UNAH
UNIVERSIDAD NACIONAL
AUTÓNOMA DE HONDURAS

Boletín Divulgativo

Carrera de Matemáticas

- Elementos finitos
- Análisis espacial
- Técnicas de regularización
- Redes neuronales
- Aproximación asintótica

OCTUBRE-2025

Presentación

Este documento fue planificado, supervisado, asesorado y desarrollado por el profesor David Méndez de la Carrera de Matemáticas de la UNAH, presenta artículos divulgativos y de investigación desarrollados por estudiantes avanzados de la Carrera de Matemáticas durante el año 2025. Se abarca una temática variada: elementos finitos, análisis espacial, técnicas de regularización, redes neuronales y aproximaciones asintóticas; en algunos de los trabajos se desarrolló una revisión bibliográfica de trabajos pertinentes y se resumió según lo comprendido por cada autor, en otros casos, se realizó experimentación original y se obtuvo resultados interesantes.

El objetivo principal de desarrollar este documento es que a futuro, en base a la experiencia obtenida y después de tener varias experiencias similares, se transforme en una revista científica de Matemáticas, cuestión que requiere de mucho trabajo por parte del equipo de profesores investigadores del programa y otros colaboradores externos; además de ser una muestra de que en la Carrera de Matemáticas se está desarrollando en los estudiantes un espíritu investigador. Cabe destacar que documentos similares se han desarrollado desde el programa de Maestría en Matemáticas de la UNAH.

Todas las revisiones bibliográficas y temas aquí presentados se encasillan dentro de las líneas de investigación de la UNAH, entre los temas prioritarios abarcados se encuentran: ciencia, cambio climático y vulnerabilidad, y productividad. Esto evidencia que la Carrera de Matemáticas está interesada en colaborar con las prioridades investigativas de la universidad y mantiene un compromiso con vincularse con la sociedad.

Se espera que a futuro se sigan desarrollando publicaciones similares, que a pesar de ser trabajos de divulgación, forman un inicio en la actividad investigativa de la Carrera de Matemáticas.

Octubre del año 2025, Ciudad Universitaria

Tegucigalpa, M.D.C., Honduras

® Carrera de Matemáticas - UNAH
Edificio F1, Segundo Piso, Ciudad Universitaria
Tegucigalpa, M.D.C. Honduras.
https://matematica.unah.edu.hn/escuela/carreras/licenciaturas/
carrera.matematica@unah.edu.hn
Tel. 2216-3055

Contenido

APLICACIÓN DEL MÉTODO DE ELEMENTO FINITO A LA ECUACIÓN DEL CALOR EN UN MEDIO ANISOTRÓPICO

LUIS HUMBERTO MARADIAGA

Resumen. A comienzos del siglo XIX, Joseph Fourier formuló la ecuación del calor en su obra *Théorie analytique de la chaleur*, estableciendo cómo varía la temperatura en un cuerpo a lo largo del tiempo. En este estudio, analizamos la versión estacionaria de dicha ecuación en un medio anisotrópico y no homogéneo en dos dimensiones con condiciones de frontera de tipo Dirichlet. El interés principal recae en materiales cuya conductividad térmica depende de la posición, tal como sucede en ciertas aleaciones, fibras compuestas o estructuras con variaciones internas. Para tratar el problema, se implementa el método de elementos finitos, aprovechando su capacidad para manejar mallas irregulares y coeficientes variables. Los resultados permitirán caracterizar la distribución de temperatura en regiones donde las propiedades térmicas cambian de forma significativa.

Abstract. At the beginning of the 19th century, Joseph Fourier formulated the heat equation in his work *Théorie analytique de la chaleur*, establishing how the temperature of a body varies over time. In this study, we analyze the stationary version of this equation in a two-dimensional anisotropic and inhomogeneous medium with Dirichlet-type boundary conditions. The main interest lies in materials whose thermal conductivity depends on position, such as certain alloys, composite fibers, or structures with internal variations. To address the problem, the finite element method is implemented, taking advantage of its ability to handle irregular meshes and variable coefficients. The results will allow to characterize the temperature distribution in regions where thermal properties change significantly.

1. Introducción

La ecuación del calor es una importante ecuación diferencial en derivadas parciales de tipo parabólicas que describe la distribución del calor o variaciones de la temperatura. Es una herramienta fundamental en la ciencia y la ingeniería, y desempeña un papel crucial en numerosos campos, como la termodinámica, la física, la ingeniería termica y muchas disciplinas más.

En este proyecto de investigación se estudiará la ecuación del calor mediante un medio anisotrópico no homogéneo en dos dimensiones con condición de Dirichlet en estado estacionario. Un medio anisotrópico se caracteriza por tener propiedades térmicas y geométricas donde se propagan de forma distinta en todas las direcciones en el espacio. La conductividad térmica va a depender de su posición, esto quiere decir que el material conducirá el calor de manera diferente en diferentes regiones

Fecha: Agosto 08, 2025.

Palabras y frases clave. Ecuación del calor, Elemento Finito, medio anisotrópico, no homogéneo, Espacios de Sobolev, Teorema de Divergencia, Identidad de Green, Soporte, Ortotrópico.

del espacio en el que estemos trabajando. Un ejemplo de un material anisotrópico puede ser la madera, cristales como el grafito y el cuarzo, fibra de vidrio y fibra de carbono, láminas de metal, hueso, etc [1].

La estrategia que utilizaremos para resolver este problema consiste en aplicar el método de elementos finitos, una técnica numérica ampliamente utilizada para discretizar dominios continuos y aproximar soluciones de ecuaciones diferenciales parciales. La elección del método de elementos finitos se justifica por su capacidad para manejar problemas con geometrías complejas y propiedades variables en el espacio. Además, es una técnica ampliamente aceptada en el campo de la modelización y simulación de fenómenos físicos, y ofrece resultados precisos y confiables. Al utilizar este enfoque, podremos obtener una descripción cuantitativa detallada de la distribución de temperatura en un medio anisotrópico no homogéneo y comprender mejor los mecanismos de transferencia de calor involucrados.

Este tipo de modelación tiene aplicaciones prácticas en el diseño y mejora de materiales industriales, sistemas de climatización, eficiencia energética en edificaciones y análisis térmico de componentes estructurales [3]. Por ejemplo en el contexto de Honduras, donde se requiere impulsar el desarrollo tecnológico, optimizar el uso de la energía y fortalecer la infraestructura, este tipo de estudios contribuye al desarrollo sostenible y se alinea con los temas prioritarios de investigación de la Universidad Nacional Autónoma de Honduras (UNAH): infraestructura y desarrollo territorial, productividad y competitividad, ciencia, cambio climático y vulnerabilidad, y desarrollo energético.

2. Antecedentes

La ecuación que describe la conducción del calor en los sólidos ocupa una posición única en la física matemática moderna. Además de ser fundamental para el análisis de problemas relacionados con la transferencia de calor en sistemas físicos, la estructura conceptual-matemática de la ecuación de conducción del calor (también conocida como ecuación de difusión del calor) ha inspirado la formulación matemática de muchos otros procesos físicos en términos de difusión. En consecuencia, las matemáticas de la difusión han facilitado la transferencia de conocimiento relacionado con la resolución de problemas entre diversas disciplinas aparentemente inconexas. El proceso transitorio de conducción del calor, descrito mediante una ecuación diferencial parcial, fue formulado por primera vez por Jean-Baptiste-Joseph Fourier (1768-1830) y presentado como manuscrito en el Instituto de Francia en 1807. En aquel entonces, la termodinámica, la teoría del potencial y la teoría de ecuaciones diferenciales eran comunes. Las etapas iniciales de su formulación. Combinando notables dotes en matemáticas puras y conocimientos sobre física observacional, Fourier abrió nuevas áreas de investigación en física matemática con su obra maestra de 1807, "Théorie de la Propagation de la Chaleur dans les Solides"[4].

El enfoque de Fourier no fue aceptado de inmediato: transcurrieron unos quince años hasta que su obra clásica, se publicó y se difundió ampliamente. Una vez disponible, su método se empleó no solo en problemas de transferencia de calor, sino también en ámbitos como la electricidad, la difusión química, el flujo en medios porosos, la genética y la economía, e impulsó mucha investigación en teoría

de ecuaciones diferenciales. Hoy, casi dos siglos después, la ecuación de conducción de calor sigue siendo la base para analizar numerosos sistemas físicos, biológicos y sociales [4].

Para entender este desarrollo, conviene repasar avances previos del siglo XVIII. Gabriel Daniel Fahrenheit (1686–1736) perfeccionó en 1714 el termómetro de mercurio y, en 1724, estableció la escala que lleva su nombre, con 32°F para la fusión del hielo y 212°F para la ebullición del agua. Más adelante, Joseph Black (1728–1799), figura clave en la química cuantitativa, observó en Glasgow que al derretirse el hielo absorbía calor sin cambiar de temperatura, introduciendo así el concepto de "calor latente". En 1779, Adair Crawford (1748–1795) publicó sus Experimentos y observaciones del calor animal y la inflamación de cuerpos combustibles, donde propuso el papel del oxígeno en la generación de calor en la respiración y presentó un procedimiento de mezclas para medir calor específico. Finalmente, Jean Baptiste Biot (1774–1862) investigó la conducción térmica en una barra delgada calentada por un extremo, observando que el calor se desplazaba longitudinalmente pero también se perdía transversalmente al ambiente. Este conjunto de aportes preparó el terreno para los trabajos de Fourier y su revolucionaria formulación de la ecuación del calor [4, 5].

Tras los desarrollos iniciales de Fourier, Siméon Denis Poisson (1781–1840) profundizó en la resolución de problemas de difusión térmica aplicando métodos de análisis potencial: introdujo la ecuación de Poisson como extensión de la ecuación de Laplace, lo que permitió estudiar fuentes internas de "calor" o carga y entender mejor las condiciones de contorno. Carl Gustav Jakob Jacobi (1804–1851) y Peter Gustav Lejeune Dirichlet (1805–1859) establecieron fundamentos rigurosos para el análisis de Fourier, formalizando las condiciones de frontera (por ejemplo, condiciones de Dirichlet y Neumann) en problemas con la ecuación del calor y desarrollando técnicas de series y transformadas para asegurar convergencia y unicidad de la solución. William Thomson (Lord Kelvin) (1824–1907) aplicó la ecuación del calor al estudio de la conducción térmica en la corteza terrestre, analizando el gradiente geotérmico y vinculándolo al segundo principio de la termodinámica y al concepto de disipación de energía térmica. James Clerk Maxwell (1831–1879), aunque más conocido por su teoría electromagnética, estableció analogías entre la difusión de calor y el transporte de partículas, sentando bases para el enfoque estadístico de procesos de difusión en gases. Estos aportes ampliaron y rigieron el entendimiento de la ecuación del calor, consolidándola como prototipo en el análisis de ecuaciones en derivadas parciales y en la modelación de fenómenos de conducción y difusión [5].

La ecuación del calor continuó profundizándose en el siglo XX y en la actualidad investigadores como Eugenio Elia Levi (1883–1917) desarrollaron teorías de existencia y unicidad para ecuaciones parabólicas; Juliusz Schauder (1899–1943) estableció estimados fundamentales para la regularidad de soluciones. Más adelante, Olga Ladyzhenskaya (1922–2004) realizó aportes esenciales al estudio de la estabilidad y comportamiento de soluciones en ecuaciones parabólicas y en mecánica de fluidos, mientras que Jacques-Louis Lions (1928–2001) impulsó el uso de métodos funcionales y formulaciones variacionales en el análisis y control de sistemas gobernados

por ecuaciones en derivadas parciales, en tiempos recientes, se han explorado extensiones a modelos con fuentes no lineales, difusión en medios heterogéneos y control óptimo del proceso. Gracias a estas contribuciones contemporáneas, la ecuación del calor sigue siendo un modelo central en el análisis de ecuaciones en derivadas parciales, en física matemática y en múltiples aplicaciones interdisciplinarias.

3. Introducción teórica

En esta sección la mayoría de los conceptos vienen de [1] y [2].

3.1. Preliminares. Antes de abordar la formulación débil y la aplicación del método de elementos finitos, es necesario repasar algunos resultados fundamentales del análisis funcional y del cálculo vectorial que se utilizarán en el desarrollo teórico.

El teorema de la divergencia, también conocido como teorema de Gauss, establece una relación entre la integral de la divergencia de un campo vectorial sobre una región y el flujo del campo a través de la frontera de dicha región. Sea $\Omega \subset \mathbb{R}^2$ un dominio abierto y acotado con frontera suave $\partial\Omega$, y sea $\mathbf{F} = (F_1, F_2)$ un campo vectorial suficientemente suave definido sobre Ω. Este resultado es fundamental para derivar la identidad de Green y para justificar el paso de una formulación fuerte a una formulación débil de problemas en derivadas parciales. Entonces, se cumple:

$$\int_\Omega \nabla \cdot \mathbf{F}\, dx = \int_{\partial\Omega} \mathbf{F} \cdot \mathbf{n}\, ds, \tag{3.1}$$

donde $\mathbf{n}$ es el vector normal exterior unitario a la frontera $\partial\Omega$ y ds denota el elemento de longitud sobre la frontera.

Esta integral tambien puede ser escrita como una integral doble y también puede denominarse integral de área.

$$\int_\Omega \nabla \cdot \mathbf{F} = \iint_\Omega \nabla \cdot \mathbf{F}(x, y)\, dA,$$

La anterior es una integral sobre la frontera de Ω, la cual es una curva, y por tanto puede referirse como una integral de línea. Para que el teorema de la divergencia se cumpla, el campo vectorial $\mathbf{F}$ debe ser lo suficientemente suave.

La identidad de Green es una consecuencia directa del teorema de la divergencia y permite relacionar integrales que involucran funciones y sus derivadas. Si $u, v \in C^1(\overline{\Omega})$, se tiene:

$$-\int_\Omega v\Delta u\, dx = \int_\Omega \nabla u \cdot \nabla v\, dx - \int_{\partial\Omega} v\frac{\partial u}{\partial n}\, ds, \tag{3.2}$$

donde $\frac{\partial u}{\partial n} = \nabla u \cdot \mathbf{n}$ es la derivada normal de u en la frontera. Esta identidad será clave en el desarrollo de la formulación variacional del problema, ya que permite

reducir el orden de las derivadas presentes en la ecuación diferencial, lo cual es esencial para trabajar en espacios funcionales más generales.

Definición 3.1. Si u es una función, su soporte es la clausura del conjunto en el que u es distinto de cero:

$$\operatorname{supp}(u) = \overline{\{(x,y) \in \mathbb{R}^2 : u(x,y) \neq 0\}} = \overline{\{x \in \mathbb{R} : u(x) \neq 0\}}$$

Si u se define en Ω y $\operatorname{supp}(u)$ es un subconjunto compacto (es decir, un subconjunto cerrado y acotado de Ω), entonces se dice que u tiene soporte compacto en Ω.

Definición 3.2. El espacio $C^k(\overline{\Omega})$ se define como el conjunto de todas las funciones u con valores reales definidas en $\overline{\Omega}$ con la propiedad de que u y sus dericadas parciales hasta el orden k son todas continuas en $\overline{\Omega}$.

Definición 3.3. El espacio C_0^∞ se define como el conjunto de todas las funciones que son infinitamente diferenciables en Ω y tienen soporte compacto en Ω.

La ecuación del calor en estado estacionario para un medio anisotrópico y no homogéneo se expresa como:

$$-\nabla \cdot (\mathbf{K}(x,y)\nabla u(x,y)) = f(x,y), \quad \text{en } \Omega \subset \mathbb{R}^2, \tag{3.3}$$

donde:

- $u(x,y)$ representa la temperatura,
- $\mathbf{K}(x,y)$ es el tensor de conductividad térmica simétrico, positivo definido, que varía espacialmente,
- $f(x,y)$ representa una fuente térmica interna,
- $\Omega \subset \mathbb{R}^2$ es el dominio bidimensional.

La anisotropía implica que el calor no se propaga de forma uniforme en todas las direcciones. En medios no homogéneos, la conductividad térmica cambia en función de la posición, lo que complica la solución analítica del problema.

En este trabajo se considerarán únicamente condiciones de Dirichlet:

$$u(x,y) = g(x,y), \quad \text{en } \partial\Omega, \tag{3.4}$$

donde $g(x,y)$ es una función dada en la frontera del dominio.

3.2. Modelo. Considere el siguiente problema de valor en la frontera

$$\text{PVF} : \begin{cases} -\nabla \cdot \mathbf{q} = f & \text{en } \Omega, \\ u(x,y) = g & \text{en } \partial\Omega. \end{cases} \tag{3.5}$$

Este modelo posibilita la comprensión de los procesos de transferencia de calor en medios caracterizados por variaciones en sus propiedades térmicas. Al comprender cómo la temperatura se distribuye en estos medios anisotrópicos no homogéneos, podemos analizar la influencia de las variaciones espaciales en la conductividad térmica.

La ecuación de calor se obtiene a partir de la ley de Fourier y del principio de conservación de la energía. Según la ley de Fourier, la razón de cambio del flujo de energía calórica por unidad de área a través de una superficie es proporcional al gradiente de temperatura negativo en la superficie:

$$q = \kappa \cdot \nabla u. \tag{3.6}$$

Entonces nos dice que $\mathbf{q}$ es el vector de calor y f es una función que depende (x, y) que simula la entrada del calor desde una fuente externa. En un medio anisotrópico, el flujo de calor $\mathbf{q}$ no es paralelo con el gradiente de temperatura, es decir κ no es coeficiente escalar, sino que un tensor, en nuestro caso que $\Omega \subset \mathbb{R}^2$, el tensor de conductividad térmica $\mathbf{K}$ se escribe como:

$$\mathbf{K}(x, y) = \begin{bmatrix} k_{11}(x, y) & k_{12}(x, y) \\ k_{21}(x, y) & k_{22}(x, y) \end{bmatrix}, \tag{3.7}$$

donde κ es simétrico.

En el caso especial de los materiales ortotrópicos, el tensor de conductividad κ asume una forma diagonal

$$\mathbf{K}(x, y) = \begin{bmatrix} k_{11}(x, y) & 0 \\ 0 & k_{22}(x, y) \end{bmatrix} \tag{3.8}$$

sustituyendo $\mathbf{q}$ en (3.5) nuetro modelo quedaría de la forma:

$$\text{PVF} : \begin{cases} -\nabla \cdot (\kappa \cdot \nabla u) = f & \text{en } \Omega, \\ u(x, y) = g & \text{en } \partial\Omega. \end{cases} \tag{3.9}$$

Definición 3.4. Los siguientes espacios funcionales son fundamentales en análisis funcional y métodos variacionales:

Espacio de funciones cuadrado-integrable:

$$L^2(\Omega) = \left\{ v : \Omega \to \mathbb{R} \;\middle|\; \int_\Omega v^2 < \infty \right\}.$$

Espacio de Sobolev:

$$H^1(\Omega) = \left\{ v \in L^2(\Omega) \;\middle|\; \frac{\partial v}{\partial x},\ \frac{\partial v}{\partial y} \in L^2(\Omega) \right\}.$$

Espacio de Sobolev con condiciones de frontera de Dirichlet:

$$H_0^1(\Omega) = \left\{ v \in H^1(\Omega) \;\middle|\; v = 0 \text{ en } \partial\Omega \right\}.$$

3.3. Método Numérico. La formulación débil para este problema no homegéneo g debe de satisfacer algunas condiciones de regularidad. Supongamos que existe $G \in H^1(\Omega)$ tal que $G = g$ en $\partial\Omega$ consideremos las funciones de prueba en el espacio $H_0^1(\Omega)$, sin embargo, dada que la solución deseada u no satisface las condiciones de Dirichlet homogeneas no puede ser $u \in H_0^1(\Omega)$, en cambio la función $w = u - G$ es cero en $\partial\Omega$ y por lo tanto la solución tiene la forma $u = w + G$, $w \in H_0^1(\Omega)$.

La formulación débil es:
Encontrar un $u = w + G$, $w \in H_0^1(\Omega)$ tal que

$$\int_\Omega \mathbf{K}(\nabla w) \cdot \nabla v = \int_\Omega fv - \int_\Omega \mathbf{K}\nabla G \cdot \nabla v \quad \forall v \in H_0^1(\Omega) \tag{3.10}$$

$G \in H^1(\Omega)$ tal que $G = g$ en $\partial\Omega$.

3.4. MEF. Este método busca una solución aproximada del problema débil en un subespacio de dimensión finita $V_h \subset H_0^1(\Omega)$. Se construyen funciones de base ϕ_i, típicamente funciones polinomiales por partes, localmente soportadas.

La solución aproximada se expresa como:

$$u_h(x, y) = \sum_{j=1}^{N} U_j \phi_j(x, y), \tag{3.11}$$

donde $\{\phi_j\}$ son funciones de forma y $\{U_j\}$ los grados de libertad asociados o coeficientes desconocidos a determinar.

La formulación débil conduce al sistema lineal:

$$\mathbf{KU} = \mathbf{F}, \tag{3.12}$$

donde:

$$K_{ij} = \int_\Omega (\nabla\phi_i)^\top \mathbf{K}(x, y)\nabla\phi_j \, dx \, dy, \tag{3.13}$$

$$F_i = \int_\Omega f\phi_i - \int_\Omega \mathbf{K}(x, y)\nabla G \cdot \nabla\phi_i \, dx \, dy. \tag{3.14}$$

La matriz $\mathbf{K}$ se construye como la suma de matrices de rigidez locales, ensambladas a partir de cada elemento del mallado.

En el caso anisotrópico, la matriz $\mathbf{K}(x, y)$ puede ser de la forma:

$$\mathbf{K}(x, y) = \begin{bmatrix} k_{11}(x, y) & k_{12}(x, y) \\ k_{21}(x, y) & k_{22}(x, y) \end{bmatrix}, \tag{3.15}$$

y se evalúa en cada elemento. Las integrales se aproximan mediante cuadratura numérica, típicamente con reglas de Gauss.

3.5. Existencia y unicidad de la solución de un problema débil.

Teorema 3.1 (Lax-Milgram)**.** *Supongamos que V es un espacio de Hilbert y $a(\cdot, \cdot)$ es una forma bilineal en V que es acotada y V-elíptica. Entonces, dado un $l \in V^*$, existe un único $u \in V$ tal que*

$$a(u, v) = l(v) \quad \forall v \in V.$$

Además, u depende continuamente de l; es decir,

$$\|u\|_V \leq \frac{1}{\alpha}\|l\|_{V^*},$$

donde $\alpha > 0$ es la constante de elipticidad.

3.6. El problema de aproximación. Sea V_h un espacio de dimensión finita tal que $V_h \subset H_0^1$, entonces el problema de aproximación de Galerkin es encontrar $w_h \in V_h$ tal que:

$$a(w_h, v) = l(v) \quad \forall v \in V_h.$$

Donde $a(w_h, v)$ es una forma bilineal y $l(v)$ es un funcional lineal definidos a partir de la forma variacional como

$$a(w_h, v) = \int_\Omega \mathbf{K}\nabla w_h \cdot \nabla v, \tag{3.16}$$

$$l(v) \int_\Omega fv - \int_\Omega \mathbf{K}\nabla G \cdot \nabla v_h.$$

Para resolver este problema tomemos el espacio de dimensión finita de funciones polinómicas de grado 1, ya que este espacio es de dimensión tres las funciones bases están dadas por

$$\Phi_1 = 1 - \xi - \eta, \quad \Phi_2 = \xi, \quad \Phi_3 = \eta. \tag{3.17}$$

Para reducir el coste computacional en la evaluación de las integrales al aplicar el método de elementos finitos, introducimos un dominio de referencia en el que todas las integrales se llevan a cabo sobre el intervalo (-1, 1). De este modo, homogeneizamos el proceso de integración y evitamos repetir cálculos en dominios distintos. Para trasladar cada elemento de su configuración física al dominio de referencia, empleamos la siguiente transformación:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} x_2 - x_1 & x_3 - x_1 \\ y_2 - y_1 & y_3 - y_1 \end{bmatrix} \begin{bmatrix} \xi \\ \eta \end{bmatrix} \tag{3.18}$$

Y con ayuda de esta transformación calcularemos las integrales de todos los elementos en el mismo dominio.

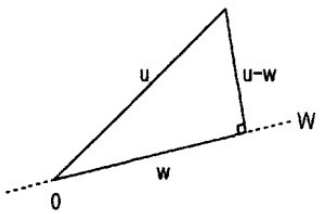


FIGURA 1. Ilustración del problema de aproximación [2]

3.7. Implementación y discusión. Para la implementación se utiliza una malla triangular con elementos de segundo orden (cuadráticos), lo cual permite una mejor aproximación del campo de temperatura en geometrías complejas. Las funciones de forma se construyen sobre el triángulo de referencia, y se trasladan a cada elemento mediante transformaciones afines.

Se implementa un procedimiento de ensamblaje del sistema global. La evaluación de cada entrada $K_{ij}^{(e)}$ requiere:

- el gradiente de las funciones de forma en coordenadas físicas,
- la matriz de conductividad evaluada en el punto de integración,
- el Jacobiano del cambio de coordenadas.

Para iniciar la implementación, se generó una malla sobre un dominio rectangular mediante el paquete `msh` de Octave. A continuación, se definieron dos funciones fundamentales:

- `phi`, encargada de proporcionar las funciones base asociadas a cada elemento de la malla.
- `grad_phi`, que calcula los gradientes de dichas funciones base.

Estas rutinas son imprescindibles para la construcción de las matrices de rigidez locales y de los vectores de carga locales.

En la etapa siguiente, se calcularon las contribuciones elementales mediante las funciones

- `get_localK`, para obtener la matriz de rigidez de cada elemento.
- `get_localF`, para obtener el vector de carga de cada elemento.

Una vez obtenidos los bloques locales, se procedió a ensamblar el sistema global con las rutinas `ensamblar_K` y `ensamblar_F`.

Para cuantificar el error de la aproximación se implementaron dos funciones:

- `L2_norma`, que evalúa la norma L^2 del error.
- `H1_semi_norma`, que calcula la semi-norma H^1.

Finalmente, la función `calculo_de_parametros` agrupa el cálculo de todos los parámetros necesarios para el modelo. Los archivos `test1`, `test2` y `test3` cargan estos parámetros, ejecutan el método de elementos finitos completo, computan los errores y las tasas de convergencia, y representan gráficamente dichos resultados.

4. Experimentos

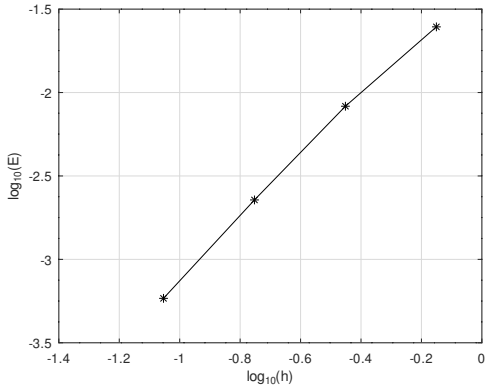
4.1. Experimento 1. Supongamos un dominio $\Omega = [0, 1] \times [0, 1]$ con un tamaño inicial de $h_x = 0, 5$, consideremos el problema de valor en la frontera en (3.5) con condición de Dirichlet homogénea, es decir, $g = 0$ y f elegida de tal forma que la solución exacta sea

$$u(x, y) = xy^2(1 - x^2)(1 - y^2)$$

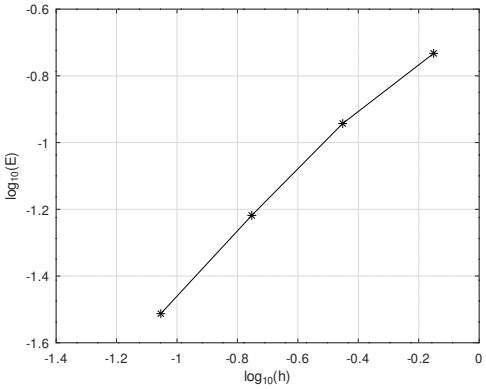
y el tensor de conductividad térmica dado por

$$\begin{bmatrix} 2x+y & x+y-1 \\ x+y-1 & 2x+y \end{bmatrix}$$

con estos datos obtenemos la gráfica del error L^2 y H^1



((A)) Figura 1: Gráfica del error L^2 experimento 1



((B)) Figura 2: Gráfica del error H^1 experimento 1

h_x	h	$\|\|u-u_h\|\|_{L^2}$	**Tasa** L^2	$\|\|u-u_h\|\|_{H^1}$	**Tasa** H^1
0.5000	0.7071	$2{,}4759 \times 10^{-2}$	0.0000	$1{,}8511 \times 10^{-1}$	0.0000
0.2500	0.3536	$8{,}2786 \times 10^{-3}$	1.5805	$1{,}1408 \times 10^{-1}$	0.6983
0.1250	0.1768	$2{,}2712 \times 10^{-3}$	1.8660	$6{,}0499 \times 10^{-2}$	0.9151
0.0625	0.0884	$5{,}8396 \times 10^{-4}$	1.9595	$3{,}0708 \times 10^{-2}$	0.9783

CUADRO 1. Norma L^2 y H^1 y tasas de convergencia para el experimento 1.

La convergencia del método con el calculo de error de la norma L^2 converge a 2 y el calculo de error de la norma H^1 converge a 1.

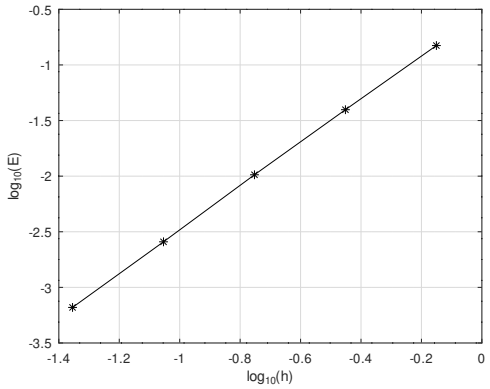
4.2. Experimento 2. Supongamos un dominio $\Omega = [0,1] \times [0,1]$ con un tamaño inicial de $h_x = 0,5$, consideremos el problema de valor en la frontera en (3.5) con condición de Dirichlet homogénea, es decir, $g = 0$ y f elegida de tal forma que la solución exacta sea

$$u(x,y) = \sin(\pi x) \cdot \sin(\pi y)$$

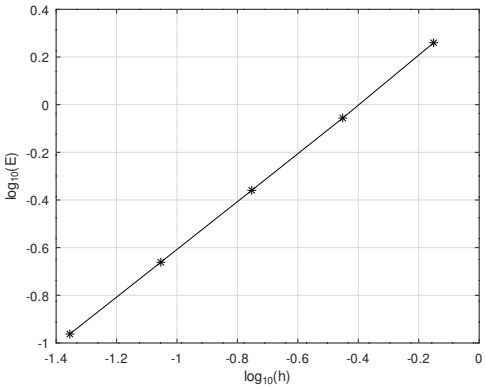
y el tensor de conductividad térmica dado por

$$\begin{bmatrix} x+y+1 & 4x+2y-1 \\ 4x+2y-1 & x+y+1 \end{bmatrix}$$

con estos datos obtenemos la gráfica del error L^2 y H^1



((C)) Figura 1: Gráfica del error L^2 experimento 2



((D)) Figura 2: Gráfica del error H^1 experimento 2

h_x	h	$\|\|u-u_h\|\|_{L^2}$	**Tasa** L^2	$\|\|u-u_h\|\|_{H^1}$	**Tasa** H^1
0.5000	0.7071	$1{,}4938 \times 10^{-1}$	0.0000	$1{,}8188 \times 10^{0}$	0.0000
0.2500	0.3536	$3{,}9577 \times 10^{-2}$	1.9162	$8{,}7849 \times 10^{-1}$	1.0499
0.1250	0.1768	$1{,}0291 \times 10^{-2}$	1.9432	$4{,}3707 \times 10^{-1}$	1.0072
0.0625	0.0884	$2{,}5684 \times 10^{-3}$	2.0025	$2{,}1816 \times 10^{-1}$	1.0025
0.0313	0.0442	$6{,}6033 \times 10^{-4}$	1.9596	$1{,}0907 \times 10^{-1}$	1.0001

CUADRO 2. Norma L^2 y H^1 y tasas de convergencia para el experimento 2.

4.3. Experimento 3 físico. Se estudia una lámina rectangular de madera con dimensiones de 4×4 metros, sobre la cual se aplica una fuente de calor descrita por la función

$$f(x,y) = A \cdot e^{-\frac{(x-x_0-w\cdot x)^2+(y-y_0)^2}{2\sigma^2}},$$

donde A representa la amplitud máxima del calor, (x_0, y_0) son las coordenadas del centro de la fuente térmica, w es un parámetro que modela la influencia de una corriente de aire en la dirección x, y σ controla el grado de dispersión de la fuente. Esta expresión permite representar una distribución térmica afectada por el viento, lo que proporciona un modelo más realista de la propagación del calor. Para esta simulación, se han utilizado los valores $A = 5$, $(x_0, y_0) = (1, 2)$, $w = 6$ y $\sigma = 2$.

$$f(x,y) = 5 \cdot e^{-\frac{(-5x-1)^2+(y-2)^2}{8}}$$

Por la propiedad ortótropica de la madera consideramos el siguiente tensor de conductividad térmica dado para madera laminada cruzada [6]

$$\boldsymbol{\kappa} = \begin{bmatrix} 0{,}638 & 0 \\ 0 & 0{,}321 \end{bmatrix} W/mK$$

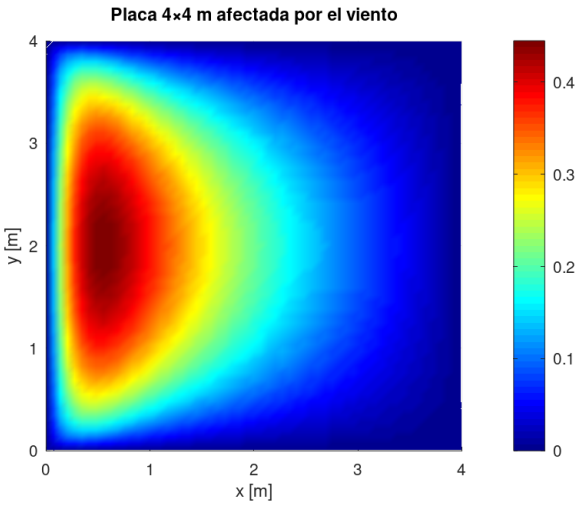


FIGURA 2. Mapa de temperatura U

Con estos datos calculamos el vector U de grados de libertad y realizamos la gráfica de calor

En el gráfico se observa como el calor sobre la lamina de madera tiende a desplazarse en la dirección positiva del eje x debido a la corriente de aire.

4.4. Experimento 4 físico. Para este experimento, consideramos una placa de madera de 3×3 metros que presenta una zona circular de humedad (radio 0,5 metros) centrada en $(1,1)$, donde la conductividad térmica se reduce al 50 %. La función fuente es una gaussiana centrada en (1,5, 1,5):

$$f(x,y) = A \cdot e^{-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}}$$

donde A representa la amplitud máxima del calor, (x_0, y_0) son las coordenadas del centro de la fuente térmica y σ controla el grado de dispersión de la fuente. Esta expresión permite representar una distribución térmica afectada por una zona humeda en el centro de la placa, lo que proporciona un modelo más realista de la propagación del calor. Para esta simulación, se han utilizado los valores $A = 8$, $(x_0, y_0) = (1{,}5, 1{,}5)$, $\sigma = 0{,}3$.

$$f(x,y) = 8 \cdot e^{-\frac{(x-1{,}5)^2+(y-1{,}5)^2}{2(0{,}3)^2}}$$

Por la propiedad ortótropica de la madera consideramos el siguiente tensor de conductividad térmica dado para madera laminada cruzada [6]

$$\boldsymbol{\kappa} = \begin{bmatrix} 0{,}638 & 0 \\ 0 & 0{,}321 \end{bmatrix} W/mK$$

Con estos datos calculamos el vector U de grados de libertad y realizamos la gráfica de calor

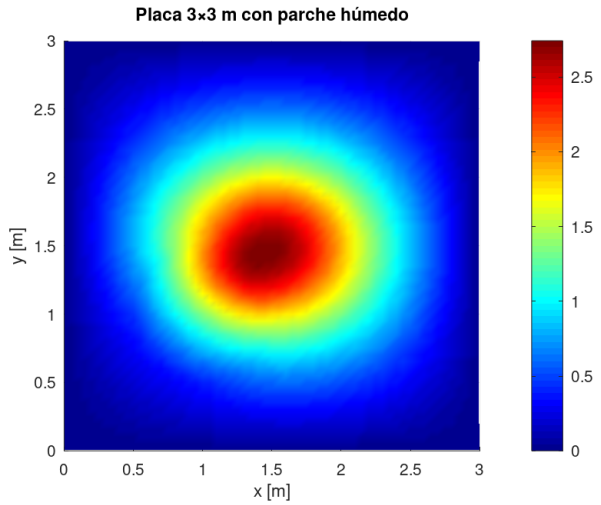


FIGURA 3. Mapa de temperatura U

En este caso, se aprecia que la humedad local frena la difusión del calor en la región circular, produciendo gradientes más pronunciados alrededor del parche húmedo.

5. CONCLUSIONES

- Se implementó con éxito el método de elementos finitos en Octave para resolver el modelo propuesto, confirmando la validez de la aproximación numérica.
- Se realizaron diversos ensayos numéricos que permitieron calcular errores y tasas de convergencia, obteniéndose resultados consistentes y satisfactorios.
- Se llevaron a cabo experimentos físicos que reproducen condiciones reales, los cuales no solo corroboraron los resultados teóricos, sino que demostraron aplicaciones prácticas del método.
- La integración del análisis matemático con herramientas de cómputo mostró su eficacia como enfoque de investigación y validación.
- Se evaluó la influencia de diferentes parámetros del material (anisotropía, no homogeneidad y humedad) en la distribución térmica, revelando comportamientos clave en la difusión del calor.
- El flujo de trabajo desarrollado (generación de malla, ensamblaje, solución y visualización) proporciona un marco reproducible y extensible para futuros estudios en transferencia de calor.

REFERENCIAS

1. F. A. Kulacki et al, eds, *Handbook of Thermal Science and Engineering* Springer, Cham, 2018.
2. M. S. Gockenbach, *Understanding and Implementing the Finite Element Method*, SIAM, Philadelphia, 2006.

3. F. P. Incropera and D. P. DeWitt, *Fundamentos de Transferencia de Calor*, 4th ed., Pearson Prentice Hall, México, 2002.
4. T. N. Narasimhan, *Fourier's heat conduction equation: history, influence, and connections*. Artículo de la revista REV. GEOPHYS. 37 (1999), No. 1, pp. 151–172, DOI 10.1029/1998RG900006.
5. R. N. Bracewell, *The Fourier Transform and Its Applications*, 3rd ed., McGraw-Hill, New York, 2000.
6. A. R. Díaz, *Multiscale modeling of the thermal conductivity of wood and its application to cross-laminated timber*. Artículo de la revista INT. J. THERM. SCI. 144 (2019), No. 6, pp. 79–92, DOI 10.1016/j.ijthermalsci.2019.05.016.

Departamento de matemáticas, Universidad Nacional Autónoma de Honduras

Dirección de correo electrónico: lhmaradiaga@unah.hn

ANÁLISIS ESPACIAL DE PRECIOS DE AIRBNB EN NYC CON GEOESTADÍSTICA Y ECONOMETRÍA

OSCAR ANDRÉE CORTÉS HERNÁNDEZ

Resumen. Este estudio analiza los precios de alojamientos de Airbnb en la Ciudad de Nueva York mediante estadística espacial y econometría. Se revisan los fundamentos de autocorrelación espacial, función de covarianza, variograma y kriging, junto con modelos de regresión espacial (SAR, SEM, SARAR). La metodología integra análisis exploratorio, estimación variográfica, predicción por kriging y modelado econométrico, con validación cruzada y métricas de desempeño. Los resultados evidencian patrones espaciales significativos y muestran que ubicación y tipo de alojamiento son determinantes clave en la formación de precios. El flujo de trabajo se implementó en R con paquetes especializados para datos georreferenciados.

Abstract. This study analyzes Airbnb listing prices in New York City using spatial statistics and econometrics. We review the foundations of spatial autocorrelation, covariance function, variogram, and kriging, along with spatial regression models (SAR, SEM, SARAR). The methodology integrates exploratory analysis, variogram estimation, kriging prediction, and econometric modeling, with cross-validation and performance metrics. Results reveal significant spatial patterns and show that location and room type are key determinants in price formation. The workflow was implemented in R using packages specialized for georeferenced data.

1. Introducción

El análisis estadístico de datos espaciales ha experimentado un crecimiento exponencial debido a la disponibilidad masiva de información georreferenciada y el desarrollo de herramientas computacionales especializadas [1,2]. Este campo combina fundamentos teóricos sólidos con aplicaciones prácticas que permiten comprender fenómenos cuya variabilidad está influenciada por su localización geográfica.

Los mercados inmobiliarios y de alojamiento representan casos de estudio paradigmáticos para el análisis espacial, ya que exhiben patrones de dependencia espacial complejos influenciados por factores de localización, accesibilidad, amenidades urbanas y características del entorno [9]. El surgimiento de plataformas digitales como Airbnb ha generado datasets ricos y granulares que permiten aplicar técnicas geoestadísticas y econométricas avanzadas para modelar la formación de precios espaciales.

Fecha 12 de Agosto de 2025.

Palabras y frases clave. Análisis espacial, Autocorrelación espacial, Geoestadística, kriging, Econometría espacial, Modelos SAR/SEM/SARAR/SLX, Indicadores LISA, Airbnb.

Los objetivos de este trabajo son: Revisar los fundamentos teóricos del análisis estadístico espacial, incluyendo la taxonomía de datos espaciales y métodos de modelado, aplicar técnicas de geoestadística y econometría espacial para analizar patrones de precios, demostrar la integración metodológica entre análisis exploratorio espacial, modelado variográfico y regresión espacial y evaluar el desempeño predictivo de diferentes enfoques.

La ciudad de Nueva York constituye un laboratorio urbano ideal para esta aplicación por su diversidad espacial, densidad de observaciones y heterogeneidad de mercados de alojamiento. Los patrones espaciales identificados y las técnicas metodológicas desarrolladas tienen aplicabilidad general para contextos urbanos similares.

Este estudio contribuye al desarrollo metodológico del análisis espacial aplicado, proporcionando un marco integrado que combina exploración, modelado y predicción espacial con implementación práctica en R [2].

2. Antecedentes

El análisis estadístico espacial ha evolucionado de manera sistemática desde mediados del siglo XX, consolidando marcos teóricos y metodológicos especializados. Los trabajos pioneros de Moran y Geary [5, 6] establecieron la base para la medición de la autocorrelación espacial mediante estadísticos diseñados para cuantificar la dependencia en datos geográficos. Posteriormente, Matheron [7] desarrolló la teoría de variables regionalizadas y el marco conceptual para analizar fenómenos espaciales continuos, lo que permitió formalizar técnicas de interpolación óptima. Ripley [8] realizó una primera síntesis exhaustiva de métodos estadísticos para datos espaciales, abordando distintos tipos de datos y técnicas asociadas.

El desarrollo de modelos de regresión espacial fue sistematizado por Anselin [9], quien estableció el marco teórico de los modelos autorregresivos y de error espacial, y posteriormente introdujo los indicadores locales de asociación espacial (LISA) [10], útiles para identificar agrupamientos y heterogeneidad espacial. En paralelo, Cressie [1] propuso una taxonomía que distingue tres tipos principales de datos espaciales geoestadísticos, de red (lattice) y patrones de puntos unificando diversas áreas mediante una notación consistente. Más recientemente, Banerjee et al. [11] incorporaron enfoques bayesianos jerárquicos para el análisis de datos espaciales a gran escala.

En años recientes, el análisis estadístico espacial ha encontrado aplicaciones relevantes en el estudio de mercados inmobiliarios y de alojamiento temporal. Investigaciones como las de Can y Dubin [13, 14] han demostrado la importancia de incluir la dependencia espacial en modelos hedónicos de precios, mientras que la disponibilidad de datos provenientes de plataformas digitales ha facilitado estudios a gran escala que combinan geoestadística y econometría espacial para comprender la formación de precios en mercados urbanos [15, 16].

3. Fundamentos Teóricos de Geoestadística y Econometría Espacial

El análisis estadístico espacial se fundamenta en el estudio de fenómenos cuya variabilidad está intrínsecamente vinculada a su localización geográfica. Su base matemática exige comprender la naturaleza de los datos espaciales, los procesos estocásticos que los generan y los métodos estadísticos adecuados para su estudio. La clasificación propuesta por Cressie [1] distingue tres categorías principales que determinan la metodología de análisis: datos geoestadísticos, datos de red y patrones de puntos.

Definición 3.1 (Datos Geoestadísticos)**.** Observaciones de un proceso estocástico espacial $\{Z(s) : s \in D \subset \mathbb{R}^d\}$ definido en un dominio espacial continuo D, donde d es la dimensión espacial. Estos datos exhiben dependencia espacial continua.

La dependencia espacial en este contexto puede modelarse mediante la *función de covarianza* (Def. 3.6) o el *variograma* (Def. 3.7), que describen la estructura de dependencia en función del vector de separación espacial h [1].

Definición 3.2 (Datos de Red)**.** Observaciones $\{Z_i : i \in I\}$ asociadas a un conjunto finito I de regiones discretas y bien definidas. La dependencia espacial se describe mediante matrices de pesos espaciales $W = \{w_{ij}\}$ que definen vecindades.

Definición 3.3 (Patrones de Puntos)**.** Conjunto aleatorio $\Phi = \{s_1, s_2, \ldots, s_N\}$ de puntos en un dominio $D \subset \mathbb{R}^d$, donde tanto las posiciones como el número de puntos pueden ser aleatorios. Se caracteriza mediante la función de intensidad $\lambda(s)$.

La medición de la dependencia espacial es fundamental. El índice I de Moran [5] es uno de los estadísticos más empleados para medir la autocorrelación espacial global.

Teorema 3.4 (Índice I de Moran)**.** *Para una variable x_i en ubicaciones $i = 1, \ldots, n$,*

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2},$$

donde w_{ij} son elementos de la matriz de pesos espaciales y $\bar{x}$ es la media muestral.

El estadístico C de Geary [6] ofrece una medida alternativa, basada en diferencias entre observaciones vecinas:

$$C \;=\; \frac{(n-1)}{2W} \, \frac{\sum_i \sum_j w_{ij} \, (x_i - x_j)^2}{\sum_i (x_i - \bar{x})^2}, \qquad W \;=\; \sum_i \sum_j w_{ij}.$$

En I de Moran, valores cercanos a 1 indican autocorrelación positiva, cercanos a -1 negativa, y próximos a 0 ausencia de autocorrelación. Para C de Geary, valores cercanos a 0 implican autocorrelación positiva, cercanos a 2 negativa, y cercanos a 1 ausencia de autocorrelación [9].

Los modelos de regresión espacial extienden la regresión clásica incorporando estructura espacial en la variable dependiente o en los errores [9].

Teorema 3.5 (Modelo Autorregresivo Espacial (SAR))**.**

$$y = \rho W y + X\beta + \epsilon,$$

donde y es la variable dependiente, ρ el coeficiente de autocorrelación, W la matriz de pesos, X la matriz de covariables, β el vector de parámetros y $\epsilon \sim N(0, \sigma^2 I)$.

El modelo de error espacial (SEM) describe dependencia en los errores:

$$y = X\beta + u, \quad u = \lambda W u + \epsilon,$$

con λ como coeficiente de autocorrelación en los errores [9].

La geoestadística, fundamentada en la teoría de Matheron [7], es esencial para datos espaciales continuos.

Definición 3.6 (Función de covarianza)**.** Sea $\{Z(s) : s \in D \subset \mathbb{R}^d\}$ un proceso estocástico espacial. La *función de covarianza* se define como

$$C(h) \;=\; \mathrm{Cov}\big(Z(s), Z(s+h)\big),$$

la cual mide la dependencia lineal entre los valores de Z en dos ubicaciones separadas por un vector de desplazamiento h.

Definición 3.7 (Variograma)**.** Sea $\{Z(s) : s \in D \subset \mathbb{R}^d\}$ un proceso estocástico espacial. El *variograma* se define como

$$\gamma(h) \;=\; \tfrac{1}{2}\,\mathrm{Var}\big(Z(s+h) - Z(s)\big),$$

y describe cómo varía la similitud de las observaciones en función de la separación espacial h.

El variograma describe cómo varía la similitud de observaciones con la distancia [1]. El estimador empírico es:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[Z(s_i) - Z(s_i + h)\right]^2.$$

Teorema 3.8 (Kriging ordinario)**.**

$$\hat{Z}(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i),$$

con $\sum_{i=1}^{n} \lambda_i = 1$ y los pesos elegidos para minimizar la varianza de predicción.

El análisis de patrones de puntos se basa en procesos puntuales [3, 4]. El proceso de Poisson homogéneo, con independencia completa, tiene:

$$P(N(A) = k) = \frac{(\Lambda(A))^k e^{-\Lambda(A)}}{k!},$$

donde $\Lambda(A) = \int_A \lambda(s)\, ds$.

Funciones como la K de Ripley y la distribución de distancias al vecino más cercano permiten describir y comparar patrones [3].

Ejemplo 3.9 (Autocorrelación espacial con KNN en NYC e implementación en R)**.** Considérese un subconjunto sintético de listados de Airbnb en NYC (coordenadas proyectadas en NAD83 / UTM 18N, EPSG:26918). Se define la matriz de pesos W a partir de los $k = 8$ vecinos más cercanos (estandarización por filas) con simetrización de la vecindad. Sobre $y = \log(\text{price})$ se evalúa la autocorrelación global con el índice

I de Moran y la autocorrelación local con LISA para identificar clústeres High–High y Low–Low. Este ejemplo ilustra la construcción de W y la interpretación de I y LISA en el mismo marco metodológico empleado en el estudio; véanse las Figuras 1 y 2 para la aplicación completa.

Fragmento en R:

```
library(sf)
library(spdep)
library(spatialreg)

# airbnb_sf: sf con columnas price, room_type, minimum_nights,
# availability_365, number_of_reviews, reviews_per_month,
# neighbourhood_group, geometry (WGS84)
airbnb_sf <- st_transform(airbnb_sf, 26918)  # NAD83 / UTM 18N (metros)

coords <- st_coordinates(airbnb_sf)
nb  <- knn2nb(knearneigh(coords, k = 8), sym = TRUE)  # simetrizar KNN
lw  <- nb2listw(nb, style = "W")

airbnb_df <- st_drop_geometry(airbnb_sf)

# OLS base
ols <- lm(log(price) ~ room_type + minimum_nights + availability_365 +
            number_of_reviews + reviews_per_month + neighbourhood_group,
          data = airbnb_df)

# Prueba de Moran en residuales OLS
moran.test(residuals(ols), lw)

# SAR por ML
sar <- lagsarlm(log(price) ~ room_type + minimum_nights + availability_365 +
                  number_of_reviews + reviews_per_month +
                  ↪  neighbourhood_group,
                data = airbnb_df, listw = lw)
summary(sar)
```

En la práctica, el análisis debe complementarse con la verificación de supuestos como estacionariedad e isotropía [1], que garantizan que las propiedades estadísticas sean invariantes ante traslaciones espaciales y que la dependencia no dependa de la dirección. La violación de estos supuestos puede requerir modelos no estacionarios o enfoques direccionales.

La validación de modelos espaciales debe considerar la dependencia inherente en los datos para evitar la subestimación de la varianza de predicción. Métodos como la validación cruzada espacialmente estructurada (dejando fuera la observación objetivo y sus vecinos) resultan más adecuados [2]. Las métricas habituales de evaluación incluyen el error cuadrático medio de predicción (RMSE), el error absoluto medio (MAE) y medidas de calibración para evaluar la calidad de las estimaciones de incertidumbre.

4. Metodología y resultados del análisis espacial de precios de Airbnb en NYC

4.1. Metodología. La investigación adopta un enfoque cuantitativo con métodos de estadística espacial para describir y modelar la estructura geográfica de los precios de alojamientos. ¿Cómo varía espacialmente el precio por noche de los alojamientos de Airbnb en Nueva York y qué factores lo explican? El objetivo general es caracterizar la dependencia espacial del precio y construir un modelo predictivo que incorpore covariables del alojamiento y de localización, apoyado en los marcos de geoestadística y econometría espacial [1, 2, 9].

Se utilizó el conjunto de datos público de listados de Airbnb para NYC (2019). Se consideraron las variables: `price` (dependiente), `longitude`, `latitude`, `room_type`, `availability_365`, `minimum_nights`, `number_of_reviews`, `reviews_per_month` y `neighbourhood_group`. Los pasos de preparación fueron:

1. Limpieza básica: se filtraron registros con `price` ≤ 0 y precios extremos (≥ 1000 USD) para atenuar la influencia de valores anómalos en la estimación y validación [2].
2. Transformación de la variable objetivo: se trabajó con log(price) para aproximar la normalidad y estabilizar la varianza, práctica habitual en modelado espacial y geoestadístico [1].
3. Estandarización de covariables continuas: se escalaron (z-score) `minimum nights`, `availability_365`, `number_of_reviews`, `reviews_per_month`, lo que facilita la interpretación y la estimación estable en presencia de diferentes escalas [2].
4. Representación espacial: los datos se convirtieron a objeto `sf` en WGS84 (EPSG:4326) y se reproyectaron a NAD83 / UTM zona 18N (EPSG:26918), un CRS métrico. Todas las parametrizaciones dependientes de distancia se especificaron en metros: variograma con cutoff = 30,000 m y bin width = 600 m, kriging con cellsize = 600 m, y hexagonado de ≈ 800 m.

El Análisis espacial exploratorio (ESE) tuvo tres componentes:

1. Mapas exploratorios: distribución espacial de listados y mapas de cuantiles de log(price) para detectar heterogeneidad espacial visible [2, 8].
2. Autocorrelación global: se aplicó el índice I de Moran sobre log(price) con matriz de pesos W construida por k-vecinos más cercanos ($k = 8$), simetrizada y estandarizada por filas, práctica común cuando no se dispone de polígonos administrativos homogéneos [2, 5].
3. Autocorrelación local (LISA): se estimaron indicadores locales para identificar clústeres significativos High–High, Low–Low y patrones High–Low/Low–High (outliers espaciales), siguiendo [9, 10].

Para facilitar comparaciones y visualizaciones, se construyó un índice de precio normalizado en $[0, 1]$ a partir del precio en escala natural:

$$\texttt{price\_index} = \frac{\text{price} - \text{mín(price)}}{\text{máx(price)} - \text{mín(price)}}.$$

El uso de transformaciones y reescalamiento de variables en flujos de trabajo espaciales está documentado en [2]. En esta aplicación no se redujeron covariables mediante PCA, aunque dicha alternativa es coherente con las prácticas descritas por este autor.

Modelado: geoestadística y econometría espacial.

Geoestadística (variograma y kriging). Sea $Z(s)$ el log-precio en la localización s. Se estimó el variograma empírico $\hat{\gamma}(h)$ y se ajustaron modelos esférico, exponencial y gaussiano; la selección se realizó por el error cuadrático del ajuste (SSE) [1, 7]. Dado el mejor modelo, se ejecutó kriging ordinario (KO) para obtener predicciones $\hat{Z}(s_0)$ y su varianza de kriging, generando una superficie continua de precios [1]. La validación cruzada (dejar-uno-fuera) evaluó RMSE ≈ 0.367, MAE ≈ 0.250 y $R^2 \approx 0.339$ en escala logarítmica:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(e_i)^2}, \quad \text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|e_i|, \quad R^2 = 1 - \frac{\sum(e_i)^2}{\sum(Z_i - \bar{Z})^2}.$$

en línea con las recomendaciones de evaluación para predictores espaciales [1].

Econometría espacial (SAR/SEM/SARAR/SLX). Como línea base se estimó un modelo OLS para log(price) con covariables del alojamiento (tipo de habitación, noches mínimas, reseñas, disponibilidad, listados por anfitrión) y dummies de localización (neighbourhood_group). Sobre sus residuales se aplicaron pruebas de dependencia espacial (Moran y LM/LM robustas) para decidir la especificación espacial adecuada [2, 9].

Se consideraron:

- SAR (rezago espacial en la dependiente): $y = \rho W y + X\beta + \varepsilon$
- SEM (error espacial): $y = X\beta + u$, con $u = \lambda W u + \varepsilon$
- SARAR (rezago y error): combinación de ambas estructuras.
- SLX (rezago en covariables): inclusión de WX para capturar efectos de vecindario en explicativas.

Las estimaciones se realizaron mediante máxima verosimilitud y procedimientos IV/GMM cuando fue pertinente (por ejemplo, `stsls` y `GMerrorsar`), conforme a [2, 9]. La comparación de modelos empleó criterios de verosimilitud (logLik) y de información (AIC/BIC), además de la ausencia de autocorrelación en residuales como condición de especificación adecuada [2].

Notas sobre reproducibilidad y software. Todo el flujo analítico (ESE, variograma/kriging y modelos espaciales) se implementó en R con los paquetes `sf`, `spdep`, `gstat`, `spatialreg`, `sphet` y `tmap`, cuyo uso aplicado y fundamentos metodológicos se discuten en [2]. La geoestadística y las propiedades de kriging siguen [1,7] mientras que la detección de asociación espacial local y los modelos SAR/SEM/SARAR se apoyan en [9,10].

4.2. Resultados. Se presentan los hallazgos empíricos agrupados en cuatro bloques que abarcan la exploración espacial de precios, el análisis de la dependencia y los clústeres locales, la caracterización de la continuidad espacial mediante variograma con predicción por kriging, y el contraste con modelos espaciales de regresión. Las interpretaciones se fundamentan en referentes teóricos y metodológicos consolidados en la literatura [1,5,7,10].

Exploración espacial de precios.

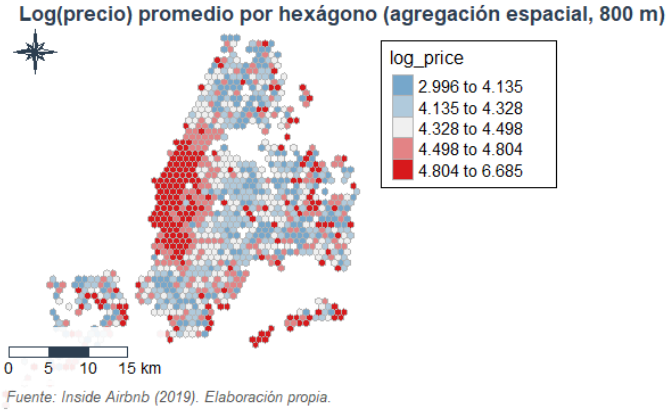


Figura 1. Distribución espacial de log(price) (cuantiles). Fuente de datos: Inside Airbnb (2019). Elaboración propia con R/`tmap`.

La Figura 1 evidencia un gradiente espacial nítido: Manhattan y corredores centrales de Brooklyn concentran los cuantiles superiores de log(price), mientras que Staten Island y sectores de Queens caen en cuantiles inferiores. La agregación hexagonal mitiga ruido local, realzando patrones de nivel medio–alto coherentes con centralidad urbana y accesibilidad, lo que coincide con la literatura sobre heterogeneidad y gradientes urbanos [2,13,14]. Este patrón sugiere dependencia espacial positiva de los precios, motivando las pruebas de autocorrelación [5,10].

Autocorrelación espacial: Moran global y LISA.

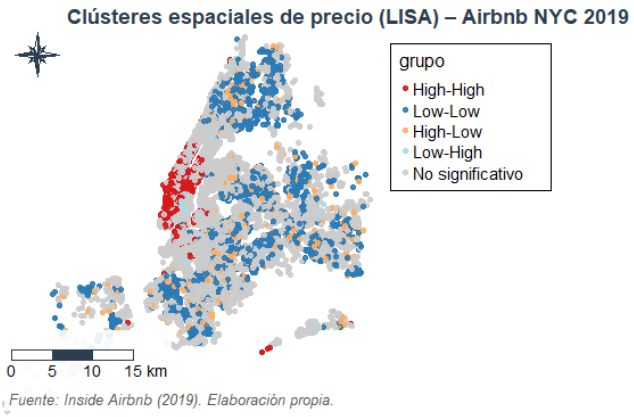


Figura 2. Clústeres locales de autocorrelación espacial (LISA) para log(price). Categorías: High–High, Low–Low, High–Low, Low–High y no significativo [10].

La Figura 2 muestra clústeres High–High en Manhattan y partes de Brooklyn, y Low–Low en Staten Island y Queens. Los patrones High–Low y Low–High son escasos y localizados, reflejando una estructura espacial relativamente homogénea dentro de cada macrozona [2, 10].

Geoestadística: variograma y kriging.

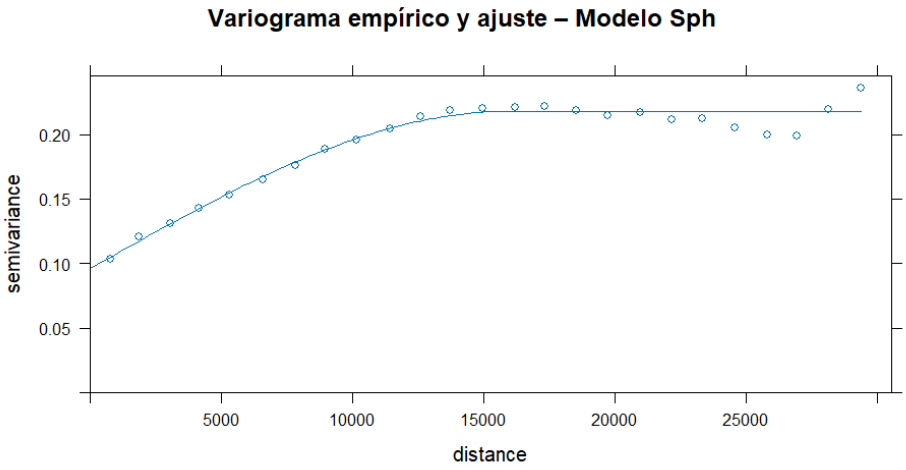


Figura 3. Variograma empírico de log(price) (bin 600 m, cutoff 30 km) con ajuste esférico (SSE mínimo). Parámetros: nugget ≈ 0.096, sill parcial ≈ 0.122, rango ≈ 15.97 km. Distancias en metros.

La Figura 3 muestra que el nugget sugiere variación micro-espacial/no observada y posible error de medición, el sill parcial indica la varianza estructural capturada por el modelo, y el rango cercano a 16 km implica que la similitud espacial decae hasta estabilizarse a esa escala intraurbana. Este comportamiento concuerda con aplicaciones urbanas de la geoestadística [1, 7], justificando el uso de kriging con un modelo esférico.

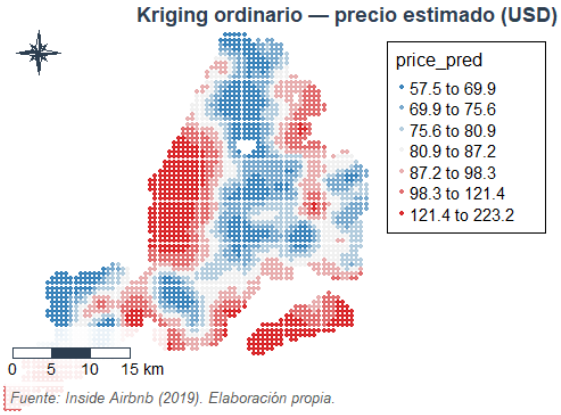


Figura 4. Superficie de precio estimado (USD) por kriging ordinario. Validación: RMSE = 0.367, MAE = 0.250, $R^2 = 0.339$ (escala log).

La Figura 4 muestra que la superficie predicha reproduce picos de precio en centralidades (especialmente Manhattan) y declives graduales hacia periferias. Las métricas LOOCV indican precisión moderada, razonable dada la heterogeneidad urbana y la escala del fenómeno, y comparables a resultados reportados en estudios afines. Este mapa es útil para *screening* territorial y para la comunicación con actores no técnicos [1, 2].

Elección y desempeño de modelos de regresión espacial (SAR/SEM/SARAR/SLX).

El análisis de regresión sobre la variable transformada $y = \log(\text{price})$ inició con un modelo OLS, el cual alcanzó un $R^2_{\text{aj}} \approx 0.502$. En este modelo, el tipo de habitación mostró un efecto claro: las room presentaron un coeficiente $\beta \approx -0.771$ y las Shared room $\beta \approx -1.164$, lo que implica reducciones aproximadas del 54 % y 69 % respectivamente en escala natural respecto a Entire home/apt. Asimismo, se observó una prima sustancial para Manhattan, cercana al 75 % respecto al Bronx. No obstante, los residuales evidenciaron autocorrelación espacial positiva significativa (Moran $p < 0.001$), lo que contraviene el supuesto de independencia y sugiere la necesidad de modelos que incorporen explícitamente la estructura espacial [9].

En consecuencia, se evaluaron modelos SAR, SEM y SARAR, seleccionando como más adecuado el modelo SARAR estimado mediante GMM (`gstsls`). Este presentó un parámetro de error espacial $\lambda \approx 0.454$, estadísticamente significativo, y residuales sin autocorrelación (Moran $I \approx -0.0047$, $p \approx 0.985$), lo que indica que la dependencia

espacial en los errores fue corregida de manera efectiva. Este ajuste, coherente con la teoría de econometría espacial [2, 9], mejora la robustez de las estimaciones y la validez de las inferencias.

Los resultados globales confirman que existe autocorrelación espacial tanto a nivel general como en forma de clústeres locales, que la continuidad intraurbana de los precios puede ser capturada de forma adecuada mediante un variograma esférico, y que el kriging ofrece una superficie de predicción interpretable con una precisión moderada. La inclusión del modelo SARAR corrige la dependencia residual y refuerza la relevancia de variables como `room_type` y `neighbourhood_group` en la formación de precios, respaldando metodologías que integran geoestadística y econometría espacial [1, 2, 9, 10].

Tabla 1. Resumen de resultados clave del análisis espacial y geoestadístico

Componente	**Resultado principal**
Moran global (log-precio)	Autocorrelación positiva significativa ($p < 0.001$).
LISA (clústeres)	High–High en Manhattan y zonas de Brooklyn; Low–Low en Staten Island y sectores de Queens.
Variograma seleccionado	Esférico; nugget 0.096; sill parcial 0.122; rango ≈ 15.97 km; SSE $\approx 9.6e{-}8$.
Kriging (CV, escala log)	RMSE $= 0.367$; MAE $= 0.250$; $R^2 = 0.339$.
OLS (log-precio)	$R^2_{\text{aj}} \approx 0.502$; efectos fuertes de tipo de habitación y ubicación.
SAR (STSLS)	ρ no significativo al 5 %; residuales con autocorrelación.
SEM (GMM)	$\lambda = 0.438$ ($z \approx 25.3$), $\sigma^2 \approx 0.196$.
SARAR (GMM)	$\lambda \approx 0.454$; residuales sin autocorrelación (Moran $I \approx -0.0047$, $p \approx 0.985$).
Efectos ilustrativos	Private room ≈ -0.77, Shared room ≈ -1.16, Manhattan ≈ 0.559, Brooklyn ≈ 0.249 (en log-precio).

Los resultados indican que existe una dependencia espacial significativa de los precios, confirmada por Moran y LISA [5, 10]. El variograma esférico captura de forma adecuada la continuidad espacial a escala intraurbana [1, 7], lo que respalda la validez de su uso en este contexto. El kriging logra una superficie de predicción coherente con la distribución observada, con un desempeño estadístico aceptable para aplicaciones exploratorias y de apoyo a la toma de decisiones [1, 2]. Finalmente, los modelos SARAR corrigen eficazmente la dependencia residual, permitiendo estimaciones más precisas y confiables de los efectos de variables clave como el tipo de habitación y la localización [2, 9].

5. Conclusiones

El presente estudio permitió responder de manera satisfactoria la pregunta central sobre la existencia, magnitud y forma de la dependencia espacial en los precios de alojamiento en Airbnb en la ciudad de Nueva York. Los resultados muestran, en primer lugar, una autocorrelación espacial positiva estadísticamente significativa, confirmada por el índice de Moran global y el análisis LISA [5, 10], lo que valida la hipótesis de que los precios no se distribuyen aleatoriamente en el espacio.

En segundo lugar, el análisis geoestadístico mediante variograma esférico [1] evidenció una continuidad espacial de los precios a una escala aproximada de 16 km, lo que permitió aplicar kriging ordinario para generar una superficie de predicción con un nivel de detalle útil para la interpretación territorial y la comunicación de resultados a audiencias no técnicas.

Finalmente, el contraste con modelos de regresión espacial reveló que la especificación SARAR estimada por GMM [9] fue la más adecuada, al corregir eficazmente la dependencia residual y proporcionar estimaciones más robustas de los efectos de variables clave como el tipo de habitación y la localización. Este hallazgo refuerza la relevancia de incorporar la dimensión espacial en el análisis de precios inmobiliarios y turísticos para mejorar la precisión predictiva y sustentar decisiones estratégicas en planificación urbana y de mercado.

Como línea futura de investigación, se sugiere ampliar el conjunto de covariables, explorar modelos espacio-temporales y métodos no paramétricos, así como técnicas bayesianas y enfoques de machine learning espacial para capturar no linealidades y efectos complejos [1]. También se recomienda replicar la metodología en otras ciudades con dinámicas de mercado diferenciadas, lo que permitiría evaluar la transferibilidad y robustez del enfoque propuesto.

Referencias

1. N. A. C. Cressie, *Statistics for Spatial Data*, Wiley-Interscience, New York, 1993.
2. R. Bivand, E. Pebesma, and V. Gómez-Rubio, *Applied Spatial Data Analysis with R*, Second edition, Springer, New York, 2013.
3. A. Baddeley, E. Rubak, and R. Turner, *Spatial Point Patterns: Methodology and Applications with R*, Chapman and Hall/CRC, Boca Raton, FL, 2015.
4. P. J. Diggle, *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*, Third edition, CRC Press, Boca Raton, FL, 2014.
5. P. A. P. Moran, Notes on continuous stochastic phenomena, *Biometrika* **37** (1950), no. 1, 17–23.
6. R. C. Geary, The contiguity ratio and statistical mapping, *The Incorporated Statistician* **5** (1954), no. 3, 115–141.
7. G. Matheron, Principles of geostatistics, *Economic Geology* **58** (1963), no. 8, 1246–1266.
8. B. D. Ripley, *Spatial Statistics*, John Wiley & Sons, New York, 1981.
9. L. Anselin, *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht, 1988.
10. L. Anselin, Local indicators of spatial association—LISA, *Geographical Analysis* **27** (1995), no. 2, 93–115.
11. S. Banerjee, B. P. Carlin, and A. E. Gelfand, *Hierarchical Modeling and Analysis for Spatial Data*, Second edition, CRC Press, Boca Raton, FL, 2014.
12. Inside Airbnb, New York City Airbnb Open Data, *Inside Airbnb*, 2019. `http://insideairbnb.com/get-the-data.html`

13. A. Can, The measurement of neighborhood dynamics in urban house prices, *Economic Geography* **66** (1990), no. 3, 254–272.
14. R. Dubin, Spatial autocorrelation: A primer, *Journal of Housing Economics* **7** (1998), no. 4, 304–327.
15. D. Wang and J. L. Nicolau, Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com, *International Journal of Hospitality Management* **62** (2017), 120–131.
16. K. Gyódi and L. Nawaro, Determinants of Airbnb prices in European cities: A spatial econometrics approach, *Tourism Management* **86** (2021), 104319.

Licenciatura en Matemática, Universidad Nacional Autónoma de Honduras

Dirección de correo electrónico: oacortes@unah.hn

TÉCNICAS DE REGULARIZACIÓN EN EL DIAGNÓSTICO DE TDAH

THELMA FONSECA

Resumen. El Trastorno por Déficit de Atención e Hiperactividad (TDAH) es una condición neuropsiquiátrica común que presenta desafíos diagnósticos debido a la complejidad de sus síntomas. Este estudio aplica técnicas de regresión penalizada, específicamente Lasso, junto con el algoritmo de k-vecinos más cercanos (KNN), para analizar un conjunto de datos clínicos tabulares relacionados con el TDAH. El objetivo es mejorar la identificación de variables clave asociadas al trastorno y construir modelos predictivos robustos que ayuden en el diagnóstico. Los resultados destacan la utilidad de la regularización y de los métodos basados en la proximidad para la selección de variables, la clasificación de los pacientes y el manejo del sobreajuste, contribuyendo así a la neurociencia clínica y demostrando la relevancia del análisis de datos en la salud mental.

Abstract. Attention Deficit Hyperactivity Disorder (ADHD) is a common neuropsychiatric condition that poses diagnostic challenges due to the complexity of its symptoms. This study applies penalized regression techniques, specifically Lasso, together with the k-nearest neighbors (KNN) algorithm, to analyze a clinical tabular dataset related to ADHD. The objective is to improve the identification of key variables associated with the disorder and to build robust predictive models that support diagnosis. The results highlight the usefulness of regularization and proximity-based methods for variable selection, patient classification, and overfitting management, thereby contributing to clinical neuroscience and demonstrating the relevance of data analysis in mental health.

1. Introducción

El Trastorno por Déficit de Atención e Hiperactividad (TDAH) es una condición neuropsiquiátrica prevalente en niños y adolescentes que afecta significativamente el rendimiento académico, social y emocional. A pesar de su alta incidencia, el diagnóstico clínico del TDAH enfrenta retos debido a la complejidad de sus síntomas y la heterogeneidad en la presentación clínica [2].

En este contexto, el análisis de datos neuropsicológicos y clínicos mediante técnicas estadísticas y de aprendizaje automático ofrece una oportunidad para mejorar la precisión diagnóstica y la comprensión de los factores asociados al trastorno. Sin embargo, los conjuntos de datos disponibles suelen contener numerosas variables con alta correlación, lo que puede afectar la interpretabilidad y el rendimiento de los modelos predictivos.

Fecha: Agosto 2025.

Palabras y frases clave. Regresión Lasso, K-vecinos más cercanos, Regularización, Selección de modelos, TDAH.

Para abordar este problema, este estudio emplea dos enfoques complementarios: (i) métodos de regresión penalizada, específicamente Lasso [1], con el objetivo de realizar selección de variables y reducir el sobreajuste, y (ii) el algoritmo de k-vecinos más cercanos (KNN), un modelo de clasificación no paramétrico basado en la proximidad de los datos. Esta comparación permite evaluar las fortalezas y limitaciones de cada técnica en el diagnóstico de TDAH, considerando tanto la selección de variables relevantes como el desempeño predictivo.

El objetivo principal de esta investigación es aplicar y comparar ambos métodos en un conjunto de datos clínicos tabulares de pacientes con sospecha o diagnóstico de TDAH, demostrando cómo las herramientas de análisis de datos pueden contribuir a la neurociencia y a la práctica clínica. Además, este trabajo resalta la importancia de las habilidades en análisis de datos para la interpretación de fenómenos complejos en neuropsicología, mostrando la sinergia entre estadística avanzada, algoritmos de aprendizaje automático y ciencia aplicada.

Los resultados esperados incluyen un modelo predictivo robusto, la identificación de variables clínicas clave mediante Lasso y una evaluación comparativa con KNN, así como una reflexión sobre las ventajas de integrar distintos métodos estadísticos y de clasificación en estudios neurocientíficos. Este enfoque interdisciplinario contribuye tanto al campo de la salud mental como al desarrollo profesional en análisis de datos, posicionando este estudio como un aporte relevante para investigadores y profesionales en ambos ámbitos. Este tema se encuentra dentro de los ejes prioritarios de investigación de la Universidad Nacional Autónoma de Honduras (UNAH), especialmente en el eje estratégico de Salud Mental y Neurociencias, y en la línea de acción de aplicación de tecnologías estadísticas y computacionales para el diagnóstico clínico.

2. ANTECEDENTES

La regresión "Lasso" y el algoritmo "k-vecinos más cercanos" (KNN) son técnicas fundamentales en el análisis estadístico y de aprendizaje automático, empleadas para mejorar la generalización, la selección de variables y la clasificación de datos complejos.

La regresión Lasso ("Least Absolute Shrinkage and Selection Operator") fue propuesta por Robert Tibshirani en 1996 como una solución a la necesidad de realizar simultáneamente regularización y selección de variables. Mediante una penalización ℓ_1, Lasso permite reducir a cero algunos coeficientes, simplificando los modelos y facilitando su interpretación. Esta propiedad ha demostrado ser especialmente útil en contextos de alta dimensionalidad, como la genómica o la neuroimagen [3]. Con el tiempo, Lasso se consolidó como un pilar del aprendizaje estadístico, originando variantes como el "Elastic Net", que combina la penalización ℓ_1 de Lasso con la ℓ_2, ofreciendo un mejor manejo de variables altamente correlacionadas [8].

El algoritmo KNN, por su parte, tiene sus raíces en los años 1950, cuando Evelyn Fix y Joseph Hodges lo propusieron como un método no paramétrico de clasificación basado en la proximidad entre puntos de datos. En 1967, Thomas Cover y Peter Hart formalizaron su aplicación en problemas de reconocimiento de patrones, demostrando que, con un número suficiente de datos, KNN puede aproximarse al clasificador Bayesiano óptimo [14]. A diferencia de los métodos de regresión, KNN no impone supuestos sobre la forma funcional de los datos, lo que le permite adaptarse a distribuciones complejas y no lineales.

Desde los años 2000, Lasso ha sido extendido a modelos de clasificación como la regresión logística penalizada con ℓ_1, siendo ampliamente utilizado en problemas de respuesta categórica. Gracias a algoritmos computacionalmente eficientes, como los métodos de coordenadas cíclicas, Lasso ha demostrado ser capaz de realizar selección de variables incluso con datos de alta dimensionalidad, identificando solo aquellas variables relevantes para la probabilidad de clase. Variantes como Elastic Net han encontrado aplicaciones destacadas en neurociencia, particularmente en la predicción de diagnósticos clínicos a partir de imágenes cerebrales, clasificación de estados mentales y análisis de conectividad funcional [8, 9].

De forma complementaria, KNN también ha evolucionado hacia aplicaciones neurocientíficas. Inicialmente utilizado como método de referencia en clasificación biomédica, ha sido empleado en tareas como la detección de patrones en señales EEG, la identificación de estados cognitivos mediante fMRI y la clasificación de pacientes con trastornos neurológicos. Su naturaleza no paramétrica lo hace especialmente útil en conjuntos de datos clínicos y neuropsicológicos, donde las relaciones entre variables suelen ser no lineales y de alta complejidad [14].

En las últimas décadas, ambas metodologías han ganado relevancia en neurociencia, un campo caracterizado por el aumento exponencial en la cantidad y complejidad de datos, como en estudios de resonancia magnética funcional (fMRI), electroencefalografía (EEG) o análisis de conectividad cerebral. La regresión Lasso ha sido utilizada para identificar regiones cerebrales clave en tareas cognitivas y emocionales, así como para la búsqueda de biomarcadores predictivos en trastornos neurológicos como el TDAH, el autismo o el Alzheimer. KNN, por su parte, se mantiene como un clasificador sencillo pero eficaz, sirviendo como referencia para validar modelos más sofisticados como SVM o redes neuronales.

Actualmente, los modelos de regularización se integran con otras técnicas de aprendizaje automático como SVM, redes neuronales y métodos de ensamble para mejorar el diagnóstico automatizado, la predicción de la progresión clínica y la segmentación de estructuras cerebrales. Variantes como Elastic Net o Group Lasso permiten, además, incorporar estructuras jerárquicas o espaciales características de los datos cerebrales.

La combinación de Lasso y KNN ha fortalecido el análisis neurocientífico, ofreciendo enfoques complementarios para la selección de variables y la clasificación de patrones cerebrales. Lasso aporta interpretabilidad y reducción de dimensionalidad al identificar biomarcadores relevantes, mientras que KNN brinda una clasificación flexible y no paramétrica basada en proximidad. En conjunto, estas técnicas impulsan el desarrollo de una neurociencia computacional predictiva, capaz de vincular datos cerebrales con conductas, síntomas clínicos o estados cognitivos de forma precisa, robusta y reproducible. Este panorama justifica la elección de Lasso y KNN como técnicas centrales en el presente trabajo.

3. Modelado del Problema de Regresión y Regularización Lasso

Para comenzar con los fundamentos del problema de regresión, es esencial comprender el contexto clínico en el que se enmarca esta investigación. El Trastorno por Déficit de Atención e Hiperactividad (TDAH) es una condición neurobiológica del desarrollo que se manifiesta a través de síntomas como inatención, hiperactividad e impulsividad. Aunque suele diagnosticarse en la infancia, también persiste en la adultez, afectando el desempeño académico, laboral y social. Su diagnóstico,

especialmente en adultos, puede ser complejo debido a la diversidad de síntomas y la presencia de trastornos comórbidos [2].

Ante esta dificultad diagnóstica, las herramientas de la estadística y la ciencia de datos se presentan como un apoyo valioso. En particular, esta investigación se enfoca en los fundamentos del problema de regresión, utilizando técnicas como la regresión logística penalizada con Lasso [1]. Esta metodología nos permite modelar la probabilidad de que un individuo presente TDAH a partir de un conjunto de variables predictoras, facilitando la identificación de factores clave y mejorando la precisión del diagnóstico.

3.1. Fundamentos del Problema de Regresión.

La regresión lineal es una de las técnicas más utilizadas en estadística y aprendizaje automático supervisado. Su objetivo central es modelar la relación entre una variable de salida (respuesta) Y y un conjunto de variables de entrada (predictoras) $X_1, X_2, \ldots, X_p$, a fin de predecir o inferir el comportamiento de Y a partir de nuevas observaciones de X [1].

El modelo más básico asume una relación lineal entre la variable respuesta y los predictores, y se expresa de la siguiente forma:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon$$

donde ε representa el término de error aleatorio, el cual se asume con distribución normal de media cero y varianza constante, es decir, $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.

Para cada observación i en el conjunto de datos, el modelo se escribe explícitamente como:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i.$$

El modelo se ajusta generalmente mediante el método de mínimos cuadrados, el cual consiste en encontrar los coeficientes $\beta_0, \beta_1, \ldots, \beta_p$ que minimizan la suma de los errores cuadrados entre los valores observados y los predichos por el modelo.

El método de mínimos cuadrados es una técnica fundamental para estimar los parámetros de un modelo lineal. Su propósito aqu í es encontrar la combinación de coeficientes que minimice la discrepancia entre los valores observados y los valores predichos por el modelo [1].

Supongamos un conjunto de datos de n observaciones (x_i, y_i), donde se desea modelar la relación entre la variable respuesta y y las variables predictoras $x_1, x_2, \ldots, x_p$ mediante el siguiente modelo lineal:

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}$$

El método de mínimos cuadrados busca los valores de $\beta_0, \beta_1, \ldots, \beta_p$ que minimicen la suma de los errores cuadrados entre los valores reales y_i y los predichos $\hat{y}_i$:

$$\text{Suma de errores cuadrados} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2$$

La regresión lineal tiene varias ventajas que la hacen muy útil. Es fácil de aplicar y de interpretar, ya que nos permite entender cómo cada variable afecta el resultado. También es rápida desde el punto de vista computacional, y si se cumplen ciertas condiciones como la linealidad o la independencia, sus estimaciones son bastante confiables. Por eso se usa mucho, sobre todo como primer paso cuando se está explorando un conjunto de datos o se quiere hacer una predicción sencilla [1].

Pero también tiene sus limitaciones. Por ejemplo, asume que la relación entre las variables es lineal, lo que no siempre ocurre en la práctica. Además, si hay valores atípicos, estos pueden afectar mucho los resultados. Otra dificultad es cuando las variables están correlacionadas entre sí, ya que eso puede hacer que los coeficientes sean inestables. Y si hay muchas variables en comparación con el número de datos, el modelo puede ajustarse demasiado y no funcionar bien con nuevos datos [1].

La Regresión logística penalizada es una técnica utilizada para modelar una variable de respuesta binaria $Y \in \{0, 1\}$ en función de un conjunto de variables predictoras $X_1, X_2, \ldots, X_p$. A diferencia de la regresión lineal, esta técnica utiliza la función logística para garantizar que las predicciones se encuentren en el intervalo $[0, 1]$, correspondiente a probabilidades [1]:

$$p(X) = \Pr(Y = 1 \mid X) = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}{1 + e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}$$

A continuación se muestra una comparación visual entre la regresión lineal y la regresión logística, que ilustra cómo difieren en su comportamiento frente a variables predictoras continuas:

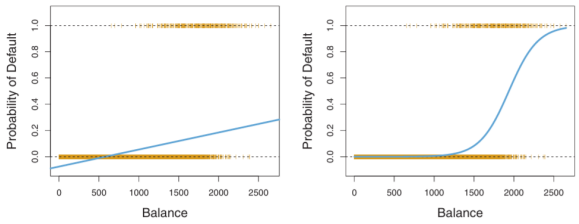


FIGURA 1. Comparación entre regresión lineal y regresión logística. La regresión lineal puede predecir cualquier valor real, mientras que la logística produce probabilidades entre 0 y 1 mediante una curva sigmoide [1].

Sin embargo, cuando el número de predictores p es grande o existe multicolinealidad entre ellos, la estimación por máxima verosimilitud puede volverse inestable o poco interpretable. Para abordar estos problemas, se incorpora un término de penalización a la función objetivo, lo que da lugar a la regresión logística penalizada.

Una forma común de penalización es Lasso, el cual añade una penalización basada en la norma ℓ_1 de los coeficientes:

$$\ell_{\text{pen}}(\beta) = \ell(\beta) - \lambda \sum_{j=1}^{p} |\beta_j|$$

donde $\lambda \geq 0$ es el parámetro de regularización que controla la magnitud de la penalización. Este término fuerza a algunos coeficientes a ser exactamente cero,

realizando así una selección automática de variables, Lasso puede generar modelos más parciales y fáciles de interpretar al eliminar predictores irrelevantes del modelo [1].

Además, desde una perspectiva bayesiana, Lasso equivale a asumir una distribución a priori de tipo Laplace (doblemente exponencial) para los coeficientes del modelo, lo cual favorece soluciones dispersas o esparsas [3].Esta técnica es particularmente útil en contextos donde se desea mejorar la interpretabilidad del modelo, controlar el sobreajuste, o cuando el número de predictores supera al número de observaciones ($p > n$).

El método Lasso, propuesto por Tibshirani en 1996, es una técnica de regularización utilizada para mejorar la precisión predictiva de los modelos lineales y reducir su complejidad al penalizar los coeficientes de regresión. A diferencia de la regresión ordinaria por mínimos cuadrados, Lasso introduce una penalización basada en la norma ℓ_1 de los coeficientes, lo que favorece soluciones dispersas, en las que algunos coeficientes pueden ser exactamente cero [3].

Definition 3.1. En el caso de un modelo de regresión lineal con predictores $X_1, X_2, \ldots, X_p$, el modelo Lasso busca los coeficientes β que minimicen la siguiente función objetivo:

$$\min_{\beta_0,\beta} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

donde $\lambda \geq 0$ es un parámetro de regularización que controla la magnitud de la penalización. Para valores grandes de λ, el modelo tiende a hacer cero muchos coeficientes β_j, lo que equivale a eliminar las variables correspondientes del modelo. Esta propiedad convierte al Lasso en una herramienta útil no solo para la regularización, sino también para la selección automática de variables [1].

Desde un enfoque estadístico, el método Lasso presenta propiedades fundamentales que explican su popularidad y utilidad en contextos de alta dimensión. Primero, destaca su capacidad para realizar selección de variables de manera consistente. Bajo ciertas condiciones teóricas, como la condición de irrepresentabilidad o la incoherencia de señales, Lasso es capaz de identificar correctamente el conjunto de variables verdaderamente relevantes cuando el tamaño de la muestra es suficientemente grande. Esta propiedad es crucial en investigaciones donde no solo se busca predecir, sino también comprender la estructura subyacente de los datos [3].

Una segunda propiedad importante es su tendencia a generar soluciones esparsas, es decir, modelos en los que muchos coeficientes estimados son exactamente cero. Esto es resultado directo de la penalización basada en la norma ℓ_1, que favorece coeficientes pequeños y tiende a eliminar aquellos predictores que no aportan significativamente a la variable respuesta. Esta característica permite simplificar el modelo sin perder precisión predictiva.

Además, el Lasso introduce un sesgo deliberado en la estimación de los coeficientes, lo que a su vez permite una reducción significativa de la varianza. Esta

reducción en la varianza mejora la capacidad de generalización del modelo frente a nuevos datos, particularmente cuando se trabaja con muestras pequeñas o conjuntos de datos ruidosos [3].

Sin embargo, esta técnica también tiene limitaciones. Una de ellas es su inestabilidad ante la presencia de predictores altamente correlacionados. En tales casos, Lasso tiende a seleccionar solo una de las variables correlacionadas y a eliminar las demás, lo cual puede llevar a modelos inestables si pequeñas variaciones en los datos generan cambios en las variables seleccionadas. Para resolver esta situación, una alternativa recomendable es el uso del método Elastic Net, que combina penalizaciones ℓ_1 y ℓ_2[8].

Por último, desde una perspectiva bayesiana, Lasso puede interpretarse como un modelo en el que se impone una distribución a priori de tipo Laplace (también conocida como doblemente exponencial) sobre los coeficientes del modelo. Esta suposición refleja la creencia de que muchos coeficientes son cero o muy cercanos a cero, favoreciendo así la construcción de modelos más parsimoniosos y fáciles de interpretar [3].

Estas propiedades convierten al Lasso en una herramienta poderosa para abordar problemas donde se busca no solo predecir con precisión, sino también entender qué variables tienen mayor relevancia dentro del fenómeno estudiado.

A diferencia de k-Vecinos Más Cercanos, que es un método no paramétrico basado en la proximidad entre las observaciones y no ajusta coeficientes β_j, Lasso es un modelo paramétrico que utiliza una penalización basada en la norma ℓ_1, lo que le permite anular algunos coeficientes completamente [1]. La siguiente tabla resume las diferencias clave:

Método	Enfoque	¿Realiza selección de variables?
KNN	Basado en distancias entre observaciones	No
Lasso	Penalización $\lambda \sum \lvert\beta_j\rvert$	Sí

Cuadro 1. Comparación entre KNN y Lasso.

La predicción en KNN se realiza a partir de los k vecinos más cercanos de una observación x_0, calculados mediante la distancia euclidiana:

$$d(x_0, x_i) = \sqrt{\sum_{j=1}^{p} (x_{0j} - x_{ij})^2},$$

donde $x_i \in \mathbb{R}^p$ representa cada observación del conjunto de datos. En problemas de clasificación, la clase predicha es:

$$\hat{C}(x_0) = \arg\max_{c \in \mathcal{C}} \sum_{i \in \mathcal{N}_k(x_0)} \mathbf{1}(y_i = c),$$

mientras que en regresión se utiliza el promedio:

$$\hat{f}(x_0) = \frac{1}{k} \sum_{i \in \mathcal{N}_k(x_0)} y_i.$$

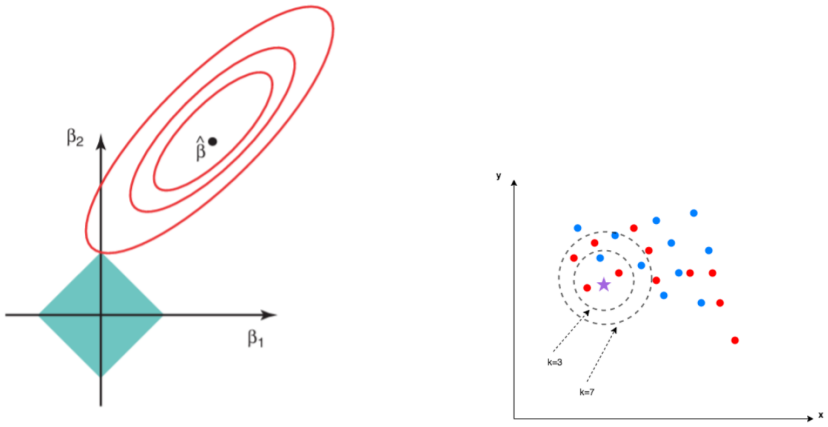


FIGURA 2. Comparación conceptual entre Lasso y KNN. Lasso ajusta un modelo lineal con regularización, mientras que KNN basa su predicción en los puntos de entrenamiento más cercanos [1].

Aunque Lasso fue desarrollado inicialmente para regresión lineal, puede extenderse a otros modelos como la regresión logística. En este caso, se penaliza la función de log-verosimilitud en lugar del error cuadrático:

$$\ell_{\text{pen}}(\beta) = \ell(\beta) - \lambda \sum_{j=1}^{p} |\beta_j|$$

Este tipo de penalización es útil en contextos de alta dimensión ($p \gg n$) como en neurociencia o genética, ya que permite construir modelos más simples, interpretables y con mejor capacidad de generalización [1]. Lasso puede realizar simultáneamente ajuste del modelo y selección de predictores, siendo una herramienta poderosa en problemas donde se necesita entender qué variables son más relevantes para la respuesta.

El parámetro de regularización λ juega un rol fundamental en el rendimiento del modelo Lasso, ya que controla el grado de penalización aplicado a los coeficientes. Elegir un valor de λ demasiado grande puede llevar a un modelo subajustado, mientras que un valor muy pequeño puede resultar en sobreajuste. Para seleccionar un valor óptimo de λ, se utiliza comúnmente el método de validación cruzada k-fold [1, 12].

Este procedimiento consiste en dividir el conjunto de datos en k subconjuntos (o *folds*) de tamaño similar. El modelo se entrena k veces, cada vez utilizando $k-1$ subconjuntos como datos de entrenamiento y el restante como conjunto de validación. Para cada valor candidato de λ, se calcula un error promedio sobre las k iteraciones, y se selecciona el valor que minimiza dicho error. Este enfoque no solo

permite encontrar el λ que mejor generaliza a datos no vistos, sino que también reduce el riesgo de seleccionar un modelo inestable [1, 12].

En la práctica, este proceso se implementa generando una secuencia de valores de λ, ajustando el modelo para cada uno y evaluando su rendimiento mediante una métrica como el error cuadrático medio. El valor de λ asociado al menor error promedio se considera óptimo y se utiliza para construir el modelo final. Esta técnica es especialmente útil en contextos con alta dimensionalidad o cuando se desea maximizar la capacidad de generalización del modelo.

El Trastorno por Déficit de Atención e Hiperactividad (TDAH) es una condición neuropsiquiátrica que afecta tanto a niños como a adultos, caracterizándose por síntomas persistentes de inatención, impulsividad e hiperactividad. En el ámbito clínico, el diagnóstico del TDAH representa un desafío debido a la heterogeneidad de sus manifestaciones y la frecuente coexistencia con otros trastornos del neurodesarrollo o del estado de ánimo. En consecuencia, los métodos tradicionales basados únicamente en observaciones clínicas y entrevistas pueden resultar insuficientes para capturar la complejidad del cuadro [11].

Frente a este escenario, el uso de herramientas estadísticas y de aprendizaje automático ha ganado relevancia, permitiendo integrar múltiples fuentes de datos —como escalas psicométricas, registros neuropsicológicos, patrones de actividad cerebral, variables demográficas o datos provenientes de sensores portátiles— con el objetivo de construir modelos predictivos más robustos. En este contexto, la regresión penalizada mediante el método Lasso ha demostrado ser una alternativa particularmente valiosa [4, 2].

Una de las principales ventajas del uso de Lasso radica en su capacidad para realizar simultáneamente la selección de variables y la estimación de parámetros. Esto resulta crucial en el análisis de datos clínicos de TDAH, donde se dispone de un gran número de posibles predictores pero un tamaño de muestra relativamente pequeño. Lasso permite identificar un subconjunto reducido de características relevantes que contribuyen significativamente a la predicción del diagnóstico, favoreciendo modelos más parcimoniosos y clínicamente interpretables [2].

Además, la esparsidad inducida por la penalización ℓ_1 facilita la eliminación automática de variables irrelevantes o redundantes, lo que mejora la generalización del modelo y reduce el riesgo de sobreajuste. Esto es especialmente importante cuando se trabaja con biomarcadores fisiológicos o variables derivadas de registros de actividad cerebral, como los datos de electroencefalografía (EEG) o imágenes por resonancia magnética funcional (fMRI), que suelen tener alta dimensionalidad y correlación entre predictores [13, 4].

Estudios recientes han empleado el método Lasso para discriminar entre pacientes con TDAH y grupos control, alcanzando desempeños competitivos en términos de sensibilidad y especificidad. Por ejemplo, se ha utilizado para extraer características relevantes a partir de la señal EEG en niños con TDAH, obteniendo modelos con un número reducido de variables pero gran capacidad de clasificación. Asimismo, en investigaciones con adultos, Lasso ha sido empleado para evaluar la importancia relativa de escalas como el ASRS, el WURS y diversas pruebas neurocognitivas en

la predicción del diagnóstico [13].

En conjunto, el enfoque Lasso representa una contribución significativa al análisis moderno de datos clínicos en TDAH, al permitir el desarrollo de modelos predictivos eficientes, estables e interpretables que integran múltiples dimensiones de información del paciente.

3.2. Ejemplo ilustrativo de Regresión Logística Penalizada con Lasso. Para ilustrar el funcionamiento del método Lasso en el contexto de una variable respuesta binaria, se generaron datos simulados con $n = 100$ observaciones y $p = 10$ variables predictoras. Sólo cinco variables $(X_1, X_3, X_7, X_9, X_{10})$ tienen un efecto real sobre la probabilidad de la respuesta, modelada mediante la función logística.

Se ajustó un modelo de regresión logística penalizada con Lasso utilizando validación cruzada k-fold para seleccionar el parámetro de regularización λ que minimiza el error de validación.

Con el λ óptimo seleccionado, los coeficientes estimados se muestran a continuación. El método Lasso realiza una selección automática de variables, anulando los coeficientes de aquellas que no contribuyen significativamente al modelo:

La siguiente tabla muestra los coeficientes estimados por el modelo Lasso para las 10 variables predictoras en el ejemplo simulado. Se observa que Lasso ha asignado coeficientes diferentes de cero únicamente a las variables relevantes X_1, X_3 , X_7 , X_9 y X_{10} mientras que el resto tienen coeficientes exactamente cero, demostrando su capacidad de selección automática:

Variable	Coeficiente Lasso
X1	0.97668711
X2	0.00000000
X3	-0.57879004
X4	0.00000000
X5	0.00000000
X6	0.00000000
X7	-0.13923859
X8	0.00000000
X9	0.72668586
X10	0.02977117

CUADRO 2. Coeficientes estimados por el modelo Lasso. Las variables con coeficiente cero fueron eliminadas durante el proceso de regularización.

Esta tabla ilustra claramente cómo el método Lasso consigue reducir la complejidad del modelo, manteniendo solo las variables más relevantes para la predicción.

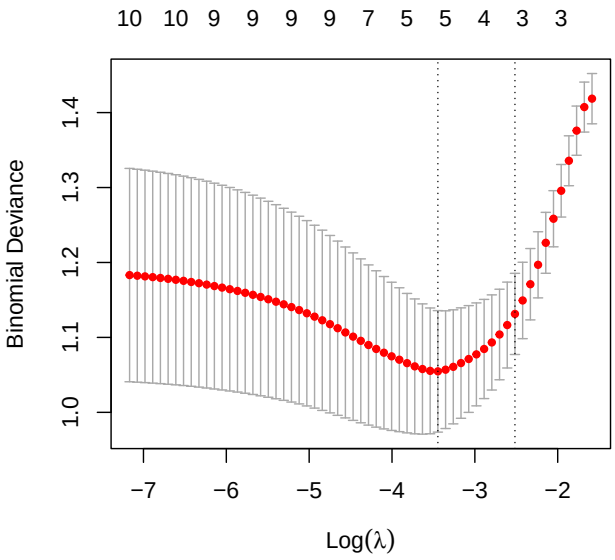


FIGURA 3. Curva de validación cruzada para Lasso. El valor de λ que minimiza el error se selecciona para ajustar el modelo final.

Este ejemplo demuestra cómo Lasso permite construir modelos parsimoniosos y con buena capacidad predictiva, especialmente útiles en problemas con alta dimensionalidad y cuando se busca interpretar qué variables son más relevantes.

Después de haber explicado los aspectos teóricos y metodológicos de esta investigación, en esta parte se presentan los experimentos realizados para evaluar el rendimiento del modelo de regresión Lasso en el diagnóstico de TDAH. Aquí se explica cómo se preparon los datos, qué criterios se utilizaron para la validación del modelo y qué métricas se tomaron en cuenta para analizar los resultados.El objetivo es presentar de manera clara la aplicación del modelo y evaluar su desempeño frente al algoritmo de k vecinos más cercanos (KNN), considerando métricas de precisión y su aplicabilidad en contextos reales.

3.3. Descripción de Datos.

Los resultados de esta sección están basados en [13]. En esta investigación se empleó el conjunto de datos HYPERAKTIV, disponible públicamente en la plataforma Kaggle. Este dataset está conformado por varios archivos en formato .xlsx, organizados en directorios que contienen información sobre actividad motora, frecuencia cardíaca, respuestas neuropsicológicas y variables clínicas relacionadas con el diagnóstico de TDAH en adultos.

La muestra incluyó 103 participantes: 51 con diagnóstico de TDAH y 52 que actuaron como controles clínicos. Las variables recopiladas abarcan sexo, edad agrupada, síntomas clínicos (ASRS, WURS, MADRS, HADS), presencia de trastornos comórbidos (como ansiedad o depresión) y condición de medicación. Asimismo, se integraron resultados del test neuropsicológico computarizado CPT-II y registros

de señales fisiológicas obtenidas mediante dispositivos de muñeca (actigrafía) y de pecho (ECG).

Para su análisis, se integraron distintos archivos Excel en un único dataset combinado, que quedó conformado por 85 observaciones y 759 variables. La variable objetivo indica si una persona presenta o no TDAH (1 para diagnóstico positivo, 0 para negativo). Entre las variables predictoras más relevantes se encuentran puntuaciones en escalas como WURS y ASRS, así como características demográficas y clínicas.

Antes de entrenar el modelo de regresión Lasso, se aplicó un proceso de preprocesamiento que incluyó la eliminación de valores faltantes, la codificación de variables categóricas cuando fue necesario, y la normalización de las variables para asegurar un entrenamiento más robusto y preciso.

Este conjunto de datos resulta especialmente valioso porque combina medidas objetivas (como la actividad física y la variabilidad de la frecuencia cardíaca) con evaluaciones clínicas subjetivas, lo que abre la posibilidad de explorar enfoques automáticos más precisos en el diagnóstico del TDAH.

Dado que el conjunto de datos HYPERAKTIV está distribuido en varios archivos .xlsx correspondientes a distintas fuentes de información (actividad motora, frecuencia cardíaca, datos clínicos, pruebas neuropsicológicas, etc.), fue necesario realizar un proceso de integración para crear una sola base de datos unificada. Para ello, se utilizó la ID única de cada participante como clave principal para unir los diferentes archivos, asegurando que toda la información correspondiente a un mismo paciente quedara correctamente asociada.

Durante este proceso también se llevó a cabo una etapa de limpieza de datos, en la que se manejaron los valores faltantes. Para las variables numéricas, se optó por rellenar los datos faltantes utilizando la media de cada variable. Esta técnica permitió mantener el tamaño de la muestra sin introducir sesgos extremos. Además, se eliminaron columnas irrelevantes o redundantes y se estandarizaron nombres de variables para facilitar su uso posterior.

Como resultado, se obtuvo un único dataset consolidado, limpio y listo para su análisis, el cual sirvió como base para la aplicación de los modelos de aprendizaje automático, en particular la regresión Lasso.

A continuación se describirán algunas de las variables que están contenidas en el dataset , y que importancia tienen para la predicción del TDAH.

Entre las variables más relevantes se encuentran: el diagnóstico final (ADHD), escalas clínicas como WURS y ASRS, evaluaciones de ansiedad y depresión (MADRS, HADS), variables demográficas como edad y sexo, y los resultados del test CPT-II.

Además, el dataset incluye numerosas variables generadas automáticamente a partir de los registros de actividad motora (ACC). Estas variables provienen de transformaciones estadísticas y espectrales, como la Transformada Rápida de Fourier (FFT), y reflejan patrones cíclicos, variaciones de intensidad y dinámicas del movimiento a lo largo del tiempo.

Dado que el dataset final contenía 759 variables, muchas de ellas altamente correlacionadas, se utilizó regresión Lasso para realizar una selección automática de las variables más informativas y reducir el riesgo de sobreajuste en los modelos predictivos.

El método Lasso es una técnica de regresión regularizada que permite realizar selección de variables y reducción de la complejidad del modelo , siendo útil especificamente cuando se trabaja con un número elevado de predictores. Este enfoque incorpora un término de penalización en la estimación de los coeficientes , el cual consiste en la suma de los valores absolutos de dichos coeficientes:

$$\min_{\beta_0,\beta} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

donde:

- y_i es la respuesta observada para la i.
- x_{ij} El valor de la variable j en la observación i.
- β_j el coeficiente asociado a la variable predictora x_{ij}.
- β_0 el término independiente o intercepto del modelo.
- n número total de observaciones en el conjunto de datos.
- p número total de variables predictoras.
- λ Parámetro de regularización que controla la fuerza de la penalización sobre los coeficientes.

3.4. Preparación de datos. El conjunto de datos utilizado consta de 85 observaciones, divididas en dos subconjuntos: 80 % para entrenamiento y 20 % para prueba. Esta división se realizó con el objetivo de evaluar la capacidad de generalización de los modelos.

Se emplearon variables predictoras relacionadas con la sintomatología de TDAH, y la variable respuesta (y) es binaria, donde:

- $y = 1$: participante con diagnóstico de TDAH.
- $y = 0$: participante sin diagnóstico de TDAH.

Para la fase de modelado se utilizó la regresión Lasso, una técnica de regularización que incorpora una penalización L_1 en la estimación de los coeficientes. La formulación matemática es:

$$\min_{\beta_0,\beta} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\},$$

donde λ es el parámetro de regularización que controla la magnitud de la penalización. Este método no solo mejora la generalización del modelo, sino que también permite la selección de variables al forzar a ciertos coeficientes β_j a ser exactamente cero.

El valor óptimo de λ se determinó utilizando validación cruzada k-fold, mediante la función cv.glmnet en R. La Figura 4 muestra la curva de validación cruzada, mientras que la Figura 4 presenta la evolución de los coeficientes en función de $\log(\lambda)$.

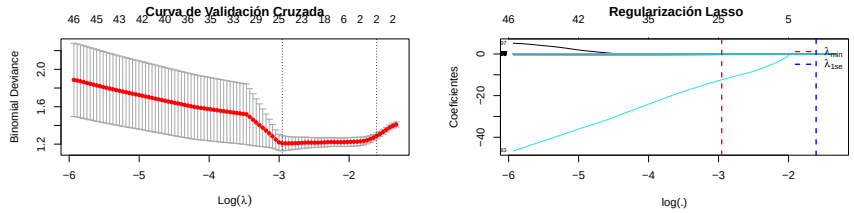


FIGURA 4. Curva de validación cruzada para selección de λ Y Evolución de los coeficientes con regularización Lasso.

En las Tabla 3.4 y Tabla 3.4 se presentan las variables seleccionadas por el modelo Lasso, junto con una breve descripción y el signo del coeficiente estimado, lo que permite interpretar su relación con la variable respuesta.

Variable	Descripción	Signo
BIPOLAR	Síntomas asociados al trastorno bipolar.	−
ANXIETY	Índice o escala de ansiedad.	−
WURS	Wender Utah Rating Scale (TDAH en infancia).	+
ASRS	Adult ADHD Self-Report Scale (TDAH en adultos).	+
HADS_D	Subescala de depresión (HADS).	−

CUADRO 3. Variables clínicas y psicométricas seleccionadas por Lasso.

Variable	Descripción	Signo
ACC__number_peaks__n_50	Número de picos en la señal (ventanas de tamaño 50).	−
ACC__ar_coefficient__coeff_4__k_10	Coeficiente 4 de un modelo autorregresivo (orden 10).	−
ACC__cwt_coefficients__coeff_9__w_2__widths_.2..5..10..20.	Coeficiente 9 de la Transformada Wavelet Continua.	+
ACC__fft_coefficient__attr_.real.__coeff_21	Parte real del coeficiente 21 de la FFT.	+
ACC__fft_coefficient__attr_.angle.__coeff_30	Fase del coeficiente 30 de la FFT.	−

CUADRO 4. Variables de series temporales (ACC__).

3.5. Evaluación del Modelo. Se evaluó el modelo Lasso utilizando el conjunto de prueba, considerando las métricas: Accuracy, Sensibilidad (Recall), Especificidad y el área bajo la curva ROC (AUC).

La Figura 5 presenta la curva ROC obtenida, mientras que la Figura 5 muestra la curva Precision-Recall (PRC).

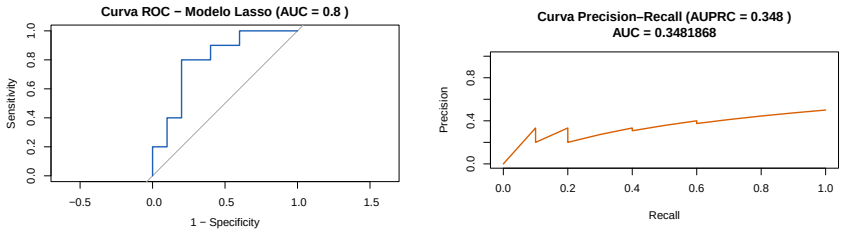


FIGURA 5. Curva ROC con AUC del modelo Lasso y Curva Precision-Recall con AUPRC.

La matriz de confusión obtenida para el conjunto de prueba fue:

	Referencia		Total
Predicción	0	1	
0	8	3	11
1	2	7	9
Total	10	10	20

CUADRO 5. Matriz de confusión del modelo Lasso.

Métricas de evaluación:

- Accuracy: 0.75 (IC 95 %: 0.509 – 0.9134).
- Kappa: 0.50.
- Sensibilidad (TPR): 0.70.
- Especificidad (TNR): 0.80.
- Valor predictivo positivo (PPV): 0.7778.
- Valor predictivo negativo (NPV): 0.7273.
- Balanced Accuracy: 0.75.
- Prueba de McNemar: $p = 1.00000$.

Para contrastar el rendimiento del modelo Lasso, se utilizó el algoritmo de K-vecinos más cercanos (KNN). Este método asigna la clase de una observación de prueba en función de las k observaciones de entrenamiento más cercanas (en términos de una métrica de distancia, comúnmente Euclidiana). Formalmente, para un punto x, la predicción se define como:

$$\hat{y}(x) = \arg \max_{c \in \{0,1\}} \sum_{i \in \mathcal{N}_k(x)} \mathbf{1}(y_i = c),$$

donde $\mathcal{N}_k(x)$ es el conjunto de los k vecinos más cercanos de x.

3.6. Resultados del modelo K-Vecinos Más Cercanos (KNN). Se implementó un modelo KNN con $k = 8$ vecinos, utilizando las mismas variables predictoras y división de datos que en el modelo Lasso.

La matriz de confusión para el conjunto de prueba es:

Predicción	Referencia 0	1	Total
0	4	2	6
1	6	8	14
Total	10	10	20

CUADRO 6. Matriz de confusión del modelo KNN con $k = 8$.

Las métricas de evaluación calculadas para KNN fueron:

- Accuracy: 0.6 (IC 95 %: 0.509 – 0.9134).
- Sensibilidad (Recall): 0.9000.
- Especificidad: 0.6000.
- Valor predictivo positivo (PPV): 0.6923
- Valor predictivo negativo (NPV): 0.8571
- Balanced Accuracy: 0.7500
- Prueba de Mcnemar's: 0.37109

Las curvas ROC y Precision-Recall se presentan en la Figura 7, respectivamente.

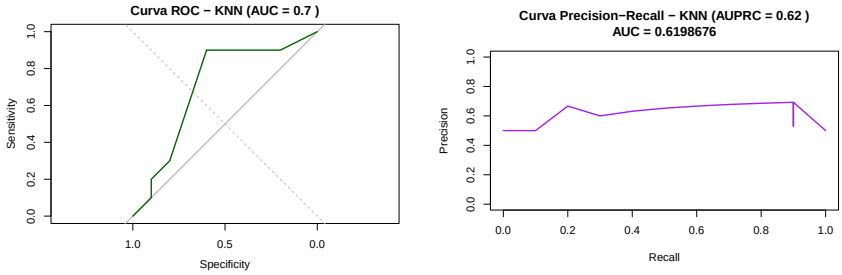


FIGURA 6. Curva ROC del modelo KNN y curva Precision-Recall del modelo KNN.

Se compararon las métricas de ambos modelos (Lasso vs. KNN) y sus respectivas curvas ROC, como se muestra en la Figura 7.

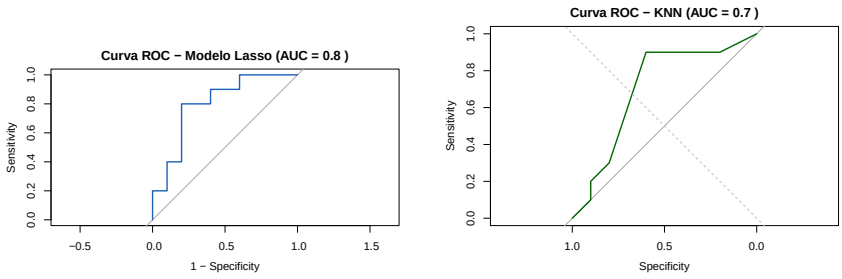


FIGURA 7. Comparación de las curvas ROC de ambos modelos.

3.7. Discusión. Los resultados obtenidos en este estudio permiten realizar una reflexión crítica sobre el uso de técnicas de aprendizaje automático para el diagnóstico del Trastorno por Déficit de Atención e Hiperactividad (TDAH), específicamente

mediante los modelos Lasso y K-vecinos más cercanos (KNN). Ambos enfoques demostraron ser válidos para el análisis del conjunto de datos clínicos utilizado, aunque con diferencias importantes en cuanto a interpretabilidad, precisión y aplicabilidad clínica.

El modelo de regresión Lasso logró un desempeño equilibrado entre sensibilidad (0.70) y especificidad (0.80), además de un área bajo la curva ROC (AUC) de 0.80. Este rendimiento, combinado con su capacidad de realizar selección automática de variables, lo convierte en una herramienta ideal para entornos clínicos donde es esencial comprender cuáles variables influyen en el diagnóstico. En particular, Lasso permitió identificar variables psicométricas clave como WURS, ASRS y HADS-D, así como características fisiológicas derivadas de transformaciones espectrales (FFT, CWT), lo que respalda el enfoque multimodal para el diagnóstico del TDAH.

En contraste, el algoritmo KNN obtuvo una mayor sensibilidad (0.90), lo que indica una mayor capacidad para detectar correctamente a los pacientes con TDAH. Sin embargo, su especificidad fue menor (0.60), lo que implica un mayor número de falsos positivos. Esta diferencia resalta el carácter no paramétrico y flexible de KNN, pero también sus limitaciones en términos de interpretabilidad, dado que no ofrece información directa sobre la importancia relativa de las variables utilizadas.

La comparación entre ambos modelos sugiere que Lasso es más apropiado cuando se requiere un modelo explicativo que facilite la toma de decisiones clínicas, mientras que KNN podría ser útil como herramienta auxiliar en tareas de detección temprana, donde se priorice la sensibilidad sobre la especificidad. La integración de ambos enfoques podría ofrecer beneficios complementarios, combinando la solidez interpretativa de Lasso con la flexibilidad predictiva de KNN.

Además, la efectividad de Lasso para reducir el número de predictores y evitar el sobreajuste resulta especialmente relevante en contextos clínicos con datos de alta dimensionalidad y tamaño muestral limitado, como ocurre frecuentemente en estudios de salud mental. Este resultado es coherente con investigaciones previas que destacan la utilidad del Lasso en neurociencia computacional y psiquiatría clínica.

En conjunto, los hallazgos de este estudio evidencian el potencial de las técnicas estadísticas penalizadas y de clasificación no paramétrica para contribuir al diagnóstico del TDAH desde un enfoque basado en datos. Esta aproximación no solo mejora el rendimiento predictivo, sino que también promueve una medicina más cuantitativa, reproducible e interpretable.

4. Conclusiones

1. La regresión Lasso permitió una selección automática y efectiva de variables relevantes, mejorando la interpretabilidad del modelo y reduciendo el sobreajuste en un conjunto de datos altamente dimensional.
2. El modelo Lasso mostró un desempeño robusto (AUC = 0.80, precisión = 75 %), mientras que KNN destacó por su alta sensibilidad (90 %), aunque con menor especificidad. Esta comparación resalta la necesidad de elegir el modelo según el objetivo clínico.
3. Lasso facilitó la construcción de un modelo más transparente, identificando variables clave para el diagnóstico de TDAH, lo cual es esencial en contextos clínicos donde la explicabilidad es crucial.

4. La integración de escalas clínicas (ASRS, WURS, HADS, MADRS) y variables fisiológicas (como ACC y FFT) potenció el enfoque multimodal, combinando datos objetivos y subjetivos en la predicción del diagnóstico.
5. La validación cruzada k-fold fue fundamental para ajustar el parámetro de regularización λ en Lasso, garantizando estabilidad del modelo y evitando el sobreajuste.
6. Este estudio evidencia cómo la ciencia de datos y la neurociencia computacional pueden complementar el juicio clínico, fortaleciendo el diagnóstico en salud mental mediante modelos predictivos interpretables.
7. La metodología puede extenderse a otros trastornos neuropsiquiátricos complejos, como depresión, autismo o Alzheimer, donde también se requiere una integración de múltiples fuentes de datos para un diagnóstico preciso.
8. Se destaca el valor clínico de combinar Lasso y KNN como herramientas complementarias: Lasso por su estructura simple e interpretabilidad, y KNN por su alta sensibilidad, lo que constituye una alternativa prometedora para la medicina de precisión.

Referencias

1. G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*, Springer, New York, 2013.
2. S. Cortese, P. Asherson, J. Buitelaar, J. M. Faraone, T. Banaschewski, et al., "The Neurobiology and Genetics of Attention-Deficit/Hyperactivity Disorder (ADHD): What Next?," *European Neuropsychopharmacology*, vol. 22, no. 2, pp. 71–81, 2012.
3. Tibshirani, R. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B*, 58(1), 267–288, 1996.
4. Zhou, F., Liu, J., & Li, H. Sparse regression models for identifying behavioral predictors of ADHD. *Journal of Attention Disorders*, 23(9), 969–981, 2019.
5. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed., American Psychiatric Publishing, Arlington, VA, 2013.
6. A. Abraham, B. Thirion, and G. Varoquaux, *Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example*, NeuroImage **147** (2017), 736–745.
7. A. Qiu, J. T. Devlin, and D. Bzdok, *The promise of LASSO for psychiatry*, Biological Psychiatry: Cognitive Neuroscience and Neuroimaging **5** (2020), no. 5, 355–357.
8. Zou, H., & Hastie, T. *Regularization and variable selection via the elastic net.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301–320, 2005.
9. Ryali, S., Supekar, K., Abrams, D. A., & Menon, V. *Sparse logistic regression for whole-brain classification of fMRI data.* NeuroImage, 51(2), 752–764, 2010.
10. Sato, J. R., Fujita, A., Cardoso, E. F., Thomaz, C. E., & Rohde, L. A. *Evaluating effective connectivity between resting-state brain regions using conditional Granger causality.* NeuroImage, 54(4), 2947–2954, 2012.
11. Zhang, X., Wang, L., Zhao, Y., Wang, X., Zhang, T., & Li, H. *Lasso-based feature selection for Alzheimer's disease prediction using fMRI data.* Frontiers in Neuroscience, 14, 565, 2020.
12. Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., & Thirion, B. *Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines.* NeuroImage, 145, 166–179, 2017.
13. arashnic. ADHD Diagnosis Data. *Kaggle Datasets*, 2024. Disponible en: `https://www.kaggle.com/datasets/arashnic/adhd-diagnosis-data`. Último acceso: 12 de julio de 2025.
14. Cover, T. M., & Hart, P. E. *Nearest neighbor pattern classification.* IEEE Transactions on Information Theory, 13(1), 21–27, 1967.

Departamento de Matemáticas, Universidad Nacional Autónoma de Honduras.
Dirección de correo electrónico: `tlrodriguez@unah.hn`

CLASIFICACIÓN DE HOGARES POR NIVEL DE POBREZA USANDO FEEDFORWARD NEURAL NETWORKS (FNN) A PARTIR DE VARIABLES SOCIOECONÓMICAS EN HONDURAS

ANTONY SMAYKOL ROMERO BEJARANO

Resumen. La pobreza sigue siendo uno de los mayores desafíos sociales y su identificación precisa es clave para diseñar políticas efectivas. Esta investigación explora cómo las Feedforward Neural Networks (FNN), una técnica de aprendizaje automático, pueden clasificar hogares por nivel de pobreza a partir de variables socioeconómicas. Esta propuesta busca aportar una herramienta moderna para el análisis social, apoyando la toma de decisiones desde una perspectiva más predictiva y menos reactiva.

Abstract. Poverty remains one of the greatest social challenges and its accurate identification is key to designing effective policies. This research explores how Feedforward Neural Networks (FNN), a machine learning technique, can classify households by poverty level based on socioeconomic variables. This proposal aims to provide a modern tool for social analysis, supporting decision-making from a more predictive and less reactive perspective.

1. Introducción

La pobreza sigue siendo uno de los principales desafíos estructurales en América Latina, y en particular en países como lo es Honduras, donde afecta a más del 60% de la población [1]. La correcta identificación y clasificación de hogares según su nivel de pobreza es importante para el diseño de políticas públicas eficaces y focalizadas [2]. Tradicionalmente, este proceso ha dependido de encuestas extensas y análisis estadísticos convencionales, los cuales, si bien son útiles, suelen ser costosos, lentos y reactivos. Este tema es uno de los ejes de investigación de la Universidad Nacional Autónoma de Honduras (UNAH) que se tomara para el Desarrollo Económico y Social del país.

Trabajar en esta temática representa una oportunidad directa para mejorar las capacidades del país en la lucha contra la pobreza. En Honduras, los recursos disponibles para políticas sociales son limitados: el gasto social real ha disminuido en los últimos años, los ingresos fiscales son bajos y la atención a sectores como salud y educación presenta problemas de formulación y ejecución presupuestaria [3]. La aplicación de modelos predictivos eficientes permite focalizar recursos de forma más justa y eficaz, reduciendo pérdidas y elevando el impacto de los programas sociales.

En los últimos años, el uso de técnicas de aprendizaje automático ha emergido como una alternativa prometedora para mejorar la precisión y eficiencia en la detección de pobreza [4]. En particular, las Redes Neuronales Feedforward (FNN) han

Fecha: January 1, 2001 and, in revised form, June 22, 2001.

Palabras y frases clave. Clasificación de Pobreza, Feedforward Neural Networks, Machine Learning, Variables Socioeconómicas, Modelado Predictivo.

demostrado un alto rendimiento en tareas de clasificación basadas en múltiples variables socioeconómicas [5]. Estas redes pueden aprender patrones complejos entre características como ingreso, acceso a servicios básicos, nivel educativo y condiciones de vivienda, y utilizarlos para predecir el nivel de pobreza de un hogar sin necesidad de reglas explícitas.

El objetivo de esta investigación es aplicar una red neuronal FNN preentrenada o de configuración sencilla para clasificar hogares hondureños según su nivel de pobreza, utilizando un conjunto de variables socioeconómicas disponibles. Esta propuesta se enmarca en el esfuerzo por modernizar los instrumentos de análisis social y facilitar la toma de decisiones en contextos de alta vulnerabilidad como el hondureño. El estudio se desarrollará en un período limitado de tres meses, por lo cual se prioriza el uso de herramientas ya existentes y accesibles, como Google Colab o plataformas de minería de datos con entornos gráficos.

A través de esta aproximación, se busca contribuir a la evidencia empírica de que las redes FNN pueden ser útiles incluso en contextos con recursos limitados, siempre que se cuente con datos confiables y una adecuada preparación de los mismos [6].

2. Antecedentes

Las redes neuronales artificiales (ANN, por sus siglas en inglés) son modelos computacionales inspirados en el funcionamiento del cerebro humano, particularmente en la forma en que las neuronas biológicas procesan información. Uno de los tipos más conocidos y utilizados es la red neuronal de tipo feedforward (FNN), en la cual la información fluye en una única dirección, desde la capa de entrada hacia la capa de salida, sin ciclos ni retroalimentaciones.

El origen conceptual de las redes neuronales se remonta a 1943, cuando McCulloch y Pitts desarrollaron un modelo matemático simplificado de neurona que era capaz de realizar operaciones lógicas básicas [7]. A partir de ese modelo, en 1958, Frank Rosenblatt propuso el perceptrón, una red neuronal de una sola capa que podía aprender a clasificar patrones linealmente separables mediante un algoritmo de ajuste de pesos [8]. Sin embargo, durante las décadas de 1960 y 1970, las redes neuronales sufrieron un periodo de estancamiento debido a las críticas del libro de Minsky y Papert, que demostraba las limitaciones del perceptrón, especialmente su incapacidad para resolver problemas no lineales como la función XOR.

El verdadero avance de las redes neuronales feedforward llegó en los años 80, con la reintroducción y formalización del algoritmo de retropropagación del error (backpropagation) por Rumelhart, Hinton y Williams en 1986 [9]. Este algoritmo permitió el entrenamiento eficiente de redes con múltiples capas ocultas, lo que se conoce como perceptrones multicapa (MLP), una forma típica de FNN. A partir de ese momento, las FNN comenzaron a usarse ampliamente en tareas de clasificación, regresión y reconocimiento de patrones.

En el contexto de la investigación social y económica, el uso de redes neuronales, y particularmente de las FNN, ha demostrado ser prometedor para analizar fenómenos complejos y multidimensionales como la pobreza. Estas redes son capaces de modelar relaciones no lineales entre variables socioeconómicas y resultados de interés, lo cual es una ventaja significativa respecto a modelos estadísticos tradicionales como la regresión logística.

En países como Brasil, se ha utilizado una red neuronal artificial multicapa (feedforward) para clasificar municipios del estado de Rio Grande do Norte según su

nivel de vulnerabilidad social. El estudio, basado en 17 variables socioeconómicas, demográficas, educativas y de salud pública, implementó redes del tipo MLP y PNN, concluyendo que el modelo FNN con seis nodos ocultos presentó el mejor desempeño de clasificación, con mayor precisión que otros métodos supervisados tradicionales [10].

A escala continental, Jean et al. utilizaron redes neuronales convolucionales (CNN) aplicadas a imágenes satelitales de alta resolución para predecir niveles de pobreza en regiones de África subsahariana, obteniendo estimaciones de ingresos y pobreza con alta precisión, incluso en zonas con escasa información de campo [4].

En India, avances recientes han demostrado el uso exitoso de aprendizaje profundo con datos satelitales y censales para mapear pobreza y condiciones de bienestar a escala local. Por ejemplo, Daoud et al. combinaron imágenes diurnas de Landsat con redes profundas para estimar índices de riqueza y condiciones vivenciales en más de 200,000 pueblos, logrando resultados comparables a metodologías tradicionales y ofreciendo resolución espacial y temporal detallada [11].

En México, se ha propuesto una innovadora técnica para mapear pobreza multidimensional a nivel de manzana residencial. Usando CNN y mapas multispectrales satelitales entrenados con datos censales de CONEVAL, los investigadores obtuvieron un coeficiente de determinación R^2 de 0.80 para la dimensión de calidad de vivienda y 0.75 para variables como agua, luz y saneamiento, demostrando que es posible generar mapeos detallados de pobreza con alta precisión [12].

Estos estudios muestran que las redes neuronales—tanto del tipo feedforward como convolucionales—pueden capturar relaciones no lineales en datos socioeconómicos y espaciales, y constituyen herramientas valiosas para clasificar hogares o regiones según su condición socioeconómica. Su aplicabilidad en Honduras parece prometedora, dados los datos censales y variables relevantes disponibles que permiten replicar enfoques similares.

3. Definiciones Fundamentales sobre las FNN

Las redes neuronales artificiales (RNA) son modelos computacionales que tratan de imitar cómo funciona el cerebro humano. Están formadas por unidades llamadas neuronas artificiales, que se organizan por capas, y que básicamente sirven para aprender patrones desde datos. Estas redes se usan bastante en cosas como clasificación, regresión o para reconocer patrones en general [7].

Una red neuronal feedforward (FNN) es un tipo de RNA en donde la información solo va en una sola dirección: de la capa de entrada, pasando por una o varias capas ocultas, hasta llegar a la salida. Lo importante es que no hay ciclos ni retroalimentación. Este tipo de red es buena para problemas de clasificación o cuando se quiere predecir algo [13].

Cada neurona artificial realiza una operación matemática parecida a esta:

$$y = f\left(\sum_{i=1}^{n} W_i x_i + b\right) \tag{3.1}$$

donde x_i son las entradas, W_i los pesos, b es el sesgo (bias) que ajusta la posición de la función de activación, y f representa la función de activación. Esta función es importante porque permite que la red pueda aprender relaciones no lineales (o sea, más complejas).

La capa de entrada recibe las variables del problema que se quiere resolver, como por ejemplo el ingreso, el nivel educativo o si se tiene acceso a servicios básicos. Cada variable se representa como una neurona. Las capas ocultas son las que están entre la entrada y la salida. Ahí es donde realmente se produce el aprendizaje, combinando entradas y pesos, y luego aplicando funciones no lineales. El número de capas y de neuronas afecta directamente qué tan compleja puede ser la red. La capa de salida genera la predicción del modelo. Para clasificación binaria (por ejemplo, sí o no), se usa una sola neurona con función sigmoide. En el caso de clasificación multiclase, se usan varias neuronas con activación softmax.

La función de activación transforma la salida lineal de una neurona en algo no lineal. Algunas funciones comunes son[14]:

- ReLU (Rectified Linear Unit): $f(x) = \max(0, x)$

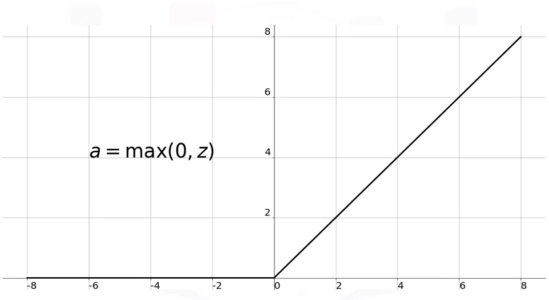


FIGURA 1. Función de activación ReLU.

- Sigmoide: $f(x) = \dfrac{1}{1 + e^{-x}}$

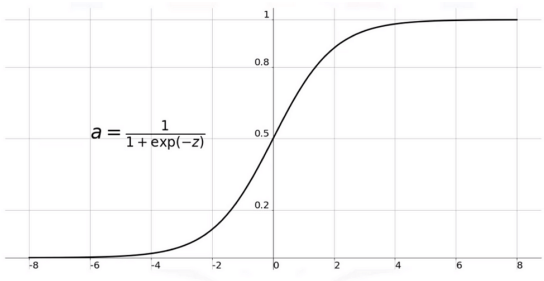


FIGURA 2. Función de activación Sigmoide.

- Softmax: convierte un vector de valores en una distribución de probabilidad (usada para clasificación multiclase).

El entrenamiento de una FNN se realiza usando el algoritmo de retropropagación del error (backpropagation) [13], el cual ajusta los pesos de la red para minimizar el error. Generalmente se usa el método de descenso de gradiente estocástico para lograr esto.

La función de pérdida es la que mide qué tanto se equivoca el modelo. En clasificación binaria, se suele utilizar la entropía cruzada binaria como función de pérdida:

$$\mathcal{L}(y, \hat{y}) = -\left[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})\right]. \tag{3.2}$$

4. Método General de Clasificación Usando FNN

Cuando se tienen datos sobre los hogares en Honduras, por ejemplo cosas como el ingreso que tienen, cuánta gente vive ahí o si tienen acceso a agua o cosas así, se puede usar una FNN para clasificarlos según qué tan pobres están. El proceso que se utiliza es el siguiente:

(1) Se ingresan los datos, es decir un vector de características que puede incluir variables como: ingreso mensual, tamaño del hogar, si hay energía eléctrica, agua, etc. a esto se le llama Entrada y de manera mas álgebraica seria de la forma:
$$x = (x_1, x_2, \ldots, x_n).$$
(2) Esos datos pasan por la red, en cada capa se hacen operaciones, se consideran los pesos de cada variable y se aplican funciones que no son lineales. Esto se llama propagación hacia adelante, y se ve así:
$$a^{(1)} = f(W^{(1)}x + b^{(1)}), \quad a^{(2)} = f(W^{(2)}a^{(1)} + b^{(2)}), ...$$
esto sirve para que la red aprenda a tomar decisiones más complejas.
(3) La red da una respuesta o predicción. Si es binaria, da 0 o 1 (por ejemplo, pobre o no). Si hay más clases, puede ser 0, 1, 2... o como se defina para el caso de clasificarlos por ejemplo como extremadamente pobre, pobre, clase media,etc. a esto se le conoce como Salida.
(4) Se entrena la red, esto se hace con datos que ya están etiquetados, es decir, que ya se sabe si son pobres o no, y la red va ajustando sus pesos para equivocarse lo menos posible aplicando el backpropagation. Todo eso se hace mediante el uso de una función de pérdida.
(5) Se mide si la red está funcionando bien o no. Para eso se usan métricas como la exactitud, precisión, sensibilidad, y otras. También se utiliza una matriz de confusión que ayuda a identificar en qué se está fallando.

5. Variables socioeconómicas

Para el desarrollo del presente modelo de predicción del nivel de pobreza, se realizó una selección específica de variables socioeconómicas que nos permitieran ver de mejor forma la situacion de los hogares hondureños. Estas variables fueron tomadas del Diccionario de Variables de la Encuesta Permanente de Hogares de Propósitos Múltiples (EPHPM), elaborado por el Instituto Nacional de Estadística (INE), correspondiente al año 2023-2024 [2], las variables son en común ya que dependiendo del año muchas de estas suelen cambiar o agregar nuevas.

El criterio de selección se centró en cuatro grandes dimensiones reconocidas en los estudios sociales y censales: ingreso, tipo de vivienda, acceso a servicios públicos y nivel educativo [2]. Estas dimensiones han demostrado ser altamente representativas para identificar condiciones de pobreza o vulnerabilidad en los hogares, tanto en Honduras como en contextos similares.

Las variables seleccionadas como ejemplo son:

- YTOTHG (Ingreso total del hogar): representa el ingreso mensual total en lempiras, lo cual constituye un indicador directo de la capacidad adquisitiva del hogar.
- V01 (Tipo de vivienda): refleja el tipo de estructura habitacional del hogar. Se codifica del 1 al 7, donde los valores corresponden a casa individual, rancho, casa improvisada, apartamento, cuartería, barrancón y local no construido para habitación pero usado como vivienda, respectivamente.
- V05 (Fuente principal de agua): describe el origen del agua que utiliza el hogar. Las opciones van del 1 al 9 e incluyen: tubería dentro de la vivienda, pozo con bomba, pozo con malacate, llave pública, río o manantial, carro cisterna, pick-up con barriles, otra vivienda u otra fuente no especificada.
- V07 (Tipo de alumbrado): indica la fuente de iluminación que utiliza el hogar, clasificada del 1 al 8. Estas categorías incluyen desde servicio público o privado hasta métodos no convencionales como lámparas de gas, energía solar, velas, ocote, entre otros.
- H07 (Tipo de sanitario): informa sobre las condiciones de saneamiento del hogar. Las categorías codificadas del 1 al 9 incluyen desde inodoros conectados a sistemas de alcantarillado o pozos sépticos, hasta letrinas, taza campesina, u hogares sin ningún tipo de sanitario.
- NIVEL (Nivel educativo alcanzado): señala el mayor nivel educativo alcanzado por algún integrante del hogar. Los valores incluyen sin nivel, primaria, secundaria y educación superior, además del código 9 para quienes no saben o no respondieron.
- POBREZA: variable binaria generada con fines ilustrativos que identifica si un hogar es considerado pobre (1) o no pobre (0), en función de sus condiciones de ingreso y otras características sociales.

La variable POBREZA presenta en la base de datos del EPHPM se categoriza en realidad por 3 valores los cuales serian extremadamente pobre (1), relativamente pobre (2) y no pobres(3), pero para este método de clasificación se decidió usar como variable binaria como fue definida anteriormente para el uso de la entropia cruzada binaria.

En la Tabla 1 se presentan datos para 20 hogares, elaborada con fines exclusivamente ilustrativos, con el propósito de mostrar cómo estas variables se estructuran dentro de la base de datos del EPHPM.

6. Preparación de datos

Antes de implementar un modelo de red neuronal, es necesario realizar una etapa de preprocesamiento de los datos, este paso resulta fundamental para garantizar que el modelo entienda la información de forma adecuada y pueda aprender de ella con mayor precisión y eficiencia. Uno de los procedimientos más comunes durante esta fase es la normalización de variables numéricas.

La normalización consiste en transformar los valores originales de las variables para que estén en una escala común, generalmente entre 0 y 1 [15], esto se vuelve necesario en las redes neuronales, ya que si algunas variables tienen rangos muy amplios (por ejemplo, ingresos en miles de lempiras) y otras tienen rangos pequeños (como niveles educativos codificados del 1 al 4), el modelo tiende a darle más importancia a las variables con mayor escala y esto puede provocar errores en el entrenamiento, dificultar la convergencia o incluso llevar a resultados inexactos.

Tabla 1. Variables socioeconómicas simuladas en 20 hogares

Hogar	YTOTHG	V01	V05	V07	H07	NIVEL	POBREZA
1	3200	2	5	5	9	1	1
2	7800	1	1	2	2	3	0
3	12000	4	1	1	1	4	0
4	5100	3	4	8	7	2	1
5	2700	6	6	7	9	1	1
6	15000	1	1	1	2	4	0
7	9000	1	2	3	3	3	0
8	6300	5	8	6	5	2	1
9	10200	4	1	1	1	4	0
10	4500	2	3	4	6	2	1
11	2800	7	9	5	9	1	1
12	13400	1	1	1	1	4	0
13	3500	3	5	6	9	1	1
14	8700	4	1	1	2	3	0
15	5600	2	4	7	7	2	1
16	11000	1	1	1	2	4	0
17	3900	6	7	5	8	1	1
18	9800	4	2	2	1	3	0
19	2900	5	6	6	7	1	1
20	7200	1	3	4	6	2	1

Una de las técnicas más utilizadas para normalizar es la normalización min-max, donde cada valor se ajusta con la siguiente fórmula:

$$x_{\text{normalizado}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

Esta fórmula garantiza que todos los valores queden comprendidos dentro del rango de 0 a 1, lo cual facilita que la red neuronal procese la información de forma balanceada y eficiente [15].

En la Tabla 2 se muestra un ejemplo de cómo se vería la normalización de la variable YTOTHG (Ingreso total del hogar) para cinco hogares ficticios.

Este tipo de transformación es aplicada a todas las variables numéricas continuas, especialmente al ingreso total, ahora para variables como V01(tipo de vivienda) tienen un nombre en específico y se conoce como las variables categóricas; y son aquellas que representan atributos en forma de categorías o códigos numéricos, pero

Tabla 2. Ejemplo de normalización de ingreso total del hogar

Hogar	YTOTHG original	Normalizado
1	3200	0.075
2	7800	0.366
3	12000	0.612
6	15000	0.796
12	13400	0.708

que no tienen un significado matemático directo [4]. Por ejemplo, el hecho de que una vivienda sea tipo 1 (casa individual) o tipo 6 (barrancón) no implica que la categoría 6 sea más que la 1 en un sentido cuantitativo, sino simplemente que son distintos tipos de vivienda.

En este tipo de variables, no es correcto aplicar técnicas como la normalización, ya que los valores no representan cantidades continuas. Por ello, una técnica muy común es la codificación one-hot (o codificación ficticia).

La codificación one-hot transforma una variable categórica en varias columnas binarias (0 o 1), una por cada categoría posible y cada fila tiene un valor de 1 en la columna correspondiente a su categoría y 0 en las demás, de esta manera, se evita imponer una relación de orden artificial entre las categorías [16].

A continuación, se presenta un ejemplo para ilustrar cómo se vería esta codificación para la variable V01 (tipo de vivienda), con datos ficticios:

Tabla 3. Codificación one-hot para la variable V01 (tipo de vivienda)

Hogar	V01 (Original)	V01_1	V01_2	V01_3	V01_4	V01_5	V01_6	V01_7
1	2	0	1	0	0	0	0	0
2	1	1	0	0	0	0	0	0
3	4	0	0	0	1	0	0	0
4	2	0	1	0	0	0	0	0
5	3	0	0	1	0	0	0	0

7. Estructura y entrenamiento del modelo

Una vez finalizado el proceso de preparación de datos, se construye el modelo FNN para llevar a cabo la clasificación binaria de hogares según su situación de pobreza donde el modelo consta de una estructura secuencial de capas:

- Una capa de entrada, cuyo número de neuronas coincide con la cantidad de variables resultantes, es decir, la suma de variables numéricas normalizadas y columnas generadas por codificación one-hot).

- Una primera capa oculta con 32 neuronas, utilizando la función de activación ReLU, que permite al modelo aprender relaciones no lineales sin saturarse con facilidad [18].
- Una segunda capa oculta con 16 neuronas, también con activación ReLU.
- Una capa de salida con una sola neurona, que emplea la función de activación sigmoide [11] para producir una probabilidad entre 0 y 1, representando la pertenencia del hogar a la categoría pobre.

Una vez definida la estructura del modelo, se procede a su entrenamiento y para ello, se compila utilizando la función de pérdida entropía cruzada binaria [18], apropiada para problemas de clasificación binaria. Como optimizador se emplea Adam, una variante del descenso de gradiente que ajusta dinámicamente la tasa de aprendizaje, lo que permite una convergencia más rápida y estable [17].

El entrenamiento se lleva a cabo sobre un conjunto de datos divididos previamente en entrenamiento y prueba donde se utilizan 100 épocas y un tamaño de lote (batch size) de 10 observaciones por iteración. Durante cada época, el modelo procesa los datos de entrenamiento, calcula la pérdida, ajusta los pesos y mejora progresivamente su capacidad de predicción [18], gracias a este proceso iterativo, el modelo aprende a identificar combinaciones de características que predicen con mayor precisión si un hogar está en situación de pobreza o no, basándose en las variables socioeconómicas proporcionadas.

8. Implementación del modelo FNN

La implementación se llevo a cabo en el lenguaje Python haciendo uso de las bibliotecas pandas, pyreadstat, scikit-learn y keras del entorno TensorFlow [19], esto se podra verse a continuación donde se mostrara el código utilizado.

```
import pyreadstat
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras import Input
```

```
# Leer archivo .sav
df, meta = pyreadstat.read_sav('Data de la Encuesta de Hogares junio 2023.sav')
variables = ['V01', 'V02', 'V03', 'V04', 'V05', 'V06', 'V07', 'V08', 'V09',
             'H03', 'H04', 'H05', 'H06', 'H07', 'H08', 'H09',
             'NIVEL', 'POBREZA']
df = df[variables]
df = df.dropna()

# Reconvertir pobreza: 1 y 2 como 1 (pobre), 3 como 0 (no pobre)
df['POBREZA'] = df['POBREZA'].replace({1: 1, 2: 1, 3: 0})
```

El código esta comentado para mejor comprensión; como se puede observar lo primero es la lectura de la base de datos EPHPM del año 2023 y se seleccionaron en este caso 20 variables centradas en el censo de Honduras como vivienda, educación,

ingreso, acceso a servicios públicos y la variable POBREZA [2] a la cual se hace una conversión donde se reemplaza 1 y 2 por 1; de igual manera 3 por 0. Se descartan todas las filas que presenten valores nulos para un entrenamiento más eficiente del modelo FNN [15].

```
# Separacion de variables predictoras y objetivo
X = df.drop('POBREZA', axis=1)
y = df['POBREZA']

# one-hot para variables categóricas
categorical_cols = ['V01', 'V02', 'V03', 'V04', 'V05', 'V06', 'V07', 'V08',
                    'H03', 'H04', 'H05', 'H06', 'H07', 'H08', 'NIVEL']
X = pd.get_dummies(X, columns=categorical_cols)
```

```
# Dividir en entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
# Normalizar variables continuas
scaler = StandardScaler()
for col in ['V09', 'H09']:
    X_train[col] = scaler.fit_transform(X_train[[col]])
    X_test[col] = scaler.transform(X_test[[col]])
```

Tendríamos ahora lo que es la preparación de datos donde aplicamos la normalización para las variables con valor escalar(continuas) [15], que en este caso solo son 2; y se aplica la codificación one-hot para el resto de variables que son catégoricas [16].Se hace una división de los datos donde un 80% son para entrenamiento y un 20% para usarlos de prueba.

```
# Implementación del modelo
model = Sequential()
model.add(Input(shape=(X_train.shape[1],)))
model.add(Dense(32, activation='relu'))
model.add(Dense(16, activation='relu'))
model.add(Dense(1, activation='sigmoid'))
```

```
# Compilación y entrenamiento
model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])
model.fit(X_train, y_train, epochs=100, batch_size=10, validation_data=(X_test, y_test))
```

Y por último seria ya la implementación del modelo; se crea la red neuronal en la primera linea y le sigue lo que es definir a la red la forma del vector a trabajar siendo la capa de entrada. En las siguientes dos lineas tendriamos las dos capas ocultas de 32 y 16 neuronas respectivamente ambas con la función de activación ReLU y la salida que tendria con función de activación la sigmoide para clasificación binaria.

Para la compilación definimos la función de pérdida que seria la entropia cruzada binaria [13] como optimizador utilizamos Adam que optimiza los pesos de forma automática [16] y el entrenamiento como se habia dicho con 100 épocas y 10 entrenamientos antes de actualizar cada peso.

9. Resultados del modelo FNN

Una vez entrenado el modelo de red neuronal con las variables seleccionadas, se procedió a evaluar su desempeño utilizando el conjunto de prueba que recordemos son el 20% de los datos. Para ello, se emplearon métricas estándar de clasificación binaria como la exactitud(accuracy), precisión(precision), sensibilidad (recall) y f1-score [18], además de la matriz de confusión, que permite visualizar en qué clases el modelo acierta o se equivoca.

La siguiente figura muestra la matriz de confusión obtenida:

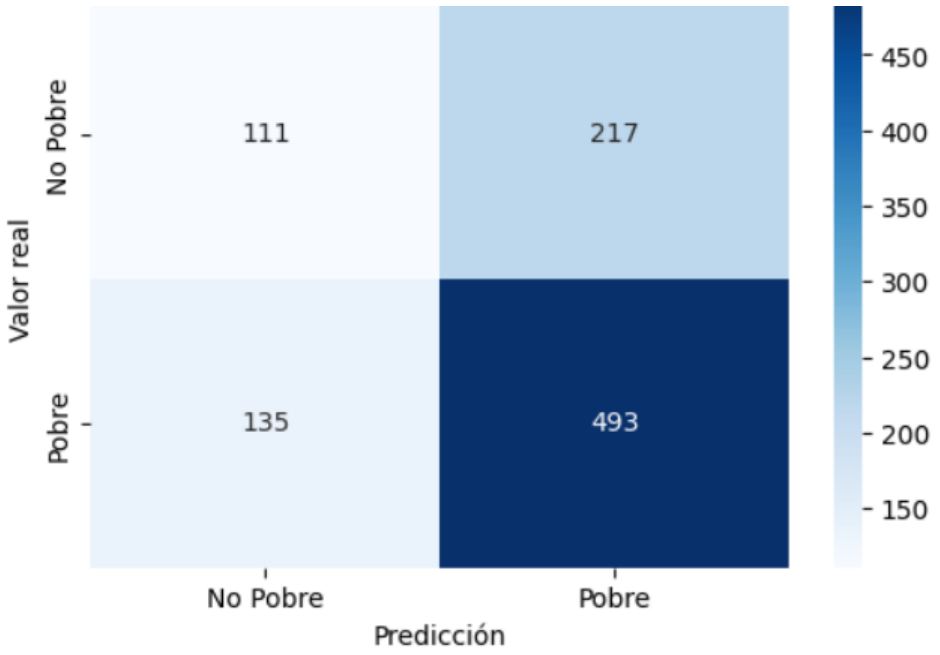


Figura 3. Matriz de confusión

La matriz indica que de los 328 hogares que realmente no eran pobres, el modelo clasificó correctamente a 111, pero se equivocó en 217, identificándolos erróneamente como pobres. En cambio, de los 628 hogares pobres, logró clasificar correctamente a 493 y se equivocó en 135.

	precision	recall	f1-score	support
0.0	0.45	0.34	0.39	328
1.0	0.69	0.79	0.74	628
accuracy			0.63	956
macro avg	0.57	0.56	0.56	956
weighted avg	0.61	0.63	0.62	956

Figura 4. Reporte de clasificación

Estos resultados se complementan con el reporte de clasificación presentado en la figura 4, donde se observa que el modelo obtuvo una exactitud global del 63%, lo que indica que predice correctamente en aproximadamente 6 de cada 10 casos. Al analizar las métricas por clase:

Para la clase "No Pobre", se obtuvo una precisión de 0.45 y un recall de 0.34, lo cual muestra que el modelo tiene dificultades para identificar correctamente a los hogares no pobres. Para la clase "Pobre", se obtuvo una precisión de 0.69 y un recall de 0.79, indicando un buen desempeño en la detección de pobreza.

Finalmente, el valor promedio ponderado del f1-score fue de 0.62, reflejando un desempeño aceptable del modelo, especialmente en la clase positiva (pobreza), aunque aún con margen de mejora en la clase negativa.

10. Conclusiones

La presente investigación demostró que es posible implementar un modelo de red neuronal feedforward para predecir la condición de pobreza de los hogares hondureños a partir de variables socioeconómicas seleccionadas. Sin embargo, durante el proceso se identificaron distintos puntos que pueden contribuir a mejorar el desempeño del modelo y la interpretación de los resultados, en primer lugar, aunque se obtuvo una exactitud aceptable del 63% y un f1-score de 0.62, se considera que el modelo puede ser mejorado a través de diferentes estrategias. Entre las más importantes se encuentran: la incorporación de una mayor cantidad de variables relevantes (por ejemplo, información del jefe del hogar, empleo o salud, hacinamiento) y la prueba de arquitecturas más especializadas; de igual manera se puede considerar la exploración de otras métricas y funciones de activación para afinar la precisión de la clasificación.

Por otro lado, cabe señalar que la base de datos utilizada, correspondiente a la Encuesta Permanente de Hogares de Propósitos Múltiples (EPHPM) del Instituto Nacional de Estadística de Honduras, contiene un total de 20,308 registros, sin embargo, al filtrar las observaciones que contenían valores nulos o sin respuesta en las variables seleccionadas, la cantidad de datos útiles se redujo ligeramente, y a pesar de esta depuración, la muestra conservó una dimensión suficiente para entrenar y validar el modelo con resultados significativos.

Finalmente, se identificó que algunas variables categóricas dentro de la encuesta han cambiado sus etiquetas o estructuras en los últimos años, lo que puede dificultar la comparación entre distintos periodos. Este aspecto resalta la necesidad de una estandarización y documentación clara de las variables utilizadas, especialmente si se busca replicar o extender este análisis en el futuro.

References

1. ECLAC, *Panorama Social de América Latina 2023*, Naciones Unidas, Santiago de Chile, 2023.
2. Instituto Nacional de Estadística, *Encuesta Permanente de Hogares de Propósitos Múltiples (EPHPM)*, INE, Tegucigalpa, 2023-2024.
3. World Bank, *Central America social expenditures and institutional review: Honduras*, World Bank, Washington, 2015.
4. N. Jean, M. Burke, M. Xie, D. W. Davis, D. B. Lobell, and S. Ermon, *Combining satellite imagery and machine learning to predict poverty*, Science **353** (2016), no. 6301, pp. 790-794. DOI 10.1126/science.aaf7894.

5. F. Nkurunziza, R. Kabanda, and P. McSharry, *Enhancing poverty classification in developing countries through machine learning: A case study of household consumption prediction in Rwanda*, Cogent Economics & Finance **13** (2024), no. 1, DOI: 10.1080/23322039.2024.2444374.
6. O. Shah and K. Tallam, *Novel machine learning approach for predicting poverty using temperature and remote sensing data in Ethiopia*, arXiv preprint (2023), DOI: 10.48550/arXiv.2302.14835.
7. W. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous activity*, Bull. Math. Biophys. **5** (1943), pp. 115–133, DOI: 10.1007/BF02478259.
8. F. Rosenblatt, *The perceptron: A probabilistic model for information storage and organization in the brain*, Psychol. Rev. **65** (1958), no. 6, pp. 386–408, DOI: 10.1037/h0042519.
9. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning representations by back-propagating errors*, Nature **323** (1986), no. 6088, pp. 533–536, DOI: 10.1038/323533a0.
10. R. L. Wingerter, S. S. Araujo, R. D. M. Silva, and A. L. M. Silva, *The use of artificial neural networks to classify the social vulnerability of municipalities in Rio Grande do Norte State, Brazil*, Cadernos de Saúde Pública **36** (2020), DOI: 10.1590/0102-311x00038319.
11. A. Daoud, F. Jordan, M. Sharma, F. Johansson, D. Dubhashi, S. Paul, and S. Banerjee, *Measuring poverty in India with machine learning and remote sensing*, arXiv preprint (2021), DOI: 10.48550/arXiv.2202.00109.
12. M. Zea-Ortiz, P. Ver, J. Salas, R. Manduchi, E. Villaseñor, A. Figueroa, and R. Suárez, *A data-driven approach to mapping multidimensional poverty at residential block level in Mexico*, Environment, Development and Sustainability (2024), pp. 1-24, DOI: 10.1007/s10668-024-05230-z.
13. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning representations by back-propagating errors*, Nature **323** (1986), no. 6088, pp. 533–536, DOI: 10.1038/323533a0.
14. H. Liu, Y. Jin, X. Song, and Z. Pei, *Rate of penetration prediction method for ultra-deep wells based on LSTM–FNN*, Applied Sciences, **12** (2022), no. 15, 7731.
15. S. Bhanja and A. Das, *Impact of Data Normalization on Deep Neural Network for Time Series Forecasting*, arXiv preprint (2018), DOI: 10.48550/arXiv.1812.05519.
16. S. Okada, M. Ohzeki and S. Taguchi, *Efficient partition of integer optimization problems with one-hot encoding*, Scientific Reports **9** (2019),no. 1, pp. 13036, DOI: 10.1038/s41598-019-49539-6.
17. D. Kingma and J. Ba, *Adam: A Method for Stochastic*, arXiv preprint (2015), DOI: 10.48550/arXiv.1412.6980
18. E. J. V. Fuentes, *Introducción a los métodos Deep Learning basados en Redes Neuronales*, Universidad Nacional de Colombia, 2019.
19. Y. Yang, X. Hao, L. Zhang and L. Ren, *Application of Scikit and Keras Libraries for the Classification of Iron Ore Data Acquired by Laser-Induced Breakdown Spectroscopy (LIBS)*, Sensors **20** (2020), no. 5, pp. 1393, DOI: 10.3390/s20051393.

APROXIMACIÓN ASINTÓTICA DE LA FUNCIÓN FACTORIAL EN UNA VARIABLE

HENRRY ALEXANDER MORAZÁN CHAVÉZ

RESUMEN. En este trabajo se estudian aproximaciones asintóticas para la función factorial $n!$, extendida al dominio real mediante la función gamma $\Gamma(n+1)$. Se analizan desarrollos clásicos, como la fórmula de Stirling y sus refinamientos, incluyendo series asintóticas y aproximaciones más precisas para valores grandes del argumento. Además, se discuten aplicaciones en probabilidad, estadística y teoría de números, así como la conexión con otras funciones especiales. Finalmente, se presentan cotas de error y comparaciones numéricas que ilustran la eficacia de estas aproximaciones en distintos contextos.

ABSTRACT. This work studies asymptotic approximations for the factorial function $n!$, extended to the real domain via the gamma function $\Gamma(n+1)$. Classical results are analyzed, including Stirling's formula and its refinements, along with asymptotic series and more precise approximations for large argument values. Additionally, applications in probability, statistics, and number theory are discussed, as well as connections to other special functions. Finally, error bounds and numerical comparisons are presented to illustrate the effectiveness of these approximations in different contexts.

1. INTRODUCCIÓN

Dentro de la programación es muy importante el costo computacional de operaciones matemáticas, más aún cuando se requiere el cálculo de factoriales relativamente grandes. Los factoriales son fundamentales en matemáticas, apareciendo en áreas como la combinatoria, la probabilidad y la física estadística [1]. Sin embargo, su crecimiento superexponencial los vuelve rápidamente inmanejables para cálculos directos. Por ejemplo, 70! supera el número de partículas en el universo observable, lo que plantea un desafío computacional enorme [7]. Aquí es donde entra en juego la aproximación asintótica, una herramienta que permite simplificar expresiones factoriales sin perder precisión [2].

Las tres métodos usados para el cálculo de factoriales relativamente grandes son la fórmula de Stirling(Aproximación básica), la Función Gamma y la más reciente Aproximación de Ramanujan [6, 3]. La fórmula de Stirling, creada por James Stirling en 1730, es un método clásico para calcular aproximaciones de factoriales de números grandes, muy usado en física y estadística. Su versión más simple es rápida pero poco precisa con números pequeños, aunque mejora significativamente al aumentar el valor. Más tarde, se desarrolló una versión refinada que incluye

Fecha: Junio 9, 2025.

Palabras y frases clave. Función factorial, función gamma, fórmula de Stirling, Aproximación de Ramanujan, aproximación asintótica.

un término adicional, ofreciendo mayor exactitud incluso con números moderados, ideal para aplicaciones prácticas donde se necesita un equilibrio entre simplicidad y precisión.

Por su parte, la función Gamma, introducida por Leonhard Euler en 1729, no es una aproximación sino una generalización exacta del factorial para números no enteros y complejos [3]. Es fundamental en matemáticas avanzadas y física teórica, aunque su cálculo requiere métodos numéricos en la mayoría de los casos.

Finalmente, la aproximación de Ramanujan, formulada a principios del siglo XX por el genio matemático indio Srinivasa Ramanujan, destaca por su asombrosa precisión, incluso con números pequeños. Su enfoque, basado en observaciones profundas más que en desarrollos tradicionales, la convierte en la opción preferida cuando se busca la mayor exactitud posible, superando a las versiones clásicas de Stirling. Desarrollada en el siglo XVIII, es la solución más famosa a este problema [16].

Las aplicaciones de esta aproximación son varias. En computación, permite calcular logaritmos de factoriales de manera eficiente, esencial en algoritmos de aprendizaje automático y estadística bayesiana. En física teórica, ayuda a modelar sistemas con un número enorme de partículas, como en termodinámica ($S = k \ln \Omega$) o mecánica cuántica. Incluso en teoría de números, el estudio asintótico de $\ln(n!)$ revela propiedades profundas sobre la distribución de primos.

Más que una herramienta útil, la aproximación de Stirling conecta dos conceptos clave: lo discreto y lo continuo. Al igual que las ecuaciones diferenciales pueden describir el crecimiento de una población de bacterias (que, aunque discreta, se modela como continua), esta fórmula nos permite aplicar técnicas del cálculo a problemas combinatorios. Su importancia va más allá de las matemáticas teóricas, extendiéndose a áreas como la criptografía, la simulación de procesos aleatorios e incluso la inteligencia artificial.

¿Por qué sigue siendo relevante hoy? En un mundo impulsado por el big data y los modelos complejos, dominar estas aproximaciones no es solo un desafío intelectual, sino una necesidad práctica. Ya sea para optimizar algoritmos, analizar redes complejas o predecir comportamientos en sistemas físicos, la habilidad de simplificar lo aparentemente imposible sigue siendo una de las mayores contribuciones de las matemáticas al progreso científico y tecnológico.

La aproximación asintótica de funciones factoriales, aunque de naturaleza teórica, posee aplicaciones prácticas directas que fortalecen la productividad y la competitividad, ejes fundamentales para el desarrollo económico sostenible de Honduras. Este trabajo se enmarca en las prioridades de investigación establecidas por la Universidad Nacional Autónoma de Honduras (UNAH), las cuales buscan responder a los desafíos globales desde una perspectiva local. En particular, este estudio se alinea con el eje de Desarrollo Económico y Social, centrándose en el tema prioritario (c) Globalización, productividad y competitividad [8]. Reconociendo la relevancia estratégica de este enfoque, ya que contribuye a generar soluciones innovadoras que impulsan el crecimiento económico y el bienestar social en Honduras.

2. Antecedentes

El estudio de las aproximaciones asintóticas para la función factorial $n!$ surgió en el siglo XVIII como respuesta a los desafíos prácticos que planteaba el cálculo de factoriales para valores grandes de n. En esa época, el cálculo exacto de factoriales para n grandes resultaba imposible con los métodos manuales disponibles, lo que motivó la búsqueda de aproximaciones eficientes.

Este problema era particularmente relevante en el contexto de la teoría de probabilidades, donde De Moivre trabajaba en la aproximación de coeficientes binomiales para grandes conjuntos de datos [9, 10]. La primera versión de lo que hoy conocemos como fórmula de Stirling fue desarrollada por De Moivre alrededor de 1730, quien estableció una relación entre factoriales y logaritmos, reconociendo que $ln(n!)$ podía aproximarse mediante integrales, sentando así las bases para el desarrollo posterior de la fórmula que lleva el nombre de Stirling.

La contribución decisiva de James Stirling en su obra "Methodus Differentialis"(1730) consistió en identificar la presencia de la constante $\sqrt{2\pi}$ en la aproximación, la aparición de π en este contexto fue un descubrimiento que reveló conexiones profundas entre el análisis discreto y continuo [6]. Este refinamiento matemático no solo mejoró la precisión de la aproximación, sino que también estableció un vínculo crucial con la integral de Gauss y la función gamma, que sería explorada posteriormente por Euler [11, 12].

La función gamma, introducida por Leonhard Euler en el siglo XVIII, generaliza el concepto de factorial a números reales y complejos mediante su integral, donde $\Gamma(n+1) = n!$ para enteros positivos n. Esta extensión permitió estudiar el comportamiento asintótico de los factoriales a través de herramientas de análisis continuo, vinculándose directamente con la famosa aproximación de Stirling [6]. La conexión entre ambas surgió al analizar el crecimiento de $\Gamma(z)$ para valores grandes de z, demostrando que la fórmula de Stirling era un caso particular de un fenómeno más amplio en el análisis asintótico [3].

Durante el siglo XIX, matemáticos como Poisson y Laplace aplicaron estas aproximaciones en diversos campos, desde la teoría de probabilidades hasta la mecánica estadística, donde era necesario trabajar con sistemas que involucraban un número enorme de partículas [13]. La necesidad de mayor precisión llevó posteriormente al desarrollo de expansiones asintóticas más completas, incorporando términos adicionales basados en los números de Bernoulli, que permitían controlar el error en aplicaciones prácticas [19].

Srinivasa Ramanujan, el genio matemático autodidacta indio del siglo XIX, revolucionó las aproximaciones factoriales con fórmulas empíricamente precisas y analíticamente elegantes con su versión mejorada, superó la precisión de Stirling para valores pequeños de n, con errores menores al 0,01 % [14]. Ramanujan logró esto combinando intuición numérica con técnicas no convencionales, incluyendo el uso de fracciones continuas y series hipergeométricas [15]. Sus contribuciones no solo refinaron las aproximaciones existentes, sino que también abrieron nuevas vías en

el análisis asintótico, influyendo en campos como la teoría de números y la computación científica. Su legado persiste en algoritmos modernos que requieren cálculos precisos de factoriales y funciones relacionadas.

Actualmente, existen varios métodos y técnicas para aproximar asintóticamente la función factorial $n!$ y su generalización la función gamma $\Gamma(z)$. Estos métodos son fundamentales en matemáticas aplicadas, física, estadística y computación. Algunos de los mas usados en la actualidad son: Aproximación de Stirling: su versión clásica y su versión mejorada, Método de Laplace, aproximación de Ramanujan [22].

En la actualidad la aproximación asintótica de los factoriales tiene diferentes usos. En probabilidad y estadística sus usos en distribuciones como la binomial, Poisson o normal, donde se calculan factoriales grandes como también para simplificar cálculos en cadenas de Markov y procesos estocásticos. En la teoría de números y algoritmos sus usos en análisis de complejidad computacional, en criptografía, para evaluar la seguridad de algoritmos basados en problemas factoriales. En optimización y machine learning en métodos de Monte Carlo y algoritmos de aproximación para distribuciones multinomiales y para aproximar logaritmos de factoriales, comunes en modelos probabilísticos [17, 18].

3. Análisis Asintótico de la Función Factorial en una Variable

3.1. Fundamentos Teóricos. Comenzaremos este análisis con algunas definiciones y conceptos: factorial, relación de recurrencia, función gamma, análisis asintótico, métodos de aproximación, comportamiento asintótico, aproximaciones clásicas y modernas.

Definición 3.1. Para un $n \geq 0, n$ factorial, se define como:

$$0! = 1$$
$$n! = (n)(n-1)(n-2)\cdots(3)(2)(1), \ \textit{para } n \geq 1$$

Además, para cada $n \geq 0, (n+1)! = (n+1)(n!)$.

Debemos tomar nota de la rapidez con la que crecen los valores de $n!$ así que antes de proseguir, intentaremos tener una idea más clara de la velocidad con que crece $n!$, podemos calcular que $10! = 3,628,800$, y éste es precisamente el numero de segundos que hay en seis semanas. En consecuencia, 11! es superior al número de segundos que tiene un año, 12! supera el número que hay en 12 años, y 13! sobrepasa la cantidad de segundos que tiene un siglo [20].

Una propiedad clave de los factoriales es su definición recursiva:

$$n! = n(n-1)!$$

Esta relación permite calcular cualquier factorial si se conoce el factorial del número anterior.

La función factorial, en su definición original, solo se aplica a los enteros no negativos. Sin embargo, en matemáticas y física, a menudo es necesario extender el concepto de factorial a números no enteros (como números reales o complejos). La función Gamma Γ es la extensión más común y utilizada para este propósito.

Definición 3.2. La función Gamma se define a través de una integral impropia conocida como la integral de Euler de segunda especie. Para un número complejo z con una parte real positiva, se define como:

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt, \quad (Re(z) > 0).$$

si evaluamos la función Γ en n natural obtenemos,

$$\Gamma(n) = (n-1)!$$

La conexión fundamental entre la función Gamma y la función factorial se establece mediante la siguiente identidad:

$$\Gamma(n+1) = (n)!$$

Esta relación demuestra que la función Gamma coincide con el factorial para todos los números enteros no negativos, pero su dominio de definición es mucho más amplio. Algunas propiedades de la función Gamma que se reflejan en las de los factoriales:
La función Gamma satisface la relación de recurrencia: $\Gamma(z+1) = z\Gamma(z)$. Donde esta es una generalización de la propiedad $n! = n(n-1)!$. Gracias a la función Gamma, es posible asignar un valor a "factoriales"de números no enteros [21].

Ejemplo 3.3. Calcule $(1/2)!$.
Utilizando la relación $\Gamma(\frac{1}{2}+1) = (\frac{1}{2})!$.
Sabemos que $\Gamma(1/2) = \sqrt{\pi}$, por lo tanto:

$$\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$$

Así, podemos decir que:

$$\left(\frac{1}{2}\right)! = \frac{\sqrt{\pi}}{2}.$$

El análisis asintótico estudia el comportamiento de funciones cuando una variable (usualmente n) tiende a un valor límite (como ∞ o 0). Las notaciones O, o, y y $\sim$, son herramientas fundamentales para comparar funciones en este contexto [2, 18].

Definición 3.4. Sea $f(n) = O(n)$, es de orden superior cuando $n \to \infty$ si existen constantes $C > 0$ y $n_0 \in N$ tales que:

$$|f(n)| \leq C \cdot |g(n)|$$

para todo $n \geq n_0$.

Ejemplo 3.5. $3n^2 + 5n + 2 = O(n^2)$, porque para $n \geq 1$,

$$3n^2 + 5n + 2 \leq 10n^2$$

.

Un caso particular: Si $f(n) = O(1)$, entonces $f(n)$ esta acotada cuando $n \to \infty$.

Definición 3.6. Sea $f(n) = o(g(n))$, es de orden inferior cuando $n \to \infty$ si:

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$$

Ejemplo 3.7. $n = o(n^2)$, porque:

$$\lim_{n\to\infty} \frac{n}{n^2} = 0$$

.

Definición 3.8. Sea $f(n) \sim g(n)$, es equivalente asintóticamente cuando $n \to \infty$ si:

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$$

.

Ejemplo 3.9. $n! \sim \sqrt{2\pi n}(\frac{n}{e})^n$.

Una de las aplicaciones en el análisis de $n!$ es la aproximación de Stirling:

$$n! \sim \sqrt{2\pi n}(\frac{n}{e})^n \ (n \to \infty).$$

Aquí la equivalencia asintótica $\sim$ captura el término dominante, mientras que O y o permiten refinamientos.
La aproximación de Stirling para el factorial $n!$ puede obtenerse mediante diferentes métodos, entre los que están: la integración por partes, el método de Laplace y la formula de Euler-Maclaurin [24].
La integración por partes es esencial para conectar la función $\Gamma(n+1)$ con $n!$.
La función Γ generaliza el factorial para números complejos, su integral es:

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt.$$

Si integramos por partes $\Gamma(z+1)$, tomando:

$$u = t^z \to du = z \cdot t^{z-1}dt,$$
$$dv = e^{-t}dt \to v = -e^{-t},$$

obtenemos:

$$\Gamma(z+1) = \Big[-t^z e^{-t} \Big]_0^\infty + z\int_0^\infty t^{z-1}e^{-t}dt = z\Gamma(z).$$

Para $z = n$ entero, demuestra la relación recursiva del factorial:

$$n! = n \cdot (n-1)!,$$

sin embargo, no da directamente la aproximación de Stirling, sino que nos ayuda a entender la estructura de $\Gamma(z)$ [21].

La integral de Laplace se utiliza para aproximar el valor de integrales de la forma:

$$\int_a^b e^{Mf(x)}dx, \quad cuando \quad M \to \infty$$

donde $f(x)$ tiene un máximo único y pronunciado en el intervalo $[a, b]$. Usando la representación integral de $\Gamma(n+1)$:

$$n! = \int_0^\infty e^{-t}t^n dt.$$

Hacemos el cambio de variable $t = ns$, para enfocarnos en el comportamiento cuando $n \to \infty$:

$$n! = n^{n+1} \int_0^\infty e^{-ns} s^n ds = n^{n+1} \int_0^\infty e^{n(lns-s)} ds.$$

Definiendo $f(s) = lns - s$. Donde el máximo de $f(s)$ ocurre cuando $s = 1$:

$$f'(1) = 0,$$
$$f''(1) = -1.$$

Aplicando el método de Laplace alrededor de $s = 1$:

$$f(s) \approx f(1) + \frac{f''(1)}{2}(s-1)^2 = -1 - \frac{(s-1)^2}{2}.$$

Entonces:

$$n! \approx n^{n+1} e^{-n} \int_{-\infty}^\infty e^{-n\frac{(s-1)^2}{2}} ds.$$

La integral es una Gaussiana:

$$\int_{-\infty}^\infty e^{-n\frac{u^2}{2}} du = \sqrt{\frac{2\pi}{n}}.$$

Por lo tanto:

$$n! \approx n^{n+1} e^{-n} \sqrt{\frac{2\pi}{n}} = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n,$$

donde esta es la formula de Stirling.
El método de Laplace es clave para obtener Stirling desde la integral de $\Gamma(n+1)$. Es muy potente para integrales exponenciales pero requiere que $f(x)$ tenga un máximo bien definido.

Teorema 3.10. *Dada $f \in C^{2m+2}[a,b]$, con $a, b \in Z$, se cumple:*

$$\sum_{k=a}^{b} f(k) = \int_a^b f(x)dx + \frac{f(a)+f(b)}{2} + \sum_{k=1}^{m} \frac{B_{2k}}{(2k!)} \left(f^{(2k-1)}(b) - f^{(2k-1)}(a) \right) + R_m.$$

Donde B_{2k} son los números de Bernoulli, R_m es un término de error, acotado por derivadas de f [23].

La fórmula de Euler-Maclaurin es otra forma de deducir Stirling usando la aproximación de sumas por integrales.
Partiendo de:

$$\ln(n!) = \sum_{k=1}^{n} \ln k.$$

Aplicamos la fórmula de Euler-Maclaurin a $f(k) = \ln k$:

$$\sum_{k=1}^{n} f(k) = \int_1^n \ln x dx + \frac{\ln n}{2} + \sum_{m=1}^{M} \frac{B_{2m}}{(2m!)} (f^{(2m-1)}(n) + C + error.$$

Donde:

$$\int \ln x dx = x \ln x - x,$$
$$f^{(2m-1)}(x) = \frac{(2m-2)!}{x^{2m-1}}.$$

Tomando $M = 1$ y despreciando términos de orden superior:

$$\ln(n!) \approx n\ln n - n + \frac{\ln n}{2} + C + \frac{B_2}{2}\frac{1}{n} = n\ln n - n + \frac{\ln n}{2} + C + \frac{1}{12n}.$$

Para determinar C, se compara con el resultado exacto de $\Gamma(n)$, obteniendo:

$$C = \sqrt{2\pi}.$$

Finalmente, exponenciando se recupera la aproximación de Stirling:

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

La fórmula de Euler-Maclaurin refina la aproximación de $ln(n!)$ mediante números de Bernoulli. Bastante precisa para sumatorias porque corrige con derivadas, pero necesita funciones suaves y cálculo de los números B_{2k} [23, 24].

El comportamiento asintótico de una función factorial $n!$ y sus aproximaciones son importantes para entender su crecimiento cuando n es grande, $n!$ es aproximadamente exponencial-exponencial, es decir, crece más rápido que cualquier función exponencial a^n, pero más lento que funciones como n^n [17].
Por lo tanto, el factorial puede acotarse mediante funciones más simples definiendo cotas para estos:
Cota inferior: $(\frac{n}{e})^n \leq n!$ (demostrable usando $e^x \geq x + 1$).
Cota superior: $n! \leq n^n$, ya que

$$n! = n \times (n-1) \times \cdots \times 1 \leq n \times n \times \cdots \times n = n^n.$$

Una cota más ajustada es:

$$(\frac{n}{e})^n \leq n! \leq e\sqrt{n}(\frac{n}{e})^n$$

3.2. Aproximaciones Clásicas y Modernas.

La fórmula de Stirling, es un método de aproximación clásica para estimar factoriales y valores de la función Gamma $\Gamma(n+1)$.
La fórmula completa es la siguiente:

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

Su versión más básica es rápida, pero poco precisa con números pequeños, aunque mejora significativamente al aumentar el valor de n, y tiene un error relativo de aproximadamente:

$$Error_{relativo} \sim \frac{1}{12n}.$$

Muy útil en física estadística y probabilidad cuando n es grande [6].

Ejemplo 3.11. Calcule el factorial para $n = 10$ usando la formula de Stirling.
Solución. Sabemos que $10! = 3,628,800$ y para aproximar mediante la formula de Stirling con $n = 10$ tenemos.

$$\sqrt{2\pi \cdot 10}\left(\frac{10}{e}\right)^{10} = 3,598,695{,}62$$

Donde la formula de Stirling nos da una aproximación con un 99,17 % de precisión y un error relativo de 0,83 %.

Más tarde, se desarrolló una versión refinada que incluye un término adicional, ofreciendo mayor exactitud incluso con números moderados, ideal para aplicaciones prácticas donde se necesita un equilibrio entre simplicidad y precisión.
La fórmula de Stirling+1, surgió como una extensión con un término correctivo y viene dada por:

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n}\right),$$

esta versión mejora la aproximación básica, como también mejora significativamente la precisión para n pequeños o moderados y reduce el error a:

$$Error_{relativo} \sim \frac{1}{288n^2},$$

siendo ideal para aplicaciones prácticas en estadística y optimización [24].

Ejemplo 3.12. Calcule el factorial para $n = 10$ usando la fórmula de Stirling+1.
Solución. Para aproximar mediante la fórmula de Stirling+1 con $n = 10$ tenemos.

$$\sqrt{2\pi \cdot 10}\left(\frac{10}{e}\right)^{10} \left(1 + \frac{1}{12n}\right) = 3,628,684{,}75$$

Donde la formula de Stirling+1 nos da una aproximación con un 99,997 % de precisión y un error relativo de 0,003 %.

La función Gamma $\Gamma(z)$ es la herramienta teórica exacta por excelencia, ya que generaliza el factorial a números complejos y no enteros, esta función no es una aproximación sino una definición exacta del factorial, para cálculos exactos o generalizaciones [12].
Cuando se requieren valores exactos o extensiones a números complejos y no enteros. Es fundamental en matemáticas avanzadas y física teórica, aunque su cálculo requiere métodos numéricos en la mayoría de los casos.

Ejemplo 3.13. Calcule el factorial para $n = 10$ usando la función Gamma.
Solución. Para aproximar mediante la función Gamma con $n = 10$ tenemos.

$$\Gamma(n+1) = \int_0^\infty t^{(n+1)-1} e^{-t} dt$$
$$\Gamma(10+1) = \int_0^\infty t^{(10+1)-1} e^{-t} dt = 3,628,800.$$

Donde la función Gamma nos da el valor exacto de 10!.

Finalmente, la fórmula de aproximación de Ramanujan es una alternativa más precisa a la fórmula de Stirling, especialmente útil para valores de n pequeños o moderados.
La fórmula es la siguiente:

$$n! \approx \sqrt{\pi}\left(\frac{n}{e}\right)^n \left(8n^3 + 4n^2 + n + \frac{1}{30}\right)^{\frac{1}{6}}.$$

Aunque Stirling es práctica para estimaciones rápidas, la aproximación de Ramanujan, basada en una profunda intuición matemática, es la opción preferida cuando se busca la máxima exactitud [16].

Ejemplo 3.14. Calcule el factorial para $n = 10$ usando la aproximación de Ramanujan.
Solución. Para aproximar mediante la aproximación de Ramanujan con $n = 10$ tenemos.

$$\sqrt{\pi}\left(\frac{n}{e}\right)^n \left(8n^3 + 4n^2 + n + \frac{1}{30}\right)^{\frac{1}{6}} = 3,628,800{,}31.$$

Donde la aproximación de Ramanujan nos da una solución casi exacta con un error de $10^{-7}\,\%$.

En resumen, cada método tiene su lugar: la fórmula de Stirling es práctica para estimaciones rápidas, su versión refinada mejora la precisión sin complicar demasiado los cálculos, la función Gamma es la herramienta teórica por excelencia, y la aproximación de Ramanujan ofrece la mayor exactitud para casos que lo requieren. Con los métodos anteriores podemos resumir lo siguiente:

CUADRO 1. Comparación de métodos de aproximación factorial

Método	**Precisión** $n = 1$	**Precisión** $n = 10$	**Complejidad**
Stirling básica	$\sim 8\,\%$ error	$\sim 0{,}8\,\%$ error	Muy simple
Stirling+1 término	$\sim 0{,}7\,\%$ error	$\sim 0{,}008\,\%$ error	Simple
Ramanujan	$\sim 0{,}01\,\%$ error	$\sim 10^{-7}\,\%$ error	Menos simple
Función Gamma	Exacta	Exacta	Evaluación numérica

A continuación se muestran tablas comparativas para los valores de $n = 1$ y $n = 10$.

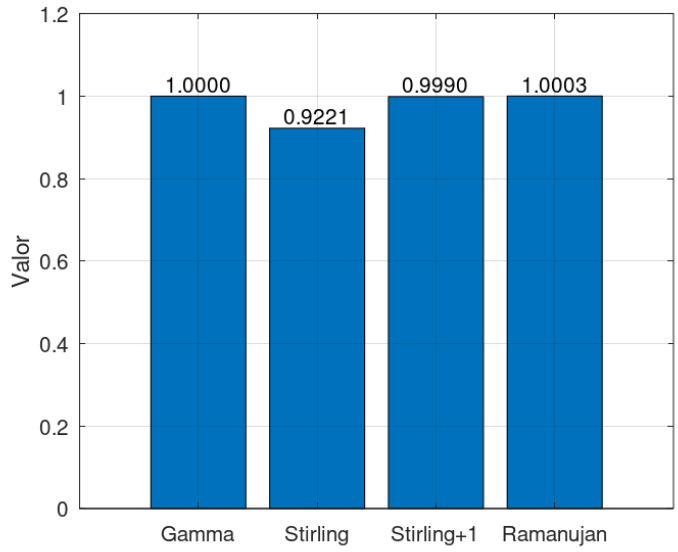


FIGURA 1. Comparación de aproximación para $n = 1$: Gamma, Stirling, Stirling+1 y Ramanujan.

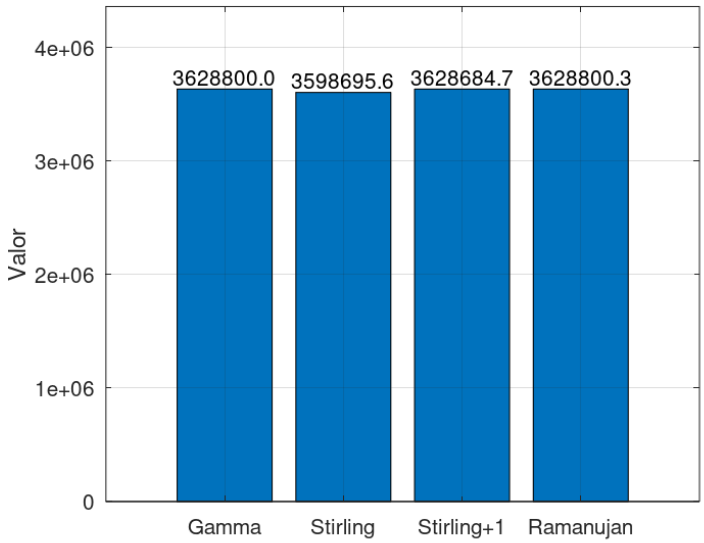


FIGURA 2. Comparación de aproximación para $n = 10$: Gamma, Stirling, Stirling+1 y Ramanujan.

4. CONCLUSIONES

El análisis asintótico de la función factorial nos ha permitido explorar distintas aproximaciones, cada una con sus ventajas y limitaciones. La clásica fórmula de Stirling, aunque sencilla y elegante, muestra errores significativos para valores pequeños de n. En cambio, las aproximaciones refinadas, como las de Ramanujan, logran una precisión impresionante incluso con n bajo, gracias a la inclusión de términos de corrección más detallados. Esto nos lleva a una conclusión práctica: la elección de la aproximación depende del contexto. Para desarrollos teóricos o análisis de comportamiento en el límite, Stirling es suficiente y más manejable. Pero en aplicaciones numéricas donde cada decimal cuenta, como en simulaciones o algoritmos de alta precisión, las fórmulas mejoradas o las de Ramanujan son la mejor opción.

Este estudio sugiere varias direcciones prometedoras para futuras investigaciones:

1. Extensión a funciones generalizadas: Aplicar estos métodos a la función Gamma incompleta y versiones multivariadas, con potenciales aplicaciones en estadística bayesiana y física teórica.
2. Análisis en el plano complejo: Investigar el comportamiento asintótico cuando z es complejo, particularmente para la función Gamma $\Gamma(z)$, lo que podría tener implicaciones en teoría de números y análisis complejo.
3. Optimización computacional: Desarrollar algoritmos eficientes basados en estas aproximaciones para el cálculo de factoriales grandes o coeficientes binomiales, relevante en criptografía y procesamiento de datos a gran escala.

4. Enfoques heurísticos al estilo de Ramanujan: Explorar métodos empíricos que combinen intuición y rigor matemático para descubrir nuevas aproximaciones de alta precisión, inspirados en el legado de Ramanujan.
5. Aplicaciones interdisciplinarias: Adaptar estas técnicas a problemas en machine learning como la optimización de hiperparámetros.

Cada una de estas líneas no solo profundizaría nuestro entendimiento teórico, sino que también podría generar herramientas prácticas para desafíos científicos y tecnológicos actuales.

En definitiva, el estudio del factorial y sus aproximaciones no es solo un ejercicio teórico, sino una herramienta poderosa con aplicaciones prácticas en múltiples áreas. A medida que avanzan las técnicas numéricas y las necesidades computacionales, seguirán surgiendo mejoras y generalizaciones que enriquecerán aún más este campo.

Referencias

1. C. Kramp, *prefacio a Elements d'arithmétique Universelle*(1808), págs. V-VI y XI-XII.
2. F. W. J. Olver, *Asymptotics and Special Functions*,Academic Press, 1974.
3. Davis, Philip J. *Leonhard Euler's Integral: A Historical Profile of the Gamma Function.* American Mathematical Monthly,(1959). 66 (10): 849–869.
4. G. Andrews, R. Askey, R. Roy, *Special Functions*, Cambridge University Press, 1999.
5. N. M. Temme, *Special Functions: An Introduction to the Classical Functions of Mathematical Physics* 1996, Wiley.
6. Hazewinkel, Michiel, ed. *Fórmula de Stirling*, Encyclopaedia of Mathematics (en inglés),2001, Springer,
7. McKie, Robin. *A brief history of Stephen Hawking.* Cosmos.(August 2007) Retrieved 13 June 2020.
8. UNAH. *Prioridades de investigación: ejes y temas prioritarios*, 2015-2019.
9. Le Cam, L. *El teorema del límite central alrededor de 1935*,Statistical Science,1986,78–96.
10. Pearson, Karl, *Nota histórica sobre el origen de la curva normal de errores*,Biometrika, 1924, 402–404
11. Weisstein, Eric W. *Gaussian Integral*. MathWorld (en inglés).
12. Davis, Philip J. *Leonhard Euler's Integral: A Historical Profile of the Gamma Function*. Am. Math. Monthly. (1959), (66): 849-869.
13. Stigler, Stephen M. *Poisson on the Poisson Distribution*. Statistics y Probability Letters(1982). 1 (1): 33-35.
14. Karatsuba, Ekatherina A. *Sobre la representación asintótica de la función gamma de Euler por Ramanujan*, Journal of Computational and Applied Mathematics, (2001), 135 (2): 225–240
15. Mortici, Cristinel. *Sobre la fórmula del argumento grande de Ramanujan para la función gamma*, Ramanujan J. (2011), 26 (2): 185– 192
16. Chao-Ping Chen. *Padé Approximant Related to Ramanujan's Formula for the Gamma Function*, Results Math, (2018),73(3), 107
17. Estrada, R. Kanwal, RP. *A distributional approach to asymptotics: theory and applications Series* , Birkhäuser Advanced Texts Basler Lehrbücher 2002.
18. Ricardo Estrada, Ram P. Kanwal. *A Distributional Approach to Asymptotics*, Springer Science and Business Media, Sep 8, 2012
19. Dilcher, K. Skula, L. Slavutskii, I. Sh. *Números de Bernoulli*. Queen's Papers in Pure and Applied Mathematics (1713-1990), (87).
20. Ralph P. Grimaldi. *Matemáticas discretas y combinatoria : una introducción con aplicaciones*. Pearson Educación, (7) 1998
21. Artin, Emil. *The Gamma function. History of Mathematics* American Mathematical Society, (30), 2006.
22. Wang, Xiang-Sheng; Wong, Roderick. *Análogos discretos de la aproximación de Laplace*. Asymptot. Anal . 54 (3-4): 165-180, 2007.
23. Gould, Henry W. *Fórmulas explícitas para números de Bernoulli*, Amer. Math. Monthly , 79 (1): 44–51, 1972.
24. Bender, C. M. and Orszag, S. A., *Advanced Mathematical Methods for Scientists and Engineers*, Springer (1999), Cap. 5.9.

Dirección de correo electrónico: `henrry.morazan@unah.hn`

APLICACIÓN DE REDES NEURONALES Y MODELOS LINEALES ESTADISTICOS PARA EL ANÁLISIS DE DATOS EXTREMOS EN EVENTOS CLIMÁTICOS

RONALD GUSTAVO ORELLANA REYES

Resumen. Este estudio enfoca la aplicación de Redes Neuronales Perceptrones Multicapa (MLP) para el análisis de eventos climáticos extremos en Honduras, utilizando datos diarios de temperatura y precipitación proporcionados por el Instituto Hondureño de Ciencias de la Tierra (IHCIT) de la UNAH. Utilizando Redes Neuronales Perceptrones Multicapa (MLP), que transforman las series de tiempo en un formato tabular mediante la transformación de datos para optimizar las variables. Esto permite a las MLP modelar patrones complejos de forma no secuencial, diferenciándose de enfoques recurrentes. Los resultados de esta investigación son fundamentales para validar la viabilidad de implementar esta metodología en diferentes estaciones meteorológicas, lo que proporcionará una visión más exacta del comportamiento de la precipitación. Estos hallazgos pueden orientar directamente la formulación de estrategias de adaptación y mitigación climática en Honduras, ofreciendo una contribución valiosa para enfrentar este desafío regional.

ABSTRACT. This study focuses on the application of Multilayer Perceptron (MLP) Neural Networks for the analysis of extreme climate events in Honduras. It uses daily temperature and precipitation data provided by the Honduran Institute of Earth Sciences (IHCIT) at UNAH. The methodology employs MLPs to transform time series data into a tabular format by optimizing variables through data transformation. This approach allows MLPs to model complex, non-sequential patterns, distinguishing them from recurrent methods. The results of this research are fundamental for validating the feasibility of implementing this methodology at various weather stations, which will provide a more accurate understanding of precipitation behavior. These findings can directly guide the formulation of climate adaptation and mitigation strategies in Honduras, offering a valuable contribution to tackling this regional challenge.

1. INTRODUCCIÓN

El cambio climático es una verdad absoluta siendo uno de los mayores desafíos en nuestra era actual, con profundas consecuencias en la economía global, sectores como la agricultura como, cambios en la calidad de los cultivos, ecosistemas y más importante en la vida de cada persona. En concreto, la frecuencia e intensidad de los eventos climáticos extremos, como sequías prolongadas, inundaciones y huracanes más devastadores están en aumento [6,7], presentando una amenaza constante para la estabilidad y el desarrollo de muchas regiones del mundo y principalmente en territorio hondureño.

Fecha: Julio del 2025.
Palabras y frases clave : GLM, Valores Extremos, Machine Learning, Cambio Climático.

Comprender la naturaleza de estos fenómenos y desarrollar herramientas robustas es esencial para construir resiliencia y mitigar el impacto de alguna manera. Los modelos lineales han sido una base para el análisis de datos, sin embargo, cuando se estudia datos de eventos climáticos extremos, sus características a menudo violan los supuestos de linealidad y normalidad que subyacen a estos modelos. Los datos extremos suelen ser asimétricos, con varianzas que dependen de la media, presentando una alta tasa de valores atípicos, lo que limita la capacidad de los modelos lineales clásicos para capturar con precisión su comportamiento. Es aquí donde los modelos lineales generalizados GLM emergen como una herramienta metodológica prometedora [1,2]. Los GLM ofrecen una flexibilidad inherente al permitir el uso de diferentes distribuciones de probabilidad y funciones de enlace, lo que hace que sea adecuado para modelar la complejidad y las distribuciones no Gaussianas inherentes a los valores extremos [8,15].

En este estudio nos adentraremos en la aplicación de diversos enfoques GLM para la estimación y análisis de eventos climáticos extremos en Honduras. Usaremos como parámetro indicador del cambio climático la temperatura y sus variaciones desde el año 1948 hasta 2017, con el objetivo de hacer una revisión exhaustiva de las características estadísticas de las variables, temperatura, humedad, altitud, región, tipo de suelo. Posteriormente aplicar modelos GLM, con el análisis y cálculo de valores extremos con el fin de estimar y predecir la ocurrencia o magnitud de tales eventos, usando diferentes familias de distribuciones (como Poisson, Binomial Negativa o Gamma) y funciones de enlace dentro del marco GLM, permitiendo ajustar así el modelo a la naturaleza específica de los valores extremos.

A pesar de la robustez de los GLM, la complejidad de los sistemas climáticos y la vasta cantidad de datos disponibles han abierto la puerta a métodos más avanzados como técnicas de Machine Learning y en particular las Redes Neuronales(NN) [14], capturando relaciones dinámicas y dependencias temporales en las series de datos climáticas. En este estudio se aprovechará la combinación de la solidez de los GLM con la capacidad de aprendizaje de patrones de la Redes Neuronales para tener una mejor capacidad de análisis y predicción de los mismos. El proceso de selección y comparativa entre diversos métodos de evaluación de modelos será crucial, estos modelos serán evaluados utilizando criterios de información como el Criterio de Información de Akaike (AIC) y el Criterio de Información Bayesiana (BIC) [22], de manera simultánea, se entrenarán y validarán diferentes arquitecturas de redes neuronales, optimizando sus hiperpárametros.

Los resultados de esta evaluación nos permitirán determinar el modelo óptimo para la estimación y predicción de eventos climáticos e iluminarán las numerosas implicaciones y significativos beneficios que esta conlleva, tales como preveer información para la toma de decisiones. fomentar el avance cinetífico y metodológico, y contribuir al desarrollo sostenible, el ambiente y la adaptación al cambio climático.

Este estudio aplica los Modelos Lineales Generalizados (GLM) y Redes Neuronales (NN) [9,10] un sistema computacional de aprendizaje supervisado, para analizar y proponer soluciones a los eventos climáticos en Honduras, siendo este parte de las áreas de investigación de la UNAH, especificamente en el tema prioritario de

Cambio Climático y Vulnerabilidad, y buscará comparar los resultados con investigaciones previas que utilizarón modelos lineales dinámicos.

2. Antecedentes

Antes de la llegada de los GLM modelos lineales generalizados, el análisis estadístico estaba anclado a la regresión lineal clásica. Esta poderosa herramienta imponía supuestos restrictivos, como la normalidad de los residuos y una relación lineal entre las variables, que a menudo no se cumplen en el mundo real, especialmente en eventos climáticos. Para abordar estas restricciones, los investigadores recurrían a transformaciones de datos, que es un proceso tedioso.

Un momento crucial en la estadística fue la publicación Generalizad Linear Models por John Nelder y Robert Wedderburn en 1972 [1]. Este trabajo innovador no solo marcó el nacimiento, sino que también unificó técnicas estadísticas como la regresión logística, regresión de Poisson, regresión Gamma, bajo un marco conceptual coherente. La propuesta radicó en identificar una estructura fundamental, una función de enlace que conectaba la media de la variable con un predictor lineal y una variable de respuesta cuya distribución pertenecía a la familia exponencial [2]. El 1970, el software GLIM (Generalizad Linear Interactive Modelling), facilitó aún más su aplicación, consolidando a los GLM como una herramienta indispensable en diversos campos como, biología, economía, y para estudios de ciencias ambientales [3, 4].

La comprensión y preparación floreció en el siglo XX. Su auge vino con Emil Gumbel en la década de 1950, quien sentó las bases de la teoría de los valores extremos (TVE), su trabajo y la de otros pioneros llevaron el desarrollo de estas distribuciones y posteriormente a la distribución generalizada de Pareto, esenciales para estimar la probabilidad y magnitud de fenómenos raros, como inundaciones y sequías [5,6]. Estas distribuciones se volvieron esenciales para estimar la probabilidad de ocurrencia y la magnitud de eventos poco frecuentes, como inundaciones históricas o sequías prolongadas, proporcionando información vital para la planificación de infraestructura y la gestión de desastres. A medida que la disponibilidad de datos climáticos de largo plazo y la capacidad computacional aumentaron, la aplicación de TVE en climatología se intensifico con más datos y capacidad computacional [7,8].

La versatilidad de los modelos Lineales Generalizados (MLG) permite modelar características de eventos extremos que no se ajustan a TVE, como la frecuencia de días con temperaturas extremas o los parámetros de las distribuciones de valores extremos [9]. La complejidad de los sistemas climáticos y la gran cantidad de datos han impulsado la emergencia de "machine learning" (ML) en redes neuronales (NN).

El concepto de redes neuronales artificiales se remonta a 1943 con el trabajo de Warren McCulloch y Walter Pitts, quienes desarrollaron el primer modelo computacional de neuronas, demostrando cómo podían realizar funciones lógicas simples [10]. Posteriormente, en 1958, Frank Rosenblatt introdujo el Perceptrón, un modelo que simulaba la capacidad de aprendizaje y que despertó un gran entusiasmo inicial [11]. Sin embargo, la investigación en NN experimentó un período de "invierno" en las décadas de 1970 y 1980 debido a las limitaciones computacionales y a las críticas

sobre las capacidades del Perceptrón, notablemente por Minsky y Papert en 1969 [12]. El resurgimiento de las redes neuronales se produjo a partir de la década de 1980, impulsado por avances teóricos como el desarrollo del algoritmo de retropropagación (backpropagation), el aumento exponencial de la potencia computacional y la disponibilidad de grandes volúmenes de datos. Figuras como John Hopfield en 1982 y el trabajo de Teuvo Kohonen con sus mapas autoorganizados en 1982, renovaron el interés en el campo [13, 14]. Desde entonces, las Redes Neuronales han demostrado una capacidad sin igual para aprender patrones no lineales y dependencias complejas en grandes conjuntos de datos, lo que las hace particularmente adecuadas para modelar la intrincada dinámica de los sistemas climáticos y sus series temporales [15]. En el análisis de eventos extremos, las NN ofrecen la posibilidad de capturar relaciones sutiles entre variables climáticas (temperatura, humedad, altitud, etc.) y la ocurrencia o magnitud de eventos extremos, complementando los enfoques estadísticos paramétricos al proporcionar una perspectiva de modelado basada en datos para la predicción y clasificación de fenómenos extremos.

3. Modelado de Eventos Climáticos Extremos con Redes Neuronales No Secuenciales

3.1. Preliminares. En esta sección, exploraremos conceptos fundamentales para comprender el análisis de eventos climáticos extremos mediante redes neuronales (NN). Discutiremos la naturaleza de los datos climáticos la importancia de los valores extremos y la capacidad de las redes neuronales para modelar estos fenómenos. Aunque nuestros datos se originan como series temporales (temperatura a lo largo del tiempo), nuestro enfoque se centrará en transformar estas secuencias en un formato tabular de características para ser procesadas por redes neuronales no secuenciales, permitiendo así una comparación directa con estudios previos que emplearon modelos lineales dinámicos.

3.1.1. La Naturaleza de los Datos Climáticos y Valores Extremos. Los datos climáticos, como la temperatura, la humedad y la precipitación, suelen recopilarse a lo largo del tiempo, formando series temporales. Este conjunto de observaciones sobre los valores que toma una variable cuantitativa en diferentes momentos del tiempo, denotados como $Y_1, Y_2, Y_3, ..., Y_T$, donde Y_t es el valor en el instante t, pueden exhibir diversos comportamientos [13].

Es común observar la presencia de una tendencia (comportamiento predominante o cambio de la media a largo plazo), un ciclo (oscilaciones de larga duración alrededor de la tendencia, como fenómenos climáticos que durán varios años), o variaciones estacionales (movimientos periódicos dentro de un periodo corto y conocido, influenciados por factores institucionales y climáticos). El componente aleatorio representa los movimientos erráticos e impredecibles que no siguen un patrón específico [13].
Sin embargo, para nuestro análisis, el interés recae en los eventos o valores extremos, que representan desviaciones significativas de las condiciones promedio y tienen impactos sustanciales. Modelar estos eventos directamente con la secuencia temporal puede ser complejo, especialmente cuando se busca un contraste con metodologías que no se basan puramente en la dinámica secuencial.
La razón por la que los datos de valores extremos pueden ser difíciles de manejar con modelos lineales clásicos radica en sus propiedades estadísticas. A menudo,

presentan distribuciones asimétricas, con varianzas que dependen de la media, y una alta tasa de valores atípicos [6,7]. Estas características violan supuestos clave de los modelos lineales tradicionales, lo que limita su capacidad para capturar con precisión el comportamiento de estos fenómenos críticos. Es por ello que enfoques más flexibles como los Modelos Lineales Generalizados (GLM) [1,2,8,15], y especialmente los modelos de "Machine Learning" como las Redes Neuronales, son particularmente adecuados para este tipo de datos.

3.2. Modelos Tradicionales en Análisis de Series de Tiempo Climáticas. Para contextualizar el concepto de este estudio, es importante reconocer las metodologías que históricamente han sido empleadas en el análisis predictivo de series de tiempo, que puede descomponerse en cuatro componentes que no son directamente observables, de los cuales únicamente se pueden obtener estimaciones. Estas cuatro componentes son: tendencia, ciclo, estacionalidad y aleatoriedad.

3.2.1. Modelos Estadísticos ARIMA. Originados en la economía, los modelos ARIMA son valorados por sus características estadísticas, su capacidad de implementar un rango de modelos exponenciales que aportan suavidad, y por la integración del método de Box-Jenkins durante la fase de entrenamiento [16]. La comprensión de estos modelos a menudo requiere el concepto del operador de retrasos B, donde para una serie de tiempo Y_t, la serie con retraso se denota por $BY_t = Y_t - 1$, y análogamente $B_k Y_t = Y_t - k$ [16]. Un modelo ARIMA (p, d, q), sus respectivos parametros puede representarse mediante la ecuación:

$$(1 - B)^d Y_t = \mu + \phi(B)(1 - B)^d Y_t = \mu + \phi(B)Z(t) + \Theta(B)\epsilon_t$$

Donde (p, d, q) son los órdenes de las componentes autorregresiva, integrada y de media móvil, respectivamente μ es una constante, $\phi(B)$ y $\theta(B)$, son polinomios en el operador de retrasos, $Z(t)$, es el proceso estacionario resultante de la diferenciación, y ϵ_t es un término de error de ruido blanco.

3.3. Redes Neuronales. Las Redes Neuronales Artificiales (RNA o NN) son un tipo de inteligencia artificial que intenta imitar la forma en que el cerebro humano funciona [9]. En lugar de utilizar un modelo digital basado en operaciones booleanas, una red neuronal opera creando conexiones entre elementos procesados, que son el equivalente computacional de las neuronas [12,14].
Las redes neuronales son particularmente efectivas en la predicción cuando se dispone de una base de datos grande previa a las situaciones a predecir. Son un conjunto de algoritmos modelados holgadamente en el cerebro humano, diseñados para reconocer patrones. Interpretan datos sensoriales a través de percepción de máquina, etiquetando o agrupando los datos de entrada. Los patrones reconocidos son numéricos, contenidos en vectores, hacia los cuales todos los datos reales (sean imágenes, sonido, texto o, para nuestro caso, datos climáticos) deben ser traducidos [14, 24]. Inspirándose en el cerebro biológico, una red neuronal artificial es una colección de unidades conectadas, también llamadas neuronas. La conexión entre las neuronas puede transportar señales entre ellas. Cada conexión transporta un valor de número real, el cual determina la fuerza de la señal [15]. A un nivel más fundamental, una neurona tiene entradas a través de sus dendritas y transmite señales de

salida mediante un axón, que entra en contacto con otras neuronas en uniones especializadas llamadas sinapsis, pasando sus señales de salida para repetir el proceso.

Los sistemas neuronales artificiales (ANS) se inspiran en la estructura del sistema nervioso biológico para desarrollar sistemas de procesamiento paralelos, distribuidos y adaptativos, que pueden exhibir cierto grado de comportamiento inteligente. Las neuronas artificiales se organizan en capas que forman una red neuronal. Estas redes neuronales, ya sea individualmente o interconectadas, pueden constituir sistemas de procesamiento global. La siguiente figura presenta un modelo básico de una neurona artificial y su correspondencia con la neurona biológica. Este modelo consta de los siguientes elementos:

- Un conjunto de pesos sinápticos w_i asociadas con cada una de las entradas $x(t)$.
- Una regla de propagación, típicamente expresada como la suma ponderada de las entradas : $h_i(t) = \sum w_i j x_j(t)$.
- Una función de activación f_i para calcular la salida de la neurona : $y_i(t) = f_i(h_i(t))$.

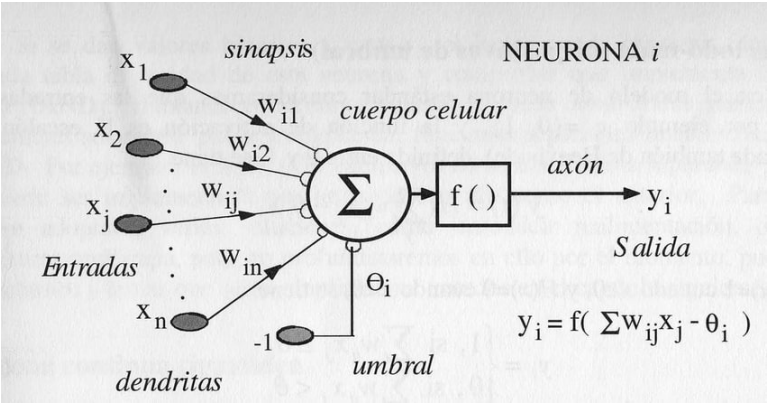


FIGURA 1. Modelo Estándar [24]

3.3.1. Deep Learning y Redes Neuronales Profundas. El deep learning (aprendizaje profundo) es una forma de "machine learning" que permite a las computadoras aprender a partir de la experiencia y entender el mundo en términos de una jerarquía de conceptos. A medida que la computadora recopila información de experiencia, no es necesario que un operador humano especifique el conocimiento requerido. La jerarquía de conceptos permite que la computadora aprenda conceptos complicados construyéndolos a partir de conceptos más sencillos [14].

Los modelos de aprendizaje profundo tienen la ventaja de ser capaces de descubrir automáticamente estructuras intrincadas en datos de altas dimensiones, una tarea que comúnmente requiere un trabajo manual exhaustivo en el caso de métodos como el bosque aleatorio. Estos modelos se construyen utilizando múltiples capas de neuronas artificiales o nodos, diseñadas para replicar la estructura y funcionamiento del cerebro humano. Al procesar datos y crear patrones para la toma de decisiones, estos modelos son la clave tecnológica detrás de aplicaciones avanzadas,

incluyendo el procesamiento de lenguaje natural, reconocimiento de voz, visión por computadora y bioinformática [16].

3.3.2. Funciones de activación. Una función de activación, que simultáneamente representa la salida de la neurona y su estado de activación, se denota como $y_i = f_i(h_i) = f_i(\sum_{j=o}^{n} w_i j x_j)$. Existen varios ejemplos de funciones de activación, entre ellos:

- Neurona todo-nada.
- Neurona continua sigmoidea.

3.3.3. Arquitecturas Relevantes para este estudio. Para representar adecuadamente las múltiples entradas de una neurona, se emplea notación matricial para los pesos W. Cada entrada individual $p_1, p_2, ..., p_R$ está asociada con pesos específicos $w_{1,1}, w_{1,2}, ..., w_1, R$ en la matriz de pesos W. La salida n puede expresarse mediante la siguiente ecuación:

$$n = w_1, 1_{P1} + w_1, 2_{P2} + ... + w_1, R_{PR} + b$$

Red Neuronal Multicapa (MLP):También conocida como red neuronal densa o feed-forward, es una arquitectura fundamental en el aprendizaje automático. Su principal característica es que la información fluye en una única dirección, desde la capa de entrada hasta la de salida, sin bucles de retroalimentación. Este modelo opera como una función no lineal que aprende y compone las transformaciones de sus neuronas para identificar relaciones complejas en los datos. Aunque la topología de la red puede variar, la configuración más utilizada se organiza en capas discretas: una capa de entrada, una o varias capas ocultas y una capa de salida. Esta estructura le permite modelar una amplia gama de problemas de predicción y clasificación, cuya arquitectura se muestra en la figura 1.

Las matrices de peso conectadas a las entradas se denominan matrices de ponderación o matriz de pesos de entrada, mientras que aquellas que utilizan la salida proveniente de otra capa se denominan matrices de pesos. Los superíndices identifican la fuente (segundo índice) y el destino (primer índice) para distinguir los pesos y otros componentes de la red. En la Figura 2 se presenta una red neuronal de entrada múltiple compuesta por una capa. Aquí, S representa la cantidad de neuronas de la capa 1, y R denota el número de elementos del vector de entrada. Se observa que para la matriz de ponderación o pesos conectada al vector de entrada p, se le asigna $I_{1,1}$, donde el segundo índice indica la fuente 1 y el primer índice también es 1. Los sesgos, las entradas netas y las salidas de la capa 1 tienen un superíndice 1, indicando que pertenecen a la primera capa.

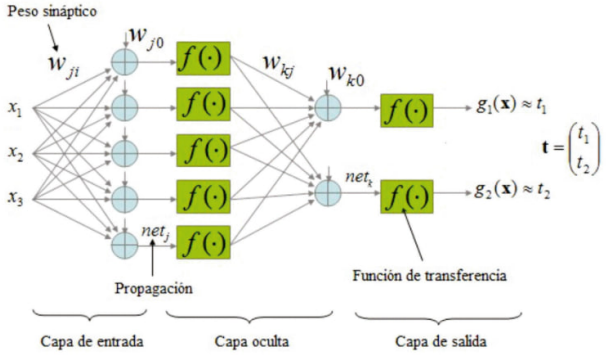


FIGURA 2. Perceptrones Multicapa

3.4. Resultados y experimentación. Se trabajó con un conjunto de datos climáticos reales, los cuales fueron recopilados como un promedio diario y estructurados en una base de datos tabular, para la fase de experimentación, se adoptó una estrategia de validación basada en la división por año, los datos se usaron para el entrenamiento de modelos, para la predicción y evaluación de su rendimiento, haciendo uso de Rstudio para entrenamiento del modelo, para entrenar el modelo, obtener diagramas y gráficos de conclusión.

El comportamiento de la Temperatura Máxima Promedio (3 días) a lo largo de los años, se observa un patrón estacional recurrente, con picos de temperatura en los meses de verano y descensos en los de invierno. A diferencia de los modelos tradicionales de series de tiempo como ARIMA, nuestro enfoque utiliza una red neuronal multicapa no secuencial para capturar este comportamiento periódico.

Nº	**Año**	**Mes**	**Día**	**PRCP**	**T_max**	**T_min**
1	1948	1	1	0.47	51	42
2	1948	1	2	0.59	45	36
3	1948	1	3	0.42	45	35
4	1948	1	12	0	41	26
5	1948	1	13	0	45	29

Nº	**Rain**	**T_max (3d)**	**PRCP (3d)**	**EE**
1	1		1.48	1
2	1		1.32	0
3	1	47	0.90	0
4	0	46	0.92	1
5	0	46.67	1.02	0

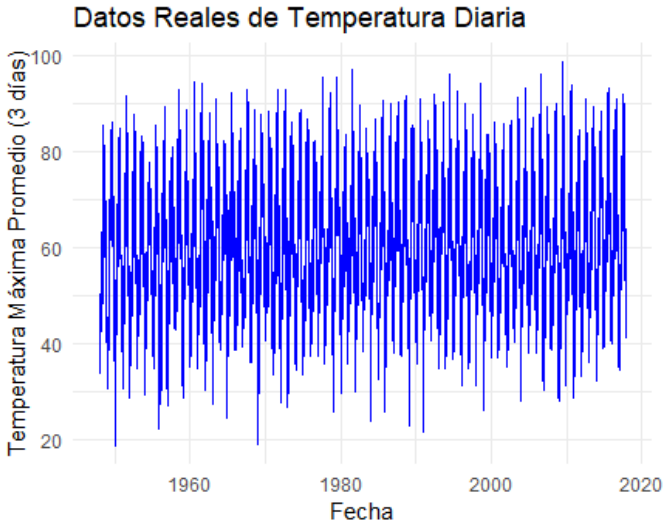


FIGURA 3. Datos de temperatura diaria

En lugar de modelar la dependencia directa del tiempo, el modelo utiliza variables como el día del año para entender y replicar la estacionalidad de la temperatura, lo que le permite hacer una estimación precisa de los valores diarios a partir de las características de la fecha y de otras variables climáticas.

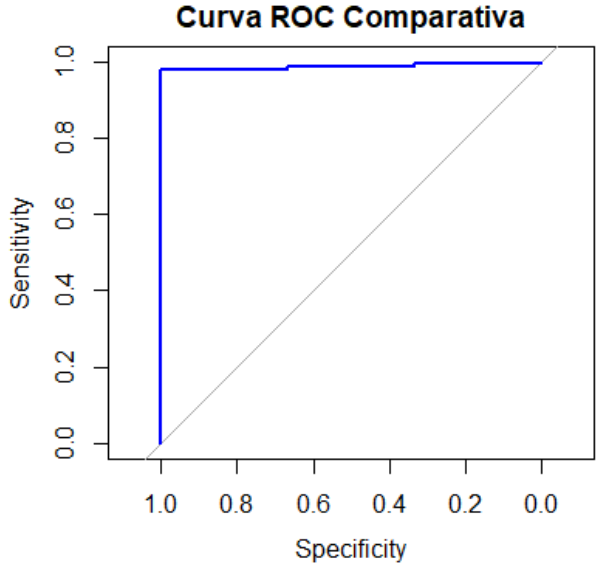


FIGURA 4. Análisis del Rendimiento del Modelo SVM

A diferencia de los modelos ARIMA que exhiben limitaciones en la predicción a largo plazo, el modelo SVM demuestra un rendimiento robusto en la tarea de

clasificación. La Curva ROC del Modelo SVM (como se ilustra en el gráfico) muestra una trayectoria casi ideal, con un Área Bajo la Curva (AUC) cercana a 1. Esto indica que el modelo logra una alta sensibilidad y especificidad de manera simultánea, lo que se traduce en una capacidad excepcional para identificar eventos climáticos extremos sin generar un número significativo de falsos positivos.

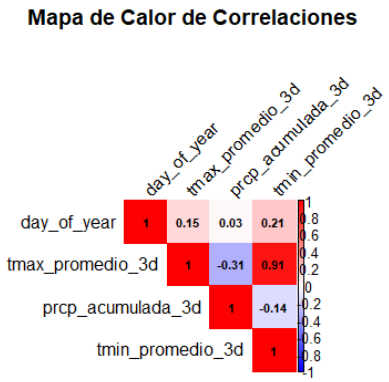


FIGURA 5. Mapa de Calor

El análisis de correlación usando la red neuronal, confirma que la temperatura máxima se relaciona fuertemente con variables temporales, esto demuestra que el modelo aprende la estacionalidad eficazmente, validando así el enfoque no secuencial.

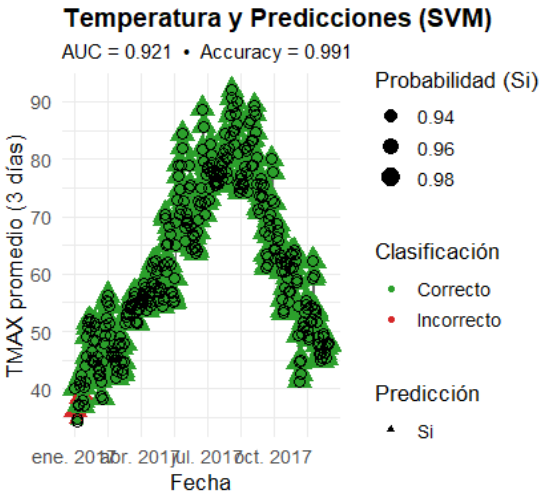


FIGURA 6. Temperatura y Predicciones

Red Neuronal prediciendo la estacionalidad de la temperatura, siguiendo de cerca los datos reales.

4. Conclusiones

A continuación, se resume el trabajo realizado y los hallazgos observados durante la experimentación.

- El proyecto demostró que la red neuronal multicapa es altamente efectiva para la estimación de la temperatura, adaptándose bien a los patrones estacionales de los datos tabulares.
- El modelo Support Vector Machine (SVM) se validó como una herramienta robusta para la clasificación de eventos de temperatura extrema.
- El uso de datos tabulares con modelos no secuenciales es una alternativa viable y precisa a las metodologías de series de tiempo para el análisis predictivo del clima.
- El análisis predictivo del clima utilizando una metodología de aprendizaje automático sobre una base de datos tabular, demostró que las redes neuronales multicapa (feed-forward) son efectivas para la estimación de valores continuos como la temperatura, adaptándose exitosamente a la estacionalidad de los datos.

Referencias

[1] Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3), 370-384.

[2] Dobson, A. J., & Barnett, A. G. (2018). *An Introduction to Generalized Linear Models*. Chapman and Hall/CRC.

[3] Baker, R. J., & Nelder, J. A. (1978). *The GLIM System, Release 3*. Royal Statistical Society.

[4] Gumbel, E. J. (1958). *Statistics of Extremes*. Columbia University Press.

[5] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.

[6] Katz, R. W. (2010). Statistics of Extremes in Climate Change. *Climatic Change*, 100(1-2), 71-87.

[7] Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences*. Academic Press.

[8] Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3), 507-554.

[9] Larranaga, P., Inza, I., & Moujahid, A. (1997). Tema 8. Redes Neuronales. *Redes Neuronales, U. del P. Vasco*.

[10] McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115-133.

[11] Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386.

[12] Minsky, M. L., & Papert, S. A. (1969). *Perceptrons*. MIT Press.

[13] Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8), 2558–2562.

[14] Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1), 59–69.

[15] Harrison, M. J., Varghese, M. M., & Jones, P. L. (2018). *Applications of Machine Learning to Weather and Climate Modeling*. Cambridge University Press.

[16] McCulloch, C. E., & Searle, S. R. (2001). *Generalized, Linear, and Mixed Models*. John Wiley & Sons.

[17] Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3), 370-384.

[18] McCulloch, C. E., & Searle, S. R. (2001). Generalized, Linear, and Mixed Models. John Wiley & Sons.

[19] Baker, R. J., & Nelder, J. A. (1978). The GLIM System, Release 3. Royal Statistical Society.

[20] Gumbel, E. J. (1958). Statistics of Extremes. Columbia University Press.

[21] Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer.

[22] Katz, R. W. (2010). Statistics of Extremes in Climate Change. *Climatic Change*, 100(1-2), 71-87.

[23] Wang, Y., & Liu, Q. (2006). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of stock–recruitment relationships. *Fisheries Research*, 77(2), 220–226.

[24] Islam, M., Chen, G., & Jin, S. (2019). An overview of neural network. *American Journal of Neural Networks and Applications*, 5(1), 7–11.

[25] Ospina, R., Botto-Maire, R., Da Silva, R., De Oliveira, J. C., Silva, A. F., Andrade, F. B., & De Oliveira, M. L. (2023). An overview of forecast analysis with ARIMA models during the COVID-19 pandemic: Methodology and case study in Brazil. *Mathematics*, 11(14), 3069.

Referencias

[1] Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3), 370-384.

[2] Dobson, A. J., & Barnett, A. G. (2018). *An Introduction to Generalized Linear Models*. Chapman and Hall/CRC.

[3] Baker, R. J., & Nelder, J. A. (1978). *The GLIM System, Release 3*. Royal Statistical Society.

[4] Gumbel, E. J. (1958). *Statistics of Extremes*. Columbia University Press.

[5] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.

[6] Katz, R. W. (2010). Statistics of Extremes in Climate Change. *Climatic Change*, 100(1-2), 71-87.

[7] Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences*. Academic Press.

[8] Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3), 507-554.

[9] Larranaga, P., Inza, I., & Moujahid, A. (1997). Tema 8. Redes Neuronales. *Redes Neuronales, U. del P. Vasco*.

[10] McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115-133.

[11] Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386.

[12] Minsky, M. L., & Papert, S. A. (1969). *Perceptrons*. MIT Press.

[13] Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8), 2558–2562.

[14] Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1), 59–69.

[15] Harrison, M. J., Varghese, M. M., & Jones, P. L. (2018). *Applications of Machine Learning to Weather and Climate Modeling*. Cambridge University Press.

[16] McCulloch, C. E., & Searle, S. R. (2001). *Generalized, Linear, and Mixed Models*. John Wiley & Sons.

[17] Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3), 370-384.

[18] McCulloch, C. E., & Searle, S. R. (2001). Generalized, Linear, and Mixed Models. John Wiley & Sons.

[19] Baker, R. J., & Nelder, J. A. (1978). The GLIM System, Release 3. Royal Statistical Society.

[20] Gumbel, E. J. (1958). Statistics of Extremes. Columbia University Press.

[21] Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer.

[22] Katz, R. W. (2010). Statistics of Extremes in Climate Change. *Climatic Change*, 100(1-2), 71-87.

[23] Wang, Y., & Liu, Q. (2006). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of stock–recruitment relationships. *Fisheries Research*, 77(2), 220–226.

[24] Islam, M., Chen, G., & Jin, S. (2019). An overview of neural network. *American Journal of Neural Networks and Applications*, 5(1), 7–11.

[25] Ospina, R., Botto-Maire, R., Da Silva, R., De Oliveira, J. C., Silva, A. F., Andrade, F. B., & De Oliveira, M. L. (2023). An overview of forecast analysis with ARIMA models during the COVID-19 pandemic: Methodology and case study in Brazil. *Mathematics*, 11(14), 3069.