



Revista de Ciencias Sociales

Depósito legal ppi 201502ZU4662
Esta publicación científica en formato
digital es continuidad de la revista impresa
Depósito Legal: pp 197402ZU789
● ISSN: 1315-9518 ● ISSN-E: 2477-9431

Universidad del Zulia. Revista de la Facultad de Ciencias Económicas y Sociales
Vol. XXXII, No. 2

Abril – Junio 2026

Revista de Ciencias Sociales

Esta publicación científica en formato
digital es continuidad de la revista impresa
Depósito Legal: pp 197402ZU789
ISSN: 1315-9518

Modelos de regresión logística y redes neuronales para la predicción de incidencia delictiva: Región Interandina – Ecuador

Gallegos-Meza, José Carlos*
Bolaños-Logroño, Paulina Fernanda**

Resumen

En la región interandina del Ecuador, donde las dinámicas sociales y económicas son heterogéneas, la integración de modelos de regresión logística y redes neuronales ofrece herramientas con proyección positiva para mitigar y anticipar la incidencia delictiva, debido a que ambos enfoques brindan insumos sólidos y significativos en la predicción de delitos. Este artículo se propuso como objetivo general describir las ventajas que ofrecen las redes neuronales y la regresión logística como herramientas heurísticas y de aplicación práctica para predicción de incidencia delictiva, particularmente bajo el contexto social. Metodológicamente, es un estudio cuantitativo comparativo de clasificación multivariante orientado a evaluar la funcionalidad de modelos socio estadísticos en la interpretación de delitos aplicados a patrones espaciotemporales de criminalidad en Ecuador. Estos hallazgos destacaron que la regresión logística multinomial tiene mejor capacidad de respuesta en la predicción de franjas horarias y provincias, como Balanced Accuracy y Kappa de Cohen, coincidiendo con otros estudiosos, quienes encontraron que los modelos lineales son excelentes en problemas de correlaciones entre variables con magnitud intermedia. Aunque la red neuronal reflejó una mejor eficacia en AUC, evidenciando su capacidad para clasificar entre clases no lineales, la regresión logística sigue siendo más efectiva en problemas con relaciones simples y moderadas.

Palabras clave: Regresión logística; redes neuronales; criminalidad; predicción del delito; región interandina del Ecuador.

* Ingeniero en Estadística. Investigador Independiente en Ecuador. E-mail: jose.gallegos@esepoch.edu.ec ORCID: <https://orcid.org/0009-0007-6831-6131>

** Doctoranda en Estadística Multivariante Aplicada en la Universidad de Salamanca, España. Magister en Sistemas de Control y Automatización Industrial. Magister Universitario en Estadística Aplicada. Ingeniera en Electrónica y Computación. Docente Investigador en la Escuela Superior Politécnica de Chimborazo, Riobamba, Chimborazo, Ecuador. Miembro del Grupo de investigación BI-DATA. E-mail: paulina.bolanos@esepoch.edu.ec ORCID: <https://orcid.org/0000-0003-3911-0461>

Logistic regression models and neural networks for predicting crime incidence: Inter-Andean Region – Ecuador

Abstract

In the Andean region of Ecuador, where social and economic dynamics are heterogeneous, the integration of logistic regression models and neural networks offers promising tools for mitigating and anticipating crime, as both approaches provide robust and significant input for crime prediction. This article aimed to describe the advantages of neural networks and logistic regression as heuristic tools with practical applications for predicting crime, particularly within the social context. Methodologically, it is a comparative quantitative study of multivariate classification designed to evaluate the functionality of socio-statistical models in interpreting crimes applied to spatiotemporal patterns of criminality in Ecuador. The findings highlighted that multinomial logistic regression has better responsiveness in predicting time periods and provinces, as demonstrated by Balanced Accuracy and Cohen's Kappa, coinciding with other studies that found linear models to be excellent for problems involving correlations between variables of intermediate magnitude. Although the neural network showed better efficiency in AUC, demonstrating its ability to classify between nonlinear classes, logistic regression remains more effective in problems with simple and moderate relationships.

Keywords: Logistic regression; neural networks; criminality; crime prediction; Andean region of Ecuador.

Introducción

La predicción del crimen se ha unificado, desde finales del siglo XX, como un dominio interdisciplinario en el que convergen la estadística, criminología, y, más recientemente, el aprendizaje automático. Estudios pioneros evidencian que los crímenes no suceden de manera casual (Sherman et al., 1989), reflejando que la criminalidad tiende a concentrarse en focos urbanos muy pequeños; mientras que Ratcliffe (2003) fortaleció el enfoque espacio temporal al demostrar la persistencia de patrones que pueden ser anticipados.

A comienzos de 2010, el aumento de bases de datos policiales digitalizadas, el uso de registros administrativos en tiempo casi real y geocodificación masiva, impulsaron una transformación profunda en este campo. Como señalan Perry et al. (2013), el conjunto de datos masivos con métodos estadísticos sofisticados permitió que los modelados

predictivos se modificaran en una herramienta sustancial para guiar la toma de decisiones en seguridad pública.

En Ecuador, las aplicaciones reglamentadas de analítica predictiva han sido nulas, algunas investigaciones destacan por su distinguida aportación derivando precedentes significativos. Chávez-Cadena y Abril-Donoso (2023) por ejemplo, realizaron la técnica de regresión logística para augurar la probabilidad de homicidios en la ciudad de Quito, destacando áreas importantes y censuras directas para la prevención situacional.

No obstante, varios investigadores señalan los peligros de los modelos que conlleva la “caja negra”. Lum e Isaac (2016), comprobaron a través de su investigación que algoritmos no transparentes pueden emularse y generar anomalías tradicionales afectando desproporcionadamente en sectores con situación de grado de alta exposición. El instituto *AI Now Institute* (Richardson et al., 2019) concluyó que los agentes policiales incumplen y desconfían de sistemas que

generan predicciones sin hipótesis claras sobre el discernimiento interno de los modelados de predicción, lo que ha llevado al descarte de herramientas como *PredPol* en diversas zonas estratégicas de Estados Unidos.

Bajo este contexto, un análisis comparativo de técnicas como la regresión logística y redes neuronales (Luna et al., 2025), es idóneo para prever la incidencia delictiva en la región interandina del Ecuador, empleando información oficial del periodo 2014-2025. El enfoque metodológico comprende dos etapas: En primer lugar, se desarrolla una fase interpretativa a través de un análisis del uso de la técnica de regresión logística; y, en segundo lugar, se lleva a cabo un estudio dirigido a maximizar el desempeño predictivo mediante redes neuronales que han sido entrenadas sobre las variables más significativas. Esta configuración nace de las sugerencias promovidas por Rudin (2019); y de Corbett-Davies et al. (2023), recomendando lo fundamental entre clasificar entre la lógica de

explicación para garantizar justicia algorítmica y reducir errores sistemáticos inadvertidos.

1. Fundamentación teórica

1.1. Teorización respecto a los procesos y procedimientos estadísticos y metodológicos utilizados

El enfoque clásico de modelado se fundamentó en la aplicación sistemática de técnicas estadísticas tradicionales para la selección y validación de variables predictoras. Se estructuró en múltiples etapas secuenciales, todas diseñadas para minimizar su alta dimensionalidad del espacio de predictores mientras se preserva la capacidad interpretativa de los modelos. Al respecto, en el Cuadro 1 se evidencia un resumen de esas etapas y técnicas aplicadas luego en tratamiento estadístico metodológico.

Cuadro 1
Descripción de las etapas secuenciales

Etapa	Breve descripción
Matriz de correlación	Se calcularon matrices de correlación con la técnica estadística de Spearman.
Factor de Inflación de la Varianza (VIF)	Se evaluó la presencia de multicolinealidad mediante el cálculo del factor de inflación de la varianza
Prueba ANOVA	La selección formal de variables se inició mediante la aplicación sistemática de pruebas de independencia chi-cuadrado entre cada predictor candidato y la variable objetivo
Tratamiento del desbalanceo de clases (He y García, 2009)	Para reducir este problema, se elaboró una estrategia de sobremuestreo de las clases minoritarias específicamente en el conjunto de entrenamiento.
Regresión logística multinomial con regularización Ridge	Para (Friedman et al., 2010) la regresión logística multinomial es una modelación generalizada del modelo de regresión logística clásica para problemas de clasificación con más de dos categorías, es por ello que se ajustó un modelo de regresión logística multinomial con regularización de tipo Ridge (Hoerl y Kennard, 1970).
Interpretación mediante razones de momios	En el modelo prevalece la interpretabilidad directa con el cálculo de razones de momios.
Red neuronal artificial multiclasa	La modelación predictiva para las tareas de clasificación multiclasa franjas horarias y provincias se realizó mediante una red neuronal multicapa (MLP) entrenada con retro propagación, correspondiendo al paradigma clásico de redes <i>feedforward</i> descrito por (Rumelhart et al., 1986).

Fuente: Elaboración propia, 2026.

1.2. Fundamentos teóricos de la arquitectura empleada

El preprocesamiento se limitó a tres operaciones fundamentales: Se realizó la codificación *one-hot* de todas las variables categóricas existentes en el conjunto $V_{cat} = \{V_1, V_2, \dots, V_q\}$, generando el conjunto completo de variables indicadoras binarias:

$$D = \{D_1, D_2, \dots, D_p\} \text{ donde } p = \sum_{i=1}^q (|V_i| - 1)$$

Y $|V_i|$ denota la cardinalidad de la variable categórica V_i .

Posteriormente se eliminaron aquellas variables indicadoras D_j para las cuales:

$$\sum_{i=1}^n D_{ij} < 30$$

Es decir, variables con menos de 30 observaciones positivas en el conjunto de datos completo. Se estandarizaron todas las variables mediante la transformación previamente mencionada, con parámetros μ_m y σ_m exclusivamente calculados sobre los conjuntos de entrenamiento de cada *fold*.

La regresión logística multinomial es una extensión del modelo de regresión logística binaria para problemas de clasificación con más de dos categorías. Se ajustó un modelo de regresión logística multinomial con regularización de tipo Ridge (Friedman et al., 2010). Si se denotan las K categorías como $1, 2, \dots, K$ y se toma la categoría K como referencia, el modelo especifica que para cada categoría $j \in \{1, 2, \dots, K-1\}$:

$$\log \left(\frac{P(Y=j|X=x)}{P(Y=K|X=x)} \right) = \beta_{j0} + \sum_{m=1}^p \beta_{jm} x_m$$

Donde: $\mathbf{x} = (x_1, x_2, \dots, x_p)$ = vector de predictores; β_{j0} = intercepto específico para la comparación entre categoría j y categoría de referencia K ; β_{jm} = coeficiente que cuantifica el efecto del predictor m -ésimo sobre el log-odds de pertenecer a la categoría j versus K .

Para prevenir el sobreajuste y mejorar la estabilidad numérica, se aplicó regularización de tipo Ridge. Esta técnica consiste en maximizar la función de log-verosimilitud penalizada:

$$\ell_{pen}(\beta) = \ell(\beta) - \lambda \sum_{j=1}^{K-1} \sum_{m=1}^p \beta_{jm}^2$$

Donde $\lambda \geq 0$ es el parámetro de regularización que controla la intensidad de la penalización. Para la regresión logística multinomial se diagnosticó con el conjunto de valores:

$$\mathcal{H}_{GLM} = \{\lambda: \lambda \in \{0, 0.0001, 0.001, 0.005, 0.01, 0.05, 0.1, 0.2\}\}$$

Con $|\mathcal{H}_{GLM}| = 8$ ajustes. El valor $\lambda = 0$ equivale al modelo que no presenta regularización, permitiendo evaluar si la regularización efectivamente maximiza el desempeño.

La red neuronal aplicada en el estudio se elaboró con una única capa oculta, configuración que se considera suficiente para asegurar la universalidad de la aproximación conforme al teorema de Cybenko (1989), quien evaluó que una red *feedforward* con una sola capa oculta y función sigmoide puede calcular cualquier función continua en un hiperespacio compacto.

$$f(x) = \sum_{j=1}^m \alpha_j \sigma(w_j^T x + b_j)$$

Así, una capa oculta constituye la arquitectura más parsimoniosa con suficiente capacidad representacional para capturar interacciones no lineales entre predictores categóricos. Para cada neurona j en la capa oculta, el potencial de activación se definió como:

$$a_j = w_j^T x + b_j$$

Y la activación correspondiente se obtiene mediante una función logística sigmoide, que introduce la no linealidad esencial:

$$z_j = \sigma(a_j) = \frac{1}{1 + e^{-a_j}}$$

El uso de la función sigmoide se consolida por su firmeza, diferenciabilidad continua y eficacia empírica en MLP clásicos para clasificar (Nasrabadi, 2006). En los

experimentos realizados, el hiperparámetro tamaño se varió entre 10 y 30 neuronas, considerando arquitecturas con diferente capacidad expresiva. El rango mencionado, incorpora distintas indicaciones metodológicas para prevenir que exista cualquier tipo de subajuste, al igual que sobreajustes (Hastie et al., 2009).

Para un problema con k categorías (tanto en franjas horarias como en provincias), la capa de salida computa un vector de *logits*:

$$a_k = \beta_k^T z + c_k, k = 1, \dots, K.$$

$$\mathcal{H}_{NN} = \{(h, \lambda): h \in \{10, 15, 20, 25, 30\}, \lambda \in \{0.001, 0.005, 0.01, 0.05, 0.1, 0.2\}\}$$

Resultando en $|\mathcal{H}_{NN}| = 30$ configuraciones distintas.

Para cada configuración $(h^*, \lambda^*) \in \mathcal{H}_{NN}$ y cada *fold* (j, s) , se entrenó una red neuronal con neuronas ocultas y parámetro de regularización λ^* , obteniendo una estimación del desempeño $\hat{\theta}_{h^*, \lambda^*}^{(j, s)}$ para una métrica objetivo θ .

2. Metodología

Se desarrolló un estudio cuantitativo comparativo de clasificación multivariante orientado a evaluar el balance entre desempeño predictivo e interpretabilidad en modelos de aprendizaje supervisado aplicados en patrones espaciotemporales de criminalidad en Ecuador (Chainey y Ratcliffe, 2013). Este estudio se desarrolló en la región interandina, una zona central de territorio ecuatoriano debido a la planificación de políticas de seguridad ciudadana en el intervalo de tiempo desde enero del 2014 a septiembre del 2025.

Se parte de tres hipótesis principales. En primer lugar, se postuló que hay factores contextuales y sociodemográficos que explican de manera sistemática la distribución del crimen en franjas horarias y provincias (Mohler et al., 2011). En segundo lugar, se propuso un modelo orientado en regresión logística multinomial que brinda un rendimiento

Los cuales se transforman en probabilidades mediante la función *softmax* (Bridle, 1990):

$$p(y = k | x) = \frac{e^{a_k}}{\sum_{l=1}^K e^{a_l}}$$

Así, el modelo estima directamente la distribución de probabilidad sobre todas las clases posibles, permitiendo trabajar con categorías nominales no ordenadas, como las franjas del día o los territorios provinciales. En la red neuronal se evaluó minuciosamente todas las combinaciones posibles en el conjunto:

comparable a una red neuronal multicapa en tareas de clasificación territorial y temporal del crimen. Finalmente, se planteó que la implementación de métodos de regularización y validación cruzada posibilita la optimización al mismo tiempo del rendimiento predictivo y de la estabilidad de las estimaciones en los dos enfoques.

También se usaron registros administrativos oficiales documentados por el Ministerio del Interior del Ecuador, que abarcan el intervalo de tiempo correspondientes entre enero de 2014 y septiembre de 2025. Estos registros fueron organizados, limpiados y ofrecidos al público mediante el portal nacional de datos abiertos del Gobierno del Ecuador (<<https://www.datosabiertos.gob.ec>>), lo que garantiza su seguimiento, transparencia y acceso conforme a criterios institucionales bajo estándares institucionales de gestión de información.

Después de un proceso de auditoría y limpieza que incluyó la validación de las áreas geográficas, identificación de valores atípicos y la comprobación de la coherencia temporal, mediante análisis de frecuencias marginales, donde se mantuvieron únicamente las variables categóricas estadísticamente significativas para desarrollar modelos explicativos y perfiles de riesgo.

a. Variables del entorno: El tipo de arma

que es utilizada, una presunta motivación para cometer el crimen, lugar específico del acontecimiento, el área donde ocurren los hechos (urbana / rural), tipo de muerte que se registra.

b. Variables sociodemográficas de la víctima: Etnia (autoidentificación), género, registro o existencia de algún antecedente penal, si presenta alguna discapacidad, nacionalidad de origen, nivel de instrucción formal y estado civil.

c. Variables temporales: Variable binaria de suceso, la cual puede estar conectada ya sea con el fin de semana o entre semana.

d. Variables predichas:

d.1. Franja horaria: El amanecer, dado entre las 00:00 am hasta las 05:59 am; la mañana, lapso entre las 06:00 am hasta las 11:59 am; la tarde, lapso entre las 12:00 pm hasta las 17:59 pm; y, finalmente, la noche, lapso entre las 18:00 pm hasta las 23:59 pm.

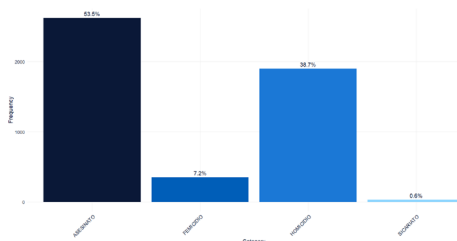
d.2. Provincia: Se establece como una clasificación nominal, entre distintas segmentaciones a nivel administrativo que tienen relación en la región interandina del Ecuador.

La unidad de estudio hace referencia a un crimen georreferenciado, el cual se encuentra de manera temporal delimitado, lo cual permite que se realice la examinación de patrones espaciotemporales que son tanto de interés operativo como para tomar acciones en la programación de estrategias de prevención, a nivel situacional y con el uso de recursos policiales.

Se establecieron procesos en esta investigación con la finalidad de asegurar la calidad de la información y su adecuación para la modelación estadística y el aprendizaje automático. Controles sistemáticos se implementaron con coherencia interna abarcando la validación de periodos en tiempos admisibles. Dada la capacidad de interpretación de la regresión logística multinomial, se elaboró un análisis detallado de los coeficientes analizados. Este análisis se detalló sobre el modelo de regresión final entrenado con el conjunto completo de datos tras implementar técnicas de sobremuestreo.

3. Resultados y discusión

Se llevó a cabo el análisis exploratorio que permitió caracterizar la distribución de las principales variables sociodemográficas de los incidentes registrados en la región interandina. Tal información ha sido trabajada en función de gráfico de barras que permiten una mejor visualización de los resultados, así como, facilidad para desarrollar los análisis, desagregando las diferentes variables involucradas en el estudio. En términos de la tipología delictiva, en el Gráfico I, el asesinato fue la modalidad con mayor frecuencia, seguido por el homicidio y en menor medida el femicidio; mientras que el sicariato constituyó un fenómeno residual.

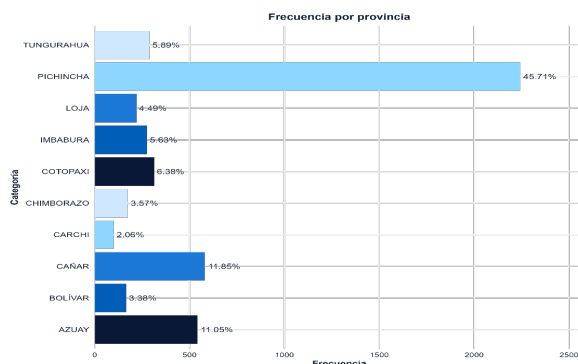


Fuente: Elaboración propia, 2026.

Gráfico I: Distribución de incidentes delictivos en la región interandina del Ecuador

El Gráfico II, presenta a Pichincha como el principal epicentro donde se genera la violencia, seguida por la provincia de Cañar y Azuay; mientras que existen otras provincias como Carchi y Bolívar que revelan

una realidad totalmente contraria, con una incidencia mucho menor en comparación al resto de provincias que pertenecen a la región interandina.

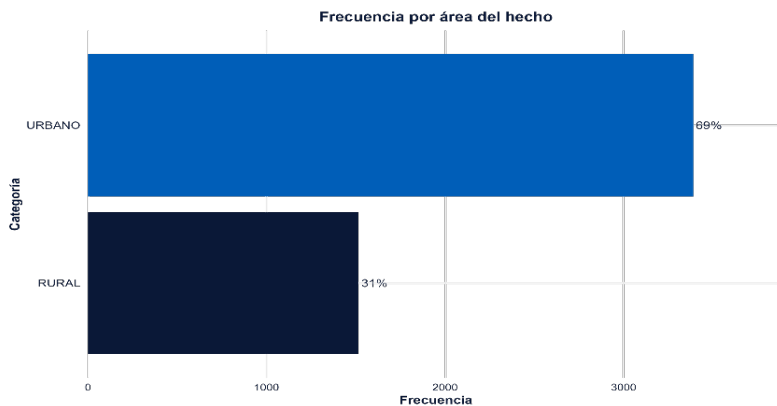


Fuente: Elaboración propia, 2026.

Gráfico II: Distribución de incidentes delictivos por provincia interandina

El Gráfico III, presenta una incidencia de alrededor del 70% con crímenes relacionados al entorno urbano, resaltando esta realidad con la alta densidad poblacional. Este hallazgo

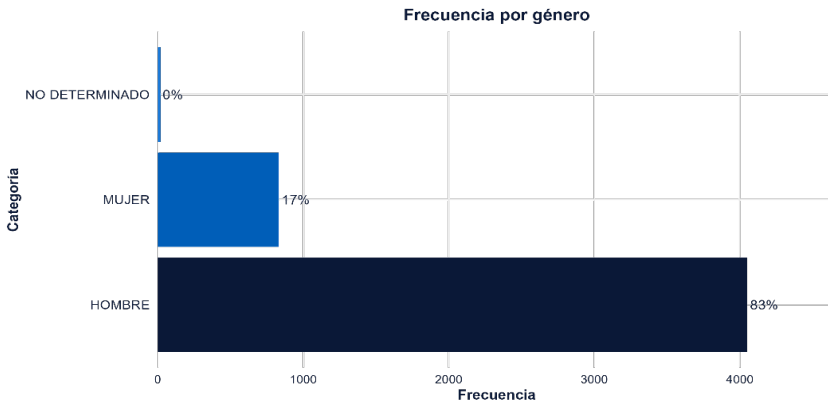
resulta muy coherente de acuerdo con la literatura que asocia a la violencia urbana en sitios donde existe una alta concentración de recursos y población en pequeñas áreas.



Fuente: Elaboración propia, 2026.

Gráfico III: Distribución de análisis territorial

En el Gráfico IV, se reflejó que la incidencia delictiva en la región interandina, la población masculina es la mayor víctima de representando más del 80% de los incidentes.

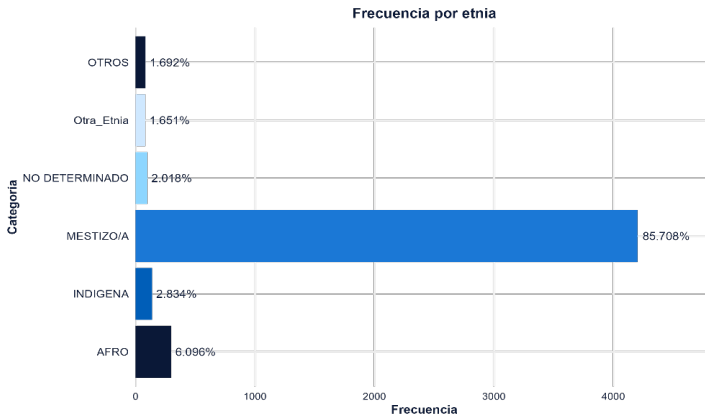


Fuente: Elaboración propia, 2026.

Gráfico IV: Distribución de la variable sexo

El Gráfico V, permite evidenciar que la mayoría de los incidentes ocurren en la población que pertenecen a la etnia mestiza, la cual representa alrededor del 90% de los registros, lo cual es consistente de acuerdo

con la composición demográfica de Ecuador. En cuanto a los afrodescendientes y la etnia indígena a pesar de que tienen menor representación, siguen constituyendo una parte significativa en la estructura social.

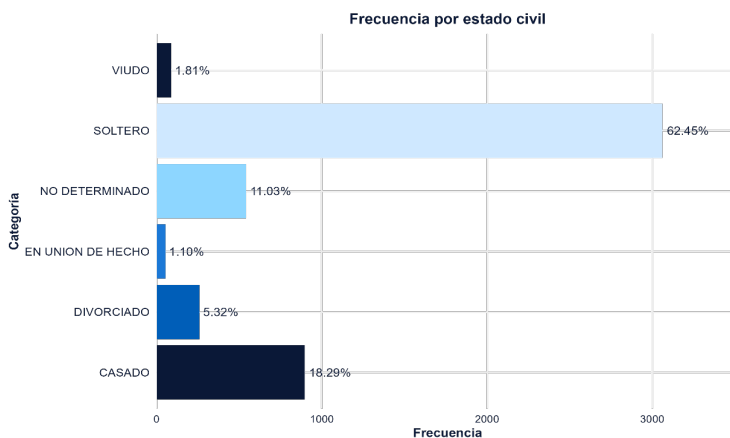


Fuente: Elaboración propia, 2026.

Gráfico V: Distribución étnica en la región interandina del Ecuador

En el Gráfico VI, se evidencia que las personas solteras representan el 60% de los casos sufridos, sugiriendo que al no existir alguna unión estable o al estar aislados en la

parte social, puede reflejarse en el aumento del riesgo para involucrarse en situaciones que resultan extremadamente violentas.

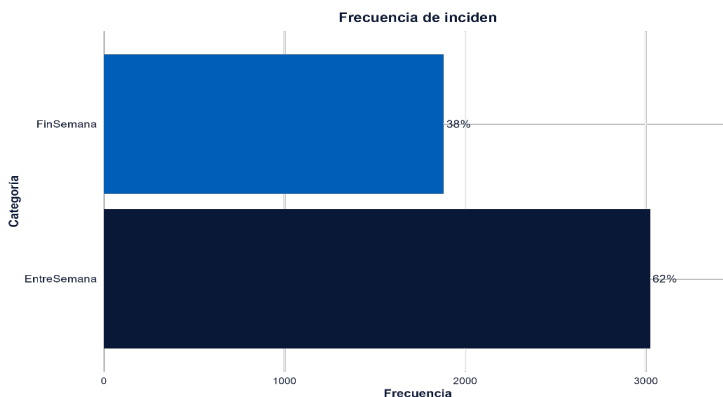


Fuente: Elaboración propia, 2026.

Gráfico VI: Distribución del estado civil

El Gráfico VII, permite observar que la distribución de los incidentes delictivos ocurrió en su mayoría durante el fin de semana, lo cual

representa el 62%, quedando un 38% restante que ocurrió en días laborables (entre semana).

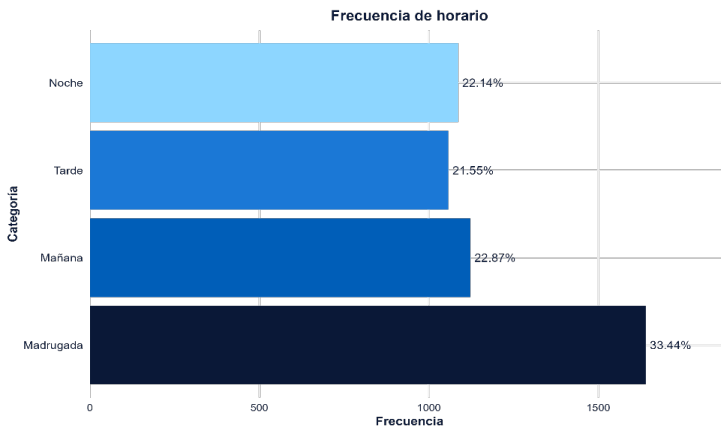


Fuente: Elaboración propia, 2026.

Gráfico VII: Distribución de si el incidente es entre fin de semana o fin de semana en la región interandina del Ecuador

En el Gráfico VIII, se evidencia la distribución de los incidentes delictivos en diferentes franjas horarias del día en las diversas provincias de la región interandina. Los resultados se distribuyen en 4 puntos,

el 21,3% de los casos fueron durante el amanecer; en la mañana se dieron el 22,7%; mientras que en la tarde se tiene un 24,4% y finalmente, durante la noche, los incidentes fueron del 23,6%.



Fuente: Elaboración propia, 2026.

Gráfico VIII: Distribución de registros en diversos horarios del día

Por otra parte, en la Tabla 1 se muestra que la “vía pública”, es la zona con mayor frecuencia de ocurrencia de crímenes en las provincias de Sierra, representando el 30,09% de los incidentes, seguido por “casas” con un 23%. “Otros lugares”, como vías públicas adicionales, suman un 11,03%. Los terrenos

baldíos también son importantes, con un 5,40% de los incidentes. Los lugares con la menor presencia de incidentes incluyen hospitales (1,73%), bosques y reservas ecológicas (1,61%), parques y plazas (1,47%), ríos (1,08%) y *night clubs* (1,06%).

Tabla 1
Frecuencia de tipos de lugares donde ocurrieron los crímenes

Distribución por tipo de lugar	Frecuencia	Porcentaje (%)
Vía publica	2.017	42,09
Casa/villa	1.128	23,00
Otro lugar	536	10,93
Terreno baldío	265	5,40
Mar	195	3,98
Centros de rehabilitación social	188	3,83
Quebradas	121	2,47
Finca	114	2,32

Cont... Tabla 1

Hospitales	85	1,73
Bosque y reservas ecológicas	79	1,61
Parques y plazas	72	1,47
Rio	53	1,08
Night clubs	52	1,06

Fuente: Elaboración propia, 2026

3.1. Estudio de interpretabilidad en la regresión logística multinomial

Para entender sobre la relación entre las variables objetivo y predictores, se entrenaron sin regularización a modelos de regresión logística multinomial utilizando las variables

significativas identificadas mediante pruebas Chi cuadrado χ^2 con un valor de alfa de 0.05. La Tabla 2, presenta distintas variables significativas, las cuales fueron identificadas mediante la regresión logística para entender acerca de la franja horaria; entre ellas se destacan:

Tabla 2
Significancia de las variables en la regresión logística multinomial en franjas horarias

Variable	Chi-cuadrado	Valor P	Significancia
Presunta motivación "Delincuencia Común"	327.504	1.11e-70	***
Lugar "Vía Pública"	241.524	4.46e-52	***
Lugar "Centros De Rehabilitación Social"	128.751	1.01e-27	***
Arma "Arma De Fuego"	110.727	7.65e-24	***
Arma "Otros"	53.750	1.27e-11	***
Lugar "Quebradas"	51.638	3.58e-11	***
Es "fin de semana"	48.882	1.38e-10	***
Área hecha "Urbano"	47.948	2.18e-10	***
Lugar "Mar"	47.743	2.41e-10	***
Lugar "Terreno Baldío"	35.701	8.66e-08	***
Tipo muerte "Homicidio"	35.280	1.06e-07	***
Arma "Constrictora"	32.398	4.31e-07	***
Lugar "Finca"	28.781	2.49e-06	***
Presunta motivación "Violencia Intrafamiliar"	27.802	4.00e-06	***
Sexo "Mujer"	27.063	5.71e-06	***
Lugar "Vía Pública"	26.715	6.75e-06	***
Arma "Arma Contundente"	25.015	1.53e-05	***
Etnia "Indígena"	21.319	9.04e-05	***
Tipo muerte "Femicidio"	19.102	2.60e-04	***
Estado civil "No Determinado"	14.263	2.57e-03	**
Estado civil "Soltero"	10.444	1.51e-02	*
Nacionalidad "No Determinado"	10.184	1.71e-02	*
Lugar "Otro lugar"	9.813	2.02e-02	*
Etnia "Mestizo/a"	9.095	2.81e-02	*
Nacionalidad "Ecuador"	8.135	4.33e-02	*

Fuente: Elaboración propia, 2026.

Entre las variables con mayor relevancia, destacan la presunta motivación de delincuencia común, mostrando un chi-cuadrado de 327.504 y un p-valor de 1.11e-70, lo cual indica una fuerte asociación con las franjas horarias, evidenciando que los crímenes relacionados con esta motivación existen en momentos específicos del día. De igual manera, los centros de rehabilitación social y la calle son otras variables que obtuvieron p-valores muy bajos, validando su importancia.

El uso de armas de fuego que son letales fue uno de los factores fundamentales en este análisis, alcanzando un valor de p correspondiente a $7.65e-24$, lo cual refleja una alta importancia en la identificación de las franjas horarias en las que ocurren la mayoría de los incidentes delictivos. En menor cantidad, el entorno marino, también revela una moderada importancia con la ocurrencia de delitos durante el transcurso del día.

Por último, otras variables como

el estado civil (soltero), la nacionalidad (Ecuador) y el uso de un arma contundente presentaron valores p demasiado bajos a 0.05, demostrando que están relacionadas significativamente, aunque con una capacidad de predicción de menor fuerza en comparación con las variables descritas.

La Tabla 3, permite analizar la variable que tiene mayor influencia en el modelo (Centros de Rehabilitación Social), generando una dirección altamente positiva, con coeficientes que son elevados, regulares y consistentes a lo largo de cada franja horaria. Las variables con relación a armas, tales como: Arma contundente, constrictora y otros, presentan efectos positivos consistentes, aunque con una reducida magnitud evaluada con otras variables como las de localización. El análisis muestra que variables relacionadas al lugar de ocurrencia son las más fuerte en el ámbito de la predicción, seguidas por las motivaciones del crimen y algunas características temporales.

Tabla 3
Variables Globales con franja horaria

Variable	Número de clases significativas	Coef. promedio	Coefmax	Dirección	Or promedio	Valor p	Importancia
Lugar "Centros de Rehabilitación Social"	3	2.528	2.939	Positiva	13.654	0.000	1.000
Presunta_motivacion_Delincuencia Común	3	1.400	1.772	Negativa	0.256	0.000	0.801
Lugar "Finca"	3	1.461	1.700	Positiva	4.453	0.000	0.789
Arma "Otros"	3	0.613	0.825	Positiva	1.870	0.000	0.640
Lugar "Mar"	2	0.962	1.678	Positiva	3.128	0.000	0.619
Arma "Arma Contundente"	3	0.424	0.521	Positiva	1.535	0.000	0.589
Es "fin de semana"	3	0.328	0.394	Negativa	0.723	0.000	0.567
Lugar "Vía Pública"	2	0.481	0.754	Negativa	0.736	0.000	0.462
Arma "Arma De Fuego"	2	0.405	0.685	Positiva	1.434	0.000	0.450
Arma "Constrictora"	2	0.429	0.632	Positiva	1.580	0.001	0.441

Fuente: Elaboración propia, 2026.

En la Tabla 4, se puede evidenciar el comportamiento que tiene cada una de las variables en la franja horaria. Durante

la madrugada, las variables de lugar que corresponden a Centros de Rehabilitación Social y Finca, son las más influyentes,

alcanzando un coeficiente promedio de 5.528 y 1.461 respectivamente, lo que muestra un aumento significativo en cuanto a la

probabilidad de crímenes se refiere; mientras que la delincuencia común muestra una reducción de probabilidad de hasta un 74%

Tabla 4
Comportamiento de variables en la franja horaria

Variable	Número de clases totales	Número de clases sig.	Coef. Promedio	Coef. sd	Coef. min	Coef. max	Consistencia	Dirección
lugar "Centros De Rehabilitacion Social"	3	3	2.528	0.535	1.924	2.939	Consistente	Positiva
Lugar "Finca"	3	3	1.461	0.324	1.092	1.700	Consistente	Positiva
Presunta motivación "Delincuencia Común"	3	3	-1.400	0.337	-1.772	-1.115	Consistente	Negativa
Arma "Otros"	3	3	0.613	0.193	0.446	0.825	Consistente	Positiva
arma "Arma Contundente"	3	3	0.424	0.117	0.294	0.521	Consistente	Positiva
Es "fin semana"	3	3	-0.328	0.096	-0.394	-0.217	Consistente	Negativa
Lugar "Mar"	3	2	0.962	0.764	0.158	1.678	Consistente	Positiva
Arma "Constrictora"	3	2	0.429	0.300	0.085	0.632	Consistente	Positiva
Lugar "Vía Pública"	3	2	-0.384	0.470	-0.754	0.145	Mixta	Negativa
Arma "Arma de Fuego"	3	2	0.303	0.424	-0.153	0.685	Mixta	Positiva
Estado civil "Soltero"	3	2	-0.268	0.108	-0.367	-0.153	Consistente	Negativa
Lugar "Terreno Baldío"	3	1	0.417	0.297	0.175	0.748	Consistente	Positiva
Lugar "Quebradas"	3	1	0.258	0.599	-0.157	0.945	Mixta	Positiva
Estado civil "No Determinado"	3	1	-0.172	0.223	-0.429	-0.027	Consistente	Negativa
Etnia "Mestizo/a"	3	1	-0.156	0.089	-0.258	-0.100	Consistente	Negativa

Fuente: Elaboración propia, 2026.

En la mañana, las armas como armas (otros) y arma contundente, y la ubicación (finca) continúan teniendo un impacto favorable en el comportamiento de variables en la franja horaria; mientras que la supuesta motivación (delincuencia común) sigue una tendencia de reducir la probabilidad del evento. En la tarde, el lugar (Mar) refleja un incremento en la probabilidad del crimen aproximado a 3; mientras que el lugar (vía pública) reflejó una dependencia negativa, disminuyendo la posibilidad del acontecimiento a investigar. Durante la noche, las variables relativas a armas constrictora y arma contundente están asociadas con un aumento en la probabilidad del suceso; mientras que supuesta motivación (delincuencia común) continúa siendo el principal indicador negativo en este estudio

dado todo el conjunto de variables estudiadas.

Como muestra la Tabla 5, en su gran mayoría, los predictores presentan efectos mixtos entre las provincias evaluadas, lo que representa una alta variación a nivel territorial en la probabilidad de ocurrencia. Entre las variables con más clases significativas como son lugar (Vía Pública), la etnia (Mestizo/a), la nacionalidad (Ecuador) y el tipo de muerte (Femicidio), revelaron que la influencia es bastante consistente en el modelo, aunque la dirección tiene cierto tipo de variaciones dependiendo de la provincia. Predictores como el área, "Urbana" y la etnia "Mestizo/a" tienden más hacia factores negativos; mientras que predictores como lugar, "Vía Pública" y tipo de muerte, "Homicidio", presentan una tendencia hacia factores positivos.

Tabla 5
Comportamiento de Variables en provincias

Variable	Número de clases totales	Número de clases sig.	Coef. Promedio	Coef. sd	Coef. min	Coef.max	Consistencia	Dirección
Etnia “Mestizo/a”	9	8	-0.607	0.818	-1.835	0.700	Mixta	Negativa
Lugar “Via Pública”	9	8	0.424	0.692	-0.354	1.451	Mixta	Positiva
Tipo “Muerte Femicidio”	9	8	0.368	0.795	-1.497	1.313	Mixta	Positiva
Nacionalidad “Ecuador”	9	8	-0.122	0.633	-1.396	0.725	Mixta	Negativa
Area “Hecho Urbano”	9	8	-0.109	0.858	-1.044	1.769	Mixta	Negativa
Lugar “Terreno Baldío”	9	8	-0.090	0.712	-1.305	0.850	Mixta	Negativa
Lugar “Casa/Villa”	9	7	0.295	0.496	-0.432	0.995	Mixta	Positiva
Lugar “Via Pública”	9	7	0.229	0.449	-0.432	0.818	Mixta	Positiva
Tipo “Muerte Homicidio”	9	7	0.208	0.888	-1.033	1.802	Mixta	Positiva
Estado civil “Soltero”	9	6	-0.283	0.208	-0.477	0.005	Mixta	Negativa
Lugar “Otro Lugar”	9	6	0.154	0.637	-1.239	0.771	Mixta	Positiva

Fuente: Elaboración propia, 2026.

La Tabla 6, evidenció que tanto factores demográficos como espaciales y situacionales tienen un peso significativo en la explicación provincial, configurando un patrón donde

algunas variables independientes actúan como factores de reducción del riesgo y otros como factores de incremento.

Tabla 6
Variables Globales en las provincias

Variable	Número de clases significativas	Coef. promedio	Coef. Max	Dirección	Or. promedio	Valor p	Importancia
Etnia “Mestizo/a”	8	0.856	1.835	Negativa	0.735	0	0.976
Área hecha “Urbano”	8	0.641	1.769	Negativa	1.360	0	0.959
Tipo muerte “Homicidio”	7	0.700	1.802	Positiva	1.767	0	0.905
Tipo muerte” Femicidio”	8	0.701	1.497	Positiva	1.773	0	0.888
Lugar “Via Pública”	8	0.650	1.451	Positiva	1.888	0	0.876
Nacionalidad “Ecuador”	8	0.489	1.396	Negativa	1.035	0	0.862
Lugar “Terreno Baldío”	8	0.598	1.305	Negativa	1.116	0	0.839
Lugar “Casa/Villa”	7	0.475	0.995	Positiva	1.492	0	0.696

Cont.... Tabla 6

Lugar "Vía Pública"	7	0.413	0.818	Positiva	1.372	0	0.650
---------------------	---	-------	-------	----------	-------	---	-------

Fuente: Elaboración propia, 2026.

Las variables globales más influyentes mantienen efectos consistentes en la mayoría de las provincias; área hecho "Urbano" y Etnia "Mestizo/a" destacan por presentar efectos predominantemente negativos y su alta importancia, mientras que, tipo muerte "Homicidio", discapacidad, lugar "Vía Pública" y tipo muerte "Femicidio"

identificaron efectos positivos, aumentando la probabilidad relativa según el *odds* ratio promedio.

Para determinar si las diferencias observadas entre modelos son estadísticamente significativas, se aplicaron *test t* pareados sobre las evaluaciones de validación cruzada, tal como se puede observar en la Tabla 7.

Tabla 7
La prueba *t* pareado sobre la validación cruzada para la predicción de franjas horarias

Métrica	Red Neuronal	Reg. Logística	Diferencia	t	p-valor	Sig.
Balanced Accuracy	0.5649	0.5784	-0.0134	-3.2684	0.0056	**
Kappa de Cohen	0.1333	0.1637	-0.0305	-3.8271	0.0018	**
F1-Score	0.3440	0.3645	-0.0204	-3.3061	0.0052	**
AUC	0.6052	0.6397	-0.0344	-7.1278	0.0000	***

Fuente: Elaboración propia, 2026.

Así mismo, el estadístico Kappa de Cohen (ver Tabla 8) demostró una diferencia estadísticamente significativa de -0.0305 , con un valor *p* de 0.0018, revelando una capacidad de la regresión logística para identificar de igual manera las relaciones correctas entre las clases. En *F1-Score*, se observó que la regresión logística obtuvo una leve ventaja con

un valor de 0.3645 en comparación a 0.3440 de las redes neuronales, lo que resulta que tiene un mejor balance entre *recall* y precisión. Finalmente, el AUC, de la regresión logística mostró una variación de -0.0344 y un valor *p* de 0.0000, demostrando una mejor capacidad para clasificar.

Tabla 8
La prueba *t* pareado sobre la validación cruzada para la predicción de provincias

Métrica	Red Neuronal	Reg. Logística	Diferencia	t	p-valor	Sig.
Balanced Accuracy	0.5860	0.6141	-0.0281	-6.6899	0.0000	***
Kappa de Cohen	0.1984	0.2714	-0.0730	-11.6151	0.0000	***
F1-Score	0.2193	0.2626	-0.0434	-7.6786	0.0000	***
AUC	0.6664	0.7288	-0.0623	-15.1084	0.0000	***

Fuente: Elaboración propia, 2026.

Los resultados de la prueba *t* pareada muestran que la regresión logística multinomial supera significativamente a la red neuronal en todas las métricas evaluadas. En *Balanced Accuracy*, la regresión logística alcanzó un valor de 0.6141; mientras que la red neuronal obtuvo 0.5860, con una diferencia de -0.0281 y un *p*-valor de 0.0000, indicando que la diferencia es altamente significativa. Este resultado sugiere que la regresión logística tiene una mejor capacidad para manejar el desbalanceo de clases y clasificar de manera más precisa en el contexto de franjas horarias y provincias.

En cuanto al Kappa de Cohen, la regresión logística mostró un valor de 0.2714; mientras que la red neuronal obtuvo 0.1984, con una diferencia de -0.0730 y un *p*-valor de 0.0000, indicando un acuerdo sustancial en las predicciones de la regresión logística en comparación con un acuerdo moderado en la red neuronal. En *F1-Score*, la regresión logística superó a la red neuronal con un valor de 0.2626, frente a 0.2193, con una diferencia de -0.0434 y un *p*-valor de 0.0000, lo que demuestra que la regresión logística mantiene un mejor equilibrio entre precisión y *recall*.

Finalmente, en AUC, la regresión logística obtuvo un valor de 0.7288; mientras que la red neuronal alcanzó 0.6664, con una diferencia de -0.0623 y un *p*-valor de 0.0000, lo que sugiere que la regresión logística tiene una mejor capacidad discriminativa entre las clases

Conclusiones

Los hallazgos revelan que tanto las redes neuronales como la regresión logística multinomial han tenido un rendimiento levemente moderado, considerando que la precisión a nivel general ha sido más baja de lo que se esperaba para tareas de predicción enfocadas en el crimen. Ambos modelos muestran limitantes, que afectan directamente a la precisión y a la generalización de las predicciones, que pueden ser medianamente útiles cuando se trata de realizar una predicción

a nivel de franja horaria y provincias.

El modelo de regresión logística multinomial logró una predicción balanceada de 0.5784, destacando en la predicción de franjas horarias. Este modelo, al ser lineal, reflejó mayor solidez y estabilidad cuando las relaciones entre las variables eran en su mayoría simples. Es por ello, que, uno de los elementos a considerar de la regresión logística es su gran capacidad de examinar coeficientes permitiendo reconocer qué variables tienen mayor impacto en la predicción de la franja horaria. Variables como tipo de muerte y el lugar del hecho, resultaron ser los más significativos estadísticamente en la predicción de crímenes durante diversas horas del día, reforzando la efectividad del modelo, ofreciendo explicaciones precisas sobre la conexión entre los resultados y las características.

Contrariamente, la red neuronal evidenció una mejor capacidad discriminativa para diferenciar en la predicción de provincias de la región interandina, con un AUC de 0.7288, subrayando una habilidad potente para manejar relaciones no lineales y patrones complejos. No obstante, la red neuronal pese a tener un rendimiento superior en términos de AUC, su interpretabilidad es muy limitada debido a la complejidad de dicho modelo, dificultando la comprensión directa de cómo las variables individuales influyen en las predicciones. A pesar de esta notable desventaja, la red neuronal destaca en la captura de esquemas geoespaciales complejos, lo cual es difícil de modelar con técnicas lineales.

No hay relación estadísticamente significativa entre algunas variables, como el estado civil y tipo de arma utilizada; algo que afecta notablemente la capacidad de predicción de los dos modelos evaluados, indicando que algunas de las variables sociodemográficas carecen de suficiente información. Este descubrimiento destaca la importancia de perfeccionar la selección de características y la comprensión de los modelos, para adquirir resultados eficientes y así maximizar la precisión y la interpretabilidad de los modelos.

A pesar de que ambas técnicas

emplearon *upsampling* para abordar el desbalanceo de clases, los resultados siguieron siendo subóptimos, indicando la necesidad de explorar modelos adicionales como los modelos *XGboost* que puedan combinar lo mejor de los enfoques lineales y no lineales, y así optimizar la precisión como la capacidad de interpretación. En contraste ambos modelos requieren de mejoras importantes en la selección de características, la comprensión de los resultados, y el manejo del desbalanceo de clases para potenciar su aplicabilidad y eficiencia práctica en cuestiones de predicción criminal

Referencias bibliográficas

- Bridle, J. S. (1990). Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In F. F. Soulié y J. Héroult (Eds.), *Neurocomputing: Algorithms, architectures and applications*. NATO ASI Series (Vol 68, pp. 227-236). Springer. https://doi.org/10.1007/978-3-642-76153-9_28
- Chainey, S., y Ratcliffe, J. (2013). *GIS and Crime Mapping*. John Wiley & Sons.
- Chávez-Cadena, M. I., y Abril-Donoso, M. E. (2023). Predicción de homicidios desde la regresión logística en el distrito metropolitano de Quito, Ecuador. *Koinonia*, VIII(16), 238-255. <https://doi.org/10.35381/r.k.v8i16.2551>
- Corbett-Davies, S., Gaebler, J. D., Nilforoshan, H., Shroff, R., y Goel, S. (2023). The Measure and Mismeasure of Fairness. *Journal of Machine Learning Research*, 24, 312. <https://doi.org/10.48550/arXiv.1808.00023>
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2, 303-314. https://web.njit.edu/~usman/courses/cs675_fall18/10.1.1.441.7873.pdf
- Friedman, J. H., Hastie, T., y Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1-22. <https://doi.org/10.18637/jss.v033.i01>
- Hastie, T., Tibshirani, R., y Friedman, J. (2009). *The Elements of Statistical Learning: Data mining, inference, and prediction*. Springer.
- He, H., y García, E. A. (2009). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263-1284. <https://doi.org/10.1109/TKDE.2008.239>
- Hoerl, A. E., y Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1), 55-67. <https://homepages.math.uic.edu/~lreyzin/papers/ridge.pdf>
- Lum, K., e Isaac, W. (2016). To Predict and Serve? *Significance*, 13(5), 14-19. <https://doi.org/10.1111/j.1740-9713.2016.00960.x>
- Luna, K. A., Melean, R. A., Andrade, C. F., y Orellana, C. P. (2025). Redes neuronales en la valoración crediticia bajo incertidumbre en pequeñas y medianas empresas de Ecuador. *Revista de Ciencias Sociales (Ve)*, XXXI(3), 313-333. <https://doi.org/10.31876/rcs.v31i3.44286>
- Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., y Tita, G. E. (2011). Self-Exciting Point Process Modeling of Crime. *Journal of the American Statistical Association*, 106(493), 100-108. <https://doi.org/10.1198/jasa.2011.ap09546>
- Nasrabadi, N. M. (2006). Pattern recognition and machine learning. *Journal of Electronic Imaging*, 16(4), 140-155. <https://doi.org/10.1117/1.2819119>

- Perry, W. L., McInnis, B., Price, C. C., Smith, S., y Hollywood, J. S. (2013). *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. Rand.
- Ratcliffe, J. H. (2003). Intelligence-led Policing. *Crime and Criminal Justice*, (248), 1-6. <https://www.aic.gov.au/sites/default/files/2020-05/tandi248.pdf>
- Richardson, R., Schultz, J., y Crawford, K. (2019). Datos erróneos, predicciones erróneas: cómo las violaciones de los derechos civiles afectan a los datos policiales, los sistemas de vigilancia predictiva y la justicia. *Nyul*, 94(192), 192-232. <https://ssrn.com/abstract=3333423>
- Rudin, C. (2019). Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*, 1, 206-215. <https://doi.org/10.1038/s42256-019-0048-x>
- Rumelhart, D. E., Hinton, G. E., y Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533-536. <https://doi.org/10.1038/323533a0>
- Sherman, L. W., Gartin, P. R., y Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology*, 27(1), 27-56. <https://doi.org/10.1111/j.1745-9125.1989.tb00862.x>