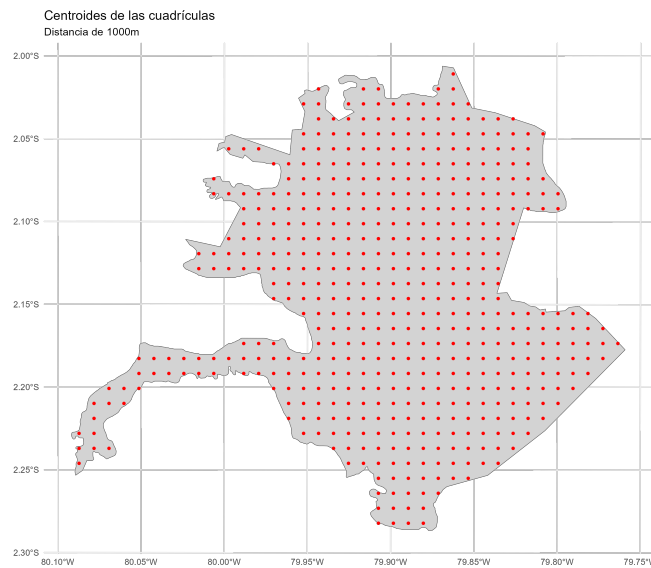


Figura 4: Centroides de las cuadrículas

3.4 Recolección de Datos

3.4.1. Incidentes

La recolección de datos se llevó a cabo a través de la obtención de la base de datos de reportes de incidentes proporcionada por el ECU911. Esta base de datos incluye variables clave como la fecha del incidente, coordenadas geográficas, tipo de delito, entidad enviada a la emergencia, ubicación, entre otras. Para el presente estudio, se decidió focalizar el análisis en cuatro tipos específicos de delitos: homicidios, tráfico de drogas, robos y extorsión, durante el período comprendido entre 2021 y 2024.

3.4.2. Factores de Riesgo

Para la obtención de las coordenadas de los establecimientos, se utilizó Google Maps, una herramienta que proporciona las coordenadas de diversos lugares o sitios de interés. A través de la construcción de algoritmos de extracción de datos mediante el API de Google, se identificaron 13 tipos de lugares que pueden considerarse como atractores o detractores de crimen, de acuerdo con la teoría de lugares de riesgo, cuyo detalle de extracción se encuentra contemplado en Tabla 1

Tipo	Especificación
Banco	Banco, banco comercial, banco estatal, caja de ahorros, cooperativa de ahorro y crédito, sucursal bancaria, oficina bancaria, entidad financiera, casa de cambio, cajeros automáticos
Bares	Bar, licorería, taberna, discoteca, night club, pub, lounge, club nocturno, cantina, coctelería, cervecería, wine bar, cocktail bar, sports bar, karaoke bar, dive bar, bar and grill, bar de tapas, bar de vinos, bar de cócteles, tienda de licores del estado, cervecería artesanal, fábrica de cerveza, tienda de bebidas alcohólicas
Paraderos de Buses	Estación de autobuses, paradero de buses, taxi stand, terminal de autobuses, transporte público, parada de taxis
Instituciones Educativas	Universidad, colegio, instituto, centro de educación, escuela pública, escuela privada, colegio bilingüe, secundaria, preparatoria, jardín de infancia
Garajes	Estacionamiento subterráneo, aparcamiento, parking, garaje, zona de aparcamiento, parking público, aparcamiento cubierto
Gasolineras	Gasolinera, estación de servicio, fuel station, service station, estación de gasolina
Centros Comerciales	Supermercado, centro comercial, hipermercado, shopping mall, department store, tienda de abarrotes, mercado
Diversión Nocturna	Club nocturno, discoteca, bar musical, sala de conciertos, speakeasy, karaoke bar, dance hall, bar con música en directo, cantina, jazz club
Centros de Recreación	Parque estatal, parque infantil, piscina, jardín, parque acuático, parque local, local park, national park, community park, playground
Comisaría de Policía	Comisaría, jefatura de policía, policía nacional, policía civil, estación de policía, oficina de seguridad
Restaurantes	Restaurante, cafetería, panadería, restaurante de comida rápida, heladería, marisquería, hamburguesería, pastelería, restaurante de fusión
Centros de Deportes	Estadio, polideportivo, campo de fútbol, basketball court, gimnasio, campo de béisbol, volleyball court, sports complex, pista de atletismo, centro deportivo
Tiendas	Tienda, tienda de abarrotes, mini market, despensa, tienda de alimentación, convenience store, farmers market

Tabla 1: Especificaciones de diferentes tipos de lugares

3.4.3. Variables

La metodología para obtener las variables que se utilizarán en el modelo de riesgo consiste en el cálculo de variables binarias basado en la cercanía de cada establecimiento al centro de las cuadrículas creadas, si se encuentran dentro de 200 m^2 , 500 m^2 y 1000 m^2 calculados a partir de las distancias euclidianas, se colocará el valor de 1, caso contrario, tendrá el valor de 0, creando tres variables binarias para cada factor. Adicionalmente, se contabilizó el total de establecimientos de cada tipo presentes en las cuadrículas. Este conteo permite obtener una variable adicional que refleja la densidad y diversidad de tipos de establecimientos en cada cuadrícula. A partir de estos cálculos, se obtuvieron un total de 52 variables. La variable dependiente Y considerada, es el conteo de delito de cada cuadrícula, para cada tipo de delito. Esta selección se llevó a cabo dado que consisten los delitos de mayor incidencia en el área.

3.5 Entrenamiento del Modelo de Riesgo

Se utilizó el modelo Elastic Net, que combina las características de regularización Lasso y Ridge. Esta técnica permite manejar situaciones con alta dimensionalidad y colinealidad entre variables, ayudando a seleccionar aquellas que tienen un impacto significativo en la variable dependiente, en este caso, el número de incidentes de acuerdo a cada tipo (Drogas, Robo, Homicidios, Extorsión). El modelo proporciona un conjunto de coeficientes que indican la fuerza y dirección de la relación entre las variables explicativas y la variable dependiente.

Basándose en los coeficientes obtenidos, se realiza una selección de variables que contribuyen significativamente al modelo. Se excluyen aquellas variables cuyos coeficientes no sean diferentes de cero, ya que no aportan información relevante al modelo.

Para optimizar aún más la selección, se implementa un procedimiento de selección de variables utilizando el criterio de información bayesiano (BIC). Durante este proceso, se prioriza la inclusión de variables relacionadas con la misma característica (por ejemplo, la proximidad a bancos) a diferentes distancias, eligiendo siempre la variable que representa la menor distancia. Este método busca minimizar el BIC al añadir o eliminar variables, seleccionando así el mejor modelo entre todos los posibles.

Con las covariables seleccionadas, se calcula el valor esperado λ de la distribución de Poisson, que representa el riesgo de que se cometa un tipo específico de delito en una zona determinada. Posteriormente, se dividen todos los valores esperados por el menor de ellos, obteniendo así una probabilidad relativa que indica el aumento del riesgo en comparación

con el área de menor riesgo.

4 Resultados y análisis

4.1 Estadística Descriptiva

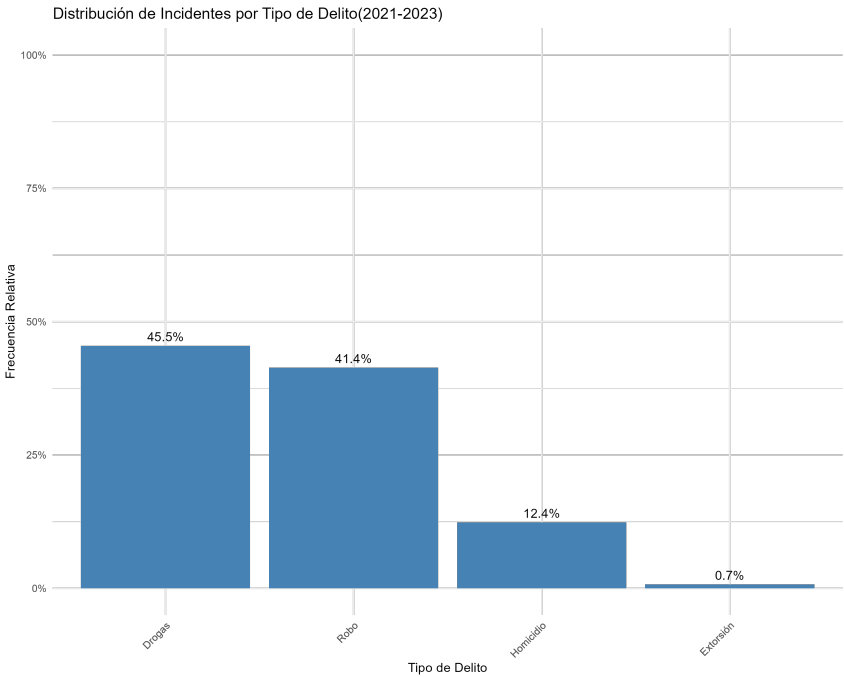


Figura 5: Frecuencia tipo de Crimen

Entre los años 2021 y 2023, el delito más reportado fue el relacionado con drogas, con un 45.5 % de los incidentes, seguido por los delitos de robo, que representaron un 41.4 %. Los homicidios constituyeron un 12.4 % del total, mientras que los delitos de extorsión fueron los menos frecuentes, con solo un 0.7 %. La Figura 5 ilustra esta distribución porcentual.

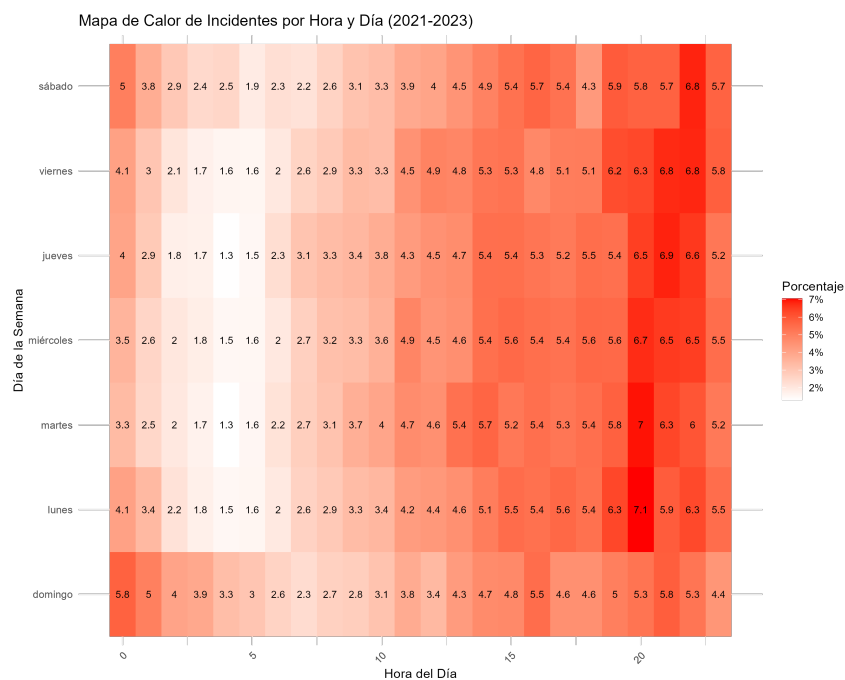


Figura 6: Mapa de Calor por Hora y Día

Se puede observar en la Figura 6 que existe una menor incidencia de los reportes durante las horas de la madrugada, entre las 00:00 y las 05:00, mientras que hay una mayor concentración de los reportes de delitos durante las horas más altas de la noche, entre las 18:00 y las 22:00.

4.2 Modelo de Riesgo

4.2.1. Valores de BIC para cada modelo

Se calcularon los valores del Criterio de Información de Bayes (BIC) para cada uno de los modelos ajustados a los datos. Los valores obtenidos para cada tipo de delito son los siguientes:

Delito	BIC
Drogas	142,757.607
Robos	25,993.200
Homicidios	1,046.728
Extorsión	370.830

Tabla 2: Criterio de Información de Bayes (BIC) para diferentes delitos

4.2.2. Drogas

Coefficient	Variable
3,432	(Intercept)
-0,689	factor_policias_dist_20011
-0,955	factor_gas_station_dist_20011
-0,397	factor_malls_dist_20011
0,319	factor_restaurants_dist_20011
-0,227	factor_sports_place_dist_20011
0,351	factor_transports_dist_20011
0,260	factor_garage_dist_20011
0,355	factor_bares_dist_50011
0,588	factor_stores_dist_50011
0,547	factor_banks_dist_100011
0,541	factor_education_dist_100011
0,600	factor_parks_dist_100011
-0,209	count_malls
-0,129	count_parks
0,059	count_stores
-0,040	count_night_centers
0,054	count_garage

Tabla 3: Coeficientes del modelo de riesgo Drogas

El coeficiente de cada variable independiente refleja el efecto multiplicativo en la tasa del delito. Por ejemplo, el intercepto de 3,432 indica el valor esperado del logaritmo natural de la tasa de delitos cuando todas las demás variables son cero, lo que indica que existe una tasa de alrededor de 30.93 delitos relacionados con drogas para cada zona en el año 2022. Si consideramos las variables binarias, como *factor_policias_dist_20011* con un coeficiente de $-0,689$, esto significa que estar dentro de 200 metros de una estación de policía reduce la tasa de delitos a la mitad, ya que $\exp(-0,689) \approx 0,502$. En contraste, un coeficiente positivo como el de *factor_bares_dist_50011* ($0,355$) muestra que estar cerca de bares aumenta la tasa de delitos en un 42.6 % ($\exp(0,355) \approx 1,426$). Además, los conteos como *count_malls* ($-0,209$) indican que cada centro comercial adicional disminuye la tasa de delitos en un 18.9 % ($\exp(-0,209) \approx 0,811$). Finalmente, un conteo positivo como el de *count_garage* ($0,054$) muestra que cada garaje adicional incrementa la tasa de delitos en un 5.6 % ($\exp(0,054) \approx 1,056$).



Figura 7: Mapa de Riesgo Drogas

4.2.3. Robos

Coefficient	Variable
2,830	(Intercept)
-0,484	factor_gas_station_dist_20011
-0,310	factor_malls_dist_20011
-0,142	factor_parks_dist_20011
0,359	factor_restaurants_dist_20011
0,297	factor_sports_place_dist_20011
0,354	factor_stores_dist_20011
0,551	factor_bares_dist_50011
0,552	factor_education_dist_50011
0,217	factor_transports_dist_50011
0,625	factor_banks_dist_100011
0,093	factor_night_centers_dist_100011
-0,138	factor_garage_dist_100011
0,041	count_education
0,169	count_gas_station
-0,020	count_restaurants
-0,032	count_night_centers
0,033	count_garage

Tabla 4: Coeficientes del modelo de riesgo para robos

Para la categoría de robo el coeficiente del intercepto es 2,830, lo que indica una tasa base esperada de delitos de robos en ausencia de otros factores, aproximadamente 16.9

delitos por unidad. La proximidad a una estación de gasolina y centros comerciales está asociada con una disminución en la tasa de robos ($-0,484$ y $-0,310$, respectivamente), lo que sugiere que estos lugares pueden tener un efecto protector. Por otro lado, estar cerca de bares y estaciones de transporte incrementa la tasa de robos, con coeficientes positivos de $0,551$ y $0,217$, respectivamente, indicando que estos lugares están asociados con un aumento en la tasa de delitos.

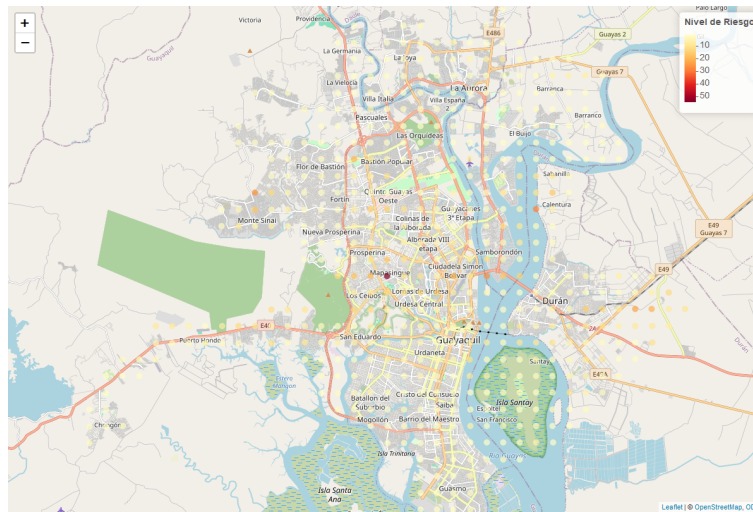


Figura 8: Mapa de Riesgo Robos

4.2.4. Homicidios

Coefficient	Variable
-1,941	(Intercept)
-0,798	factor_malls_dist_20011
-0,988	factor_parks_dist_20011
1,035	factor_stores_dist_50011
1,103	factor_bares_dist_100011
0,927	factor_sports_place_dist_100011
-1,200	factor_night_centers_dist_100011
0,259	count_policias
-0,129	count_garage

Tabla 5: Coeficientes del modelo de riesgo para Homicidios

Para el delito de homicidios el coeficiente del intercepto es $-1,941$, lo que implica que la tasa base esperada de homicidios es bastante baja cuando no se consideran otras variables,

aproximadamente 0.14 homicidios por unidad. La proximidad a centros comerciales y parques muestra efectos negativos en la tasa de homicidios con coeficientes de $-0,798$ y $-0,988$, respectivamente, lo cuál indica que estos lugares podrían estar asociados con una reducción en la tasa de homicidios. Por otro lado, estar cerca de tiendas y bares incrementa la tasa de homicidios, con coeficientes positivos de $1,035$ y $1,103$, indicando que estos lugares podrían estar vinculados con un aumento en la incidencia de homicidios.

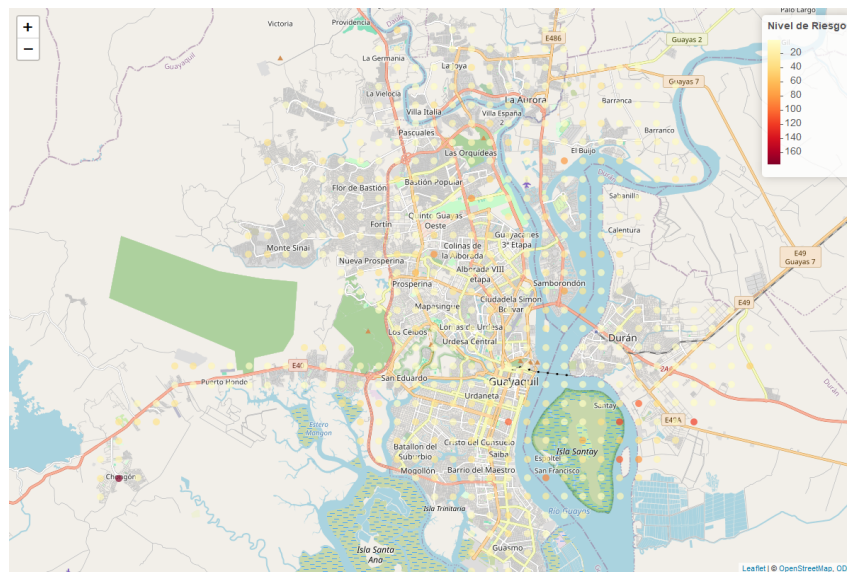


Figura 9: Mapa de Riesgo Homicidios

4.2.5. Extorsión

Coefficient	Variable
$-3,175$	(Intercept)
$0,846$	factor_parks_dist_20011
$0,957$	factor_education_dist_50011
$0,895$	factor_transports_dist_100011

Tabla 6: Coeficientes del modelo de riesgo para Extorsión

En el caso de los delitos de extorsiones el coeficiente del intercepto es $-3,175$, indicando que, en ausencia de otras variables, la tasa base esperada de extorsiones es extremadamente baja, cerca de 0.042 por unidad. El estar cerca a parques y centros educativos está asociada con un incremento en la tasa de extorsiones, con coeficientes de $0,846$ y $0,957$, respectivamente. Así mismo, la cercanía a centros de transporte también está asociada con un aumento en la tasa de extorsiones, con un coeficiente de $0,895$.



Figura 10: Mapa de Riesgo Extorsión

4.3 Modelo Predictivo

Se calcularon los criterios WAIC, DIC y la log-verosimilitud marginal de la a posteriori de la variable de respuesta para los delitos de Homicidios y Tráfico de drogas de donde se consiguieron los siguientes resultados:

Modelo	WAIC	DIC	MLIK
Hurdle	20213.06	20209.054	-10186.754
ZIP	10735.04	9993.171	-4707.174
NegBinom	11554.72	10008.774	-4703.056

Tabla 7: Criterios de Información de modelos predictivos para Homicidios

Modelo	WAIC	DIC	MLIK
Hurdle	119679.72	116014.35	-59199.32
ZIP	89195.24	86203.66	-44399.60
NegBinom	80298.21	80330.20	-40713.57

Tabla 8: Criterios de Información de modelos predictivos para Tráfico de Drogas

Los criterios sugieren que entre las distribuciones consideradas para ajustar los conteos de Homicidios, la de Poisson con ceros inflados (ZIP) resulta ser la mejor, mientras que para las incidencias de Tráfico de drogas resulta ser la mejor la Binomial Negativa.

A continuación se presentan las predicciones de cada tipo de delito en la primera semana de julio para todas las cuadrículas construidas.

4.3.1. Drogas

Las incidencias de delitos relacionados con drogas se concentran en el centro del área urbana de la Zona 8 extendiéndose al oeste por Vía a la Costa.

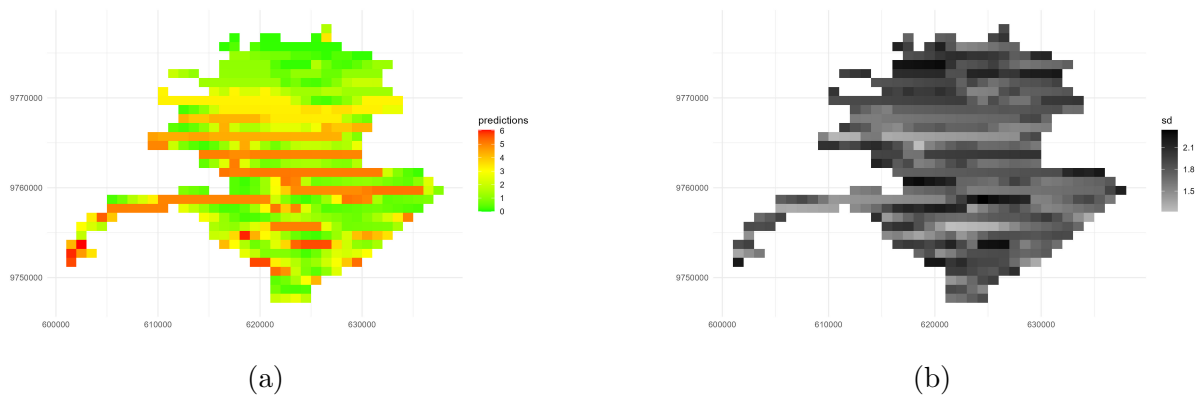


Figura 11: Predicciones de medias (a) y desviaciones estándar (b) a posteriori para las incidencias de Tráfico de Drogas

4.3.2. Robos

Se presenta concentración de incidencias de Robos en lo que corresponde a las fronteras de lo que corresponde al área delimitada, más específicamente al norte, oeste y ligeramente en el este.

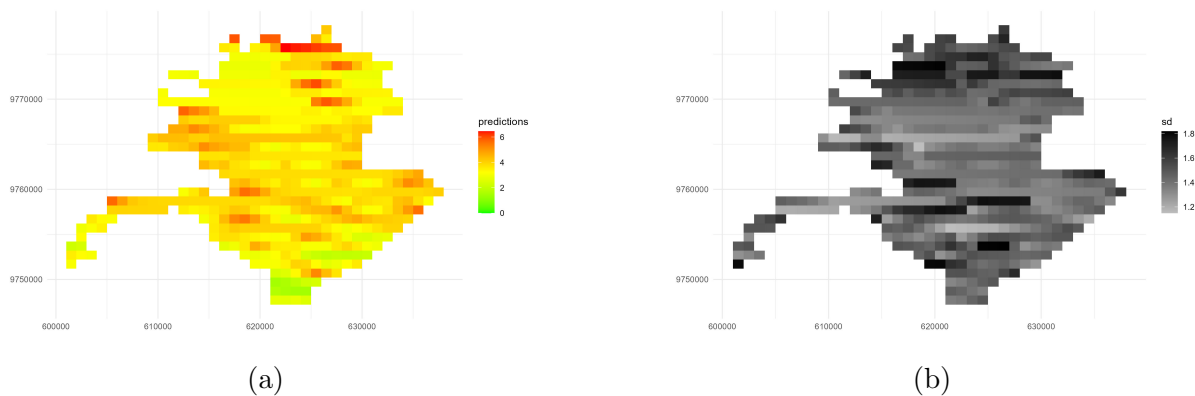


Figura 12: Predicciones de medias (a) y desviaciones estándar (b) a posteriori para las incidencias de Robos

4.3.3. Homicidios

La principal concentración de casos de homicidios se presenta en el norte del área delimitada y un comportamiento similar al sur extremo, aunque estas medias posteriores son muy cercanas a cero, pero en comparación al resto de cuadrículas, estas denotan probabilidades mayores para incidentes superiores a cero.

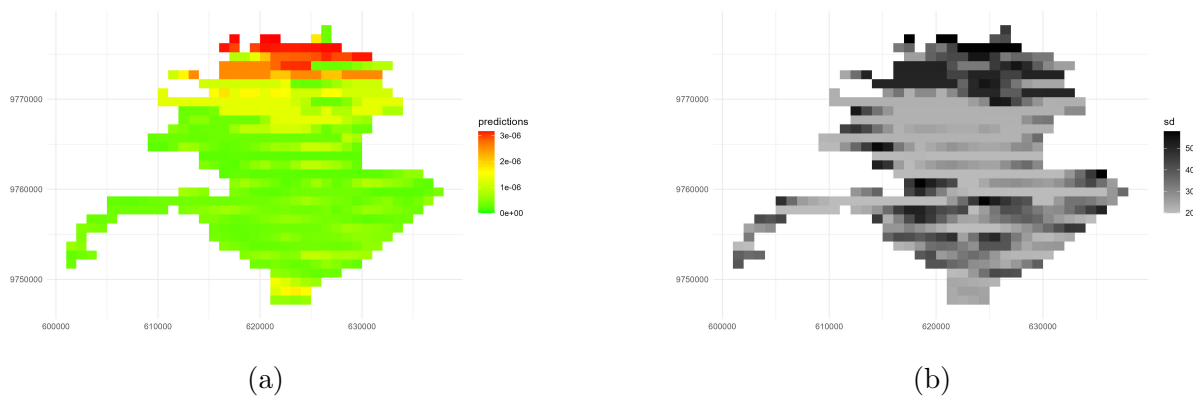


Figura 13: Predicciones de medias (a) y desviaciones estándar (b) a posteriori para las incidencias de Homicidios

4.3.4. Extorsiones

Se presenta un comportamiento similar a los homicidios acumulándose en el norte resaltando un punto caliente de delitos.

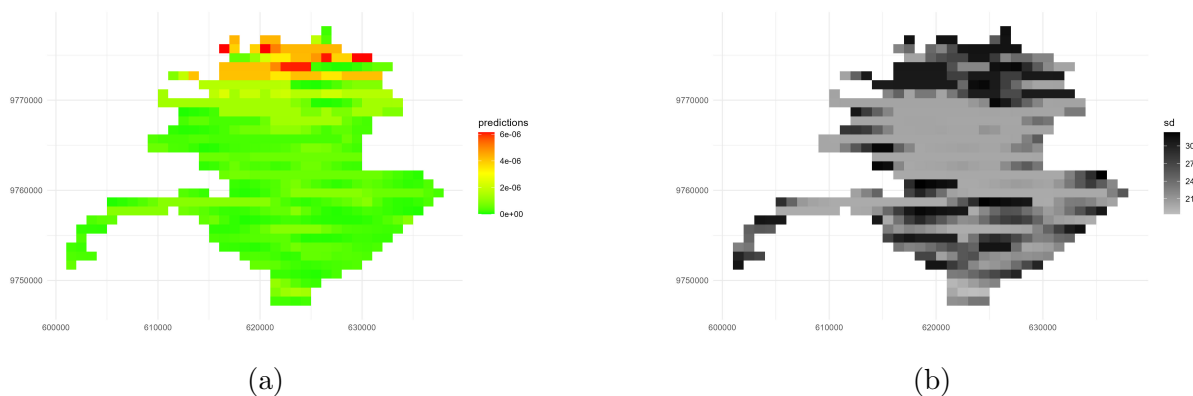


Figura 14: Predicciones de medias (a) y desviaciones estándar (b) a posteriori para las incidencias de Extorsiones

En general, en lo que respecta a las desviaciones estándar, los mapas denotan alta

variación de incidentes en lo que respecta al norte del área delimitada

5 Consideraciones finales

5.1 Conclusiones

En lo que corresponde a RTM se puede concluir que efectivamente hay una relación de conteo y cercanía con ciertos establecimientos que pueden ser considerados como atractores, detractores o generadores de crímenes dando como resultado un índice que caracterice una coordenada en base a factores ambientales.

De los modelos predictivos se pudo rescatar que, a pesar que en todos los tipos de delitos se consideran conteos de incidentes, estos tienen comportamientos distintos pero que sí podrían estar guardando cierta relación como sucede con los homicidios y las extorsiones los cuales tuvieron ambas altas concentraciones de incidencia promedio en el área norte de la zona delimitada. Por otra parte, lo que se puede capturar de las desviaciones estándar es que el área norte posee alta variabilidad a través del tiempo, pues se da una idea de que tal vez hay saltos relativamente grandes entre una semana y otra en lo que respecta a conteo de incidentes.

5.2 Recomendaciones

Se sugiere que en estudios posteriores se exploren nuevas propuestas de distribuciones a priori, pues como es de conocimiento, los métodos de carácter bayesiano dependen en gran parte del conocimiento previo de los parámetros a estimar.

Los modelos están diseñados también de tal manera que puedan ser extendidos a zonas geográficas más extensas, por lo que se sugiere que se considere expandir el estudio empezando por incluir el área rural de la Zona 8 hasta llegar al punto de poder capturar relaciones espacio-temporales en el resto del Ecuador, o al menos las ciudades de mayor índice delictivo.

Se recomienda también considerar en estudios próximos añadir variables explicativas socioeconómicas para complementar la información ambiental de los establecimientos que se consideraron en este proyecto.

Por último, se pretende también estudiar a futuro la correlación entre delitos de tal manera que la variable de respuesta, en este caso conteo de incidentes, se modele de

manera multivariante.

Referencias

- Bravo, M. (2024). Modelización espacio-temporal para la predicción de homicidios. caso de estudio: Guayaquil. Technical report, Escuela Superior Politécnica del Litoral, Guayaquil, Guayas, Ecuador.
- Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*.
- Britt, C., Rocque, M., and Zimmerman, G. (2017). The analysis of bounded count data in criminology. *Journal of Quantitative Criminology*, pages 1–19.
- Caplan, J. and Kennedy, L. (2016). *RISK TERRAIN MODELING*. The Regents of the University of California, Oakland, California.
- Chandra, B., Gupta, M., and Gupta, M. P. (2008). A multivariate time series clustering approach for crime trends prediction. In *Institute of Electric and Electronic Engineers (IEEE)*, pages 99–106.
- Chen, P., Yuan, H., and Shu, X. (2008). Forecasting crime using the arima model. *Institute of Electric and Electronic Engineers (IEEE)*, pages 1320–1330.
- Cressie, N. (1993). Statistics for spatial data. *Wiley*.
- Cressie, N. and Hsin-Cheng, H. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*.
- Diggle, P. J., Heagerty, P., Liang, K.-Y., and Zeger, S. L. (2002). *Analysis of Longitudinal Data*. Oxford University Press.
- Gneiting, T., Genton, M., and Guttorp, P. (2006). *Statistical methods for spatio-temporal systems*. Chapman and Hall/CRC.
- González, B., Munilla, T., Ayala, N., and Famea, B. (2021). Usos de inteligencia artificial y machine learning en. In *Universidad Tecnológica Nacional, Facultad Regional Buenos Aires*.
- Hossain, S., Abtahee, A., Kashem, I., Hoque, M. M., and Sarker, I. (2020). Crime prediction using spatio-temporal data. *Communications in Computer and Information Science (CCIS)*, pages 213–227.
- Hui, Z. and Trevor, H. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*.

- Kianmehr, K. and Alhajj, R. (2006). Crime hot-spots prediction using support vector machine. In *Institute of Electric and Electronic Engineers (IEEE)*, pages 487–492.
- Luo, L., Deng, M., Shi, Y., Gao, S., and Liu, N. (2022). Associating street crime incidences with geographical environment in space using a zero-inflated negative binomial regression model. *Cities*, pages 153–165.
- Martino, S. and Riebler, A. (2019). Integrated nested laplace approximations (inla). *arXiv.org*.
- McClendon, L. and Meghanathan, N. (2015). Using machine learning algorithms to analyze crime data. *Machine Learning and Applications: An International Journal (MLAIJ)*, pages 1–12.
- Muneer, U., Letchmunan, S., Hafnaz, F., Ali, M., Baqir, A., Wei, T., and Raza, H. (2021). Spatio-temporal crime predictions by leveraging artificial intelligence for citizens security in smart cities. *IEEE*, pages 1121–1133.
- Mwikali, J., Mwalili, S., and Wanjoya, A. (2019). Penalized poisson regression model using elastic net and least absolute shrinkage and selection operator (lasso) penalty. *International Journal of Data Science and Analysis*.
- OECD (2023). Boletín anual de homicidios. Obtenido de <https://oeco.pdf.org/wp-content/uploads/2024/04/OECO.-BOLETIN-ANUAL-DE-HOMICIDIOS-2023.pdf>.
- Ordoñez-Eraso, H.-A., Pardo-Calvache, C.-J., and Cobos-Lozada, C.-A. (2020). Detección de tendencias de homicidios en colombia usando machine learning. *Universidad Pedagógica y Tecnológica de Colombia (UPTC)*, pages 1–15.
- Porter, J., Capellan, J., and Chintakrindi, S. (2015). Social disorganization. *The Encyclopedia of Crime amp; Punishment*.
- Rummens, A., Hardyns, W., and Pauwels, L. (2017). The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context. *Applied Geography*, pages 120–131.
- Shengping, Y. and Berdine, G. (2015). Poisson regression. *The Southwest Respiratory and Critical Care Chronicles*.
- Shrivastav, A. K. and Ekata, D. (2012). Applicability of soft computing technique for crime forecasting: A preliminary investigation. *International Journal of Computer Science & Engineering Technology (IJCSET)*, pages 99–105.

- Vaillant, J. (1991). Negative binomial distributions of individuals and spatio-temporal cox processes. *Scandinavian Journal of Statistics*, pages 117–130.
- Yahya, A. Z. and Muhammad, H. L. (2015). Penalized poisson regression model using adaptive modified elastic net penalty. *Electronic Journal of Applied Statistical Analysis*, pages 236–245.
- Young, P. M. and Trevor, H. (2007). L1-regularization path algorithm for generalized linear models. *Journal of the Royal Statistical Society*.
- Zevallos, N., Mujica, J., and Vásquez, C. C. (2024). Amenazas a la seguridad del estado en ecuador. Obtenido de https://www.prosegurresearch.com/dam/jcr:0d29c565-b6cb-41c4-8d29-0ba16f1f908f/Informe_Ecuador.pdf.