

Análisis del desempleo en Ecuador con modelos econométricos clásicos y propuesta usando Machine Learning

An Econometric Analysis of Unemployment in Ecuador: Classical Approaches and a Machine Learning-Based Proposal

Iván A. Narváez-P¹, Erick P. Herrera-Granda²

¹ Centro de Posgrado, Universidad Politécnica Estatal del Carchi Calle Antisana y Av. Universitaria, Tulcán, Ecuador

² Departamento de Informática y Ciencias de la Computación, Escuela Politécnica Nacional, Quito, Ecuador; Zip: 170525

Recibido: 12/06/2024, Aceptado: 30/07/2025

Autor de correspondencia: Andres Narvaez: andres.narvaez@upec.edu.ec

DOI: <https://doi.org/10.53358/ideas.v7i2.1293>



PALABRAS CLAVE

Modelado del desempleo para Ecuador,
Redes neuronales,
Deep learning,
regresión lineal.

RESUMEN

Este estudio se centra en modelar y predecir el desempleo en Ecuador mediante redes neuronales, contrastando su eficacia frente a métodos econométricos tradicionales. La variable dependiente, desempleo (y), definida como el porcentaje de la población activa sin empleo, se analiza en función de siete predictores macroeconómicos: Producto Interno Bruto real (PIB, x_1), inflación (x_2 , variación porcentual del índice de precios al consumidor), exportaciones (x_3) e importaciones (x_4 , valoradas en millones de USD), salario mínimo real (x_5), gasto público (x_6 , porcentaje del PIB) y tasa de interés (x_7). Investigaciones previas, basadas en la Ley de Okun, identificaron correlaciones lineales entre crecimiento del PIB y reducción del desempleo mediante Mínimos Cuadrados Ordinarios (MCO) y modelos de cointegración. Sin embargo, estas técnicas convencionales, incluyendo regresiones múltiples, mostraron limitaciones para capturar dinámicas no lineales inherentes al desempleo. En contraste, las redes neuronales (Neural Network (NN)) especialmente las arquitecturas secuenciales demostraron superioridad al modelar interacciones complejas entre variables. El estudio evalúa modelos de deep learning con una y dos capas ocultas, identificando una configuración óptima de una capa oculta (16 neuronas), alcanzando un loss de 0.2846, precisión del 75% y Error Cuadrático Medio (MSE) de 0.2064 en datos de prueba no vistos. Los resultados destacan la capacidad de las redes neuronales para modelar el desempleo con mayor precisión que métodos lineales, atribuyendo esto a su adaptabilidad para aprender patrones no lineales en datos multivariados. Este estudio demuestra que variables como el PIB, la inflación, el gasto público y la tasa de interés influyen significativamente en el desempleo. Las redes neuronales (NN) captan eficazmente estas relaciones complejas. Se resalta su utilidad para predecir fenómenos socioeconómicos en países en desarrollo como Ecuador, superando las limitaciones de modelos tradicionales.

KEYWORDS

Unemployment modeling to Ecuador,
Neural networks (NN),
Deep learning,
Linear regression.

ABSTRACT

This study focuses on modeling and predicting unemployment in Ecuador using neural networks, contrasting their efficacy against traditional econometric methods. The dependent variable, unemployment (y), defined as the percentage of the active population without employment, is analyzed in relation to seven macroeconomic predictors: real Gross Domestic Product (GDP, x_1), inflation (x_2 , percentage change in the Consumer Price Index), exports (x_3) and imports (x_4 , valued in millions of USD), real minimum wage (x_5), public spending (x_6 , percentage of GDP), and interest rate (x_7). Previous research, based on Okun's Law, identified linear correlations between GDP growth and unemployment reduction using Ordinary Least Squares (OLS) and cointegration models. However, these conventional techniques, including multiple regressions, showed limitations in capturing the nonlinear dynamics inherent to unemployment. In contrast, particularly sequential architectures (NN) demonstrated superiority in modeling complex interactions between variables. The study evaluates deep learning models with one and two hidden layers, identifying the optimal configurations of one hidden layer (16 neurons), achieving a loss of 0.2846, 75 % accuracy, and a Mean Squared Error (MSE) of 0.2064 on unseen test data. The results highlight neural networks' ability to model unemployment with greater precision than linear methods, attributing this to their adaptability in learning nonlinear patterns within multivariate data. This study demonstrates that variables such as GDP, inflation, public spending, and interest rate significantly influence unemployment. NN effectively capture these complex relationships. Their utility in predicting socioeconomic phenomena in developing countries like Ecuador is emphasized, overcoming the limitations of traditional models.

1. Introducción

El análisis de la literatura existente reveló una evolución significativa en las metodologías para modelar y predecir el desempleo en Ecuador, empleando diversas herramientas econométricas y estadísticas multivariantes. Los estudios pueden clasificarse en varias categorías fundamentales según sus enfoques metodológicos y objetivos de investigación. Así mismo, se enfatiza que el desempleo es un variable socioeconómica que esta determinada por factores exógenos y endógenos, que la definen directa e indirectamente.

Como antecedentes se tiene que un primer grupo de investigaciones se centró en la aplicación de la Ley de Okun para comprender la relación entre crecimiento económico y desempleo. En [1] investigaron la correlación entre crecimiento económico y desempleo durante 2012-2021, utilizando el coeficiente de correlación de Pearson y un modelo econométrico basado en la ecuación de Okun, encontrando una correlación lineal negativa de -0.84 entre PIB y desempleo. De manera similar, en [2] examinaron la validez de la Ley de Okun para el período 2007-2020, implementando modelos estáticos mediante Mínimos Cuadrados Ordinarios (MCO) y modelos dinámicos de cointegración, revelando que un incremento del 1 % en el PIB reduce el desempleo entre 0.07 % y 0.28 %. En [3] aplicaron la Ley de Okun para el período 1990-2010, utilizando MCO y modelos logarítmicos, y encontraron una correlación negativa de -0.266 entre PIB y desempleo. Complementariamente, [4] profundizaron en el análisis diferenciado, investigando los efectos del PIB sobre el desempleo en mujeres y jóvenes durante 2007-2019, evidenciando coeficientes de Okun significativamente diferentes por género y grupo etario. Un segundo grupo de investigaciones exploró la relación entre inflación y desempleo, principalmente mediante la Curva de Phillips. En [5] analizaron esta relación para Ecuador, Latinoamérica y el Mundo durante 1991-2015, implementando modelos de regresión lineal múltiple con variables de control. Sorprendentemente, encontraron una relación positiva contraria a la teoría, donde un incremento del 1 % en el desempleo se asociaba con un aumento del 4.39 % en la inflación. En [6] utilizaron Python y bibliotecas especializadas para modelar esta dinámica, combinando un modelo de regresión lineal múltiple basado en la Curva de Phillips con un modelo ARIMA(1,1,1), obteniendo un coeficiente de determinación R^2 de 0.380.

La aplicación de técnicas de machine learning y redes neuronales representó un avance metodológico significativo. En [7] propusieron un modelo híbrido innovador que combina redes neuronales recurrentes y algoritmos genéticos para predecir la tasa de desempleo. Su metodología incluyó preprocesamiento de datos mediante normalización y detección de valores atípicos, implementando modelos BiLSTM, GRU y un modelo híbrido GA-LSTM. Los resultados demostraron la superioridad del modelo híbrido, con un error cuadrático medio (MSE) de 0.052 y un error porcentual absoluto

medio (MAPE) de 3.797%. [7] y amplió este enfoque aplicando diversos modelos de machine learning, incluyendo árboles de decisión, bosques aleatorios, modelos logit y probit, máquinas de soporte vectorial, redes neuronales y análisis discriminante. El análisis discriminante resultó ser el modelo más efectivo, alcanzando una precisión del 95.90% en la clasificación de individuos empleados y desempleados.

Un tercer grupo de investigaciones se enfocó en aspectos multivariantes más amplios del desempleo. En [8] implementaron técnicas estadísticas multivariantes que incluyeron regresión lineal múltiple y modelos Probit para analizar los factores determinantes del desempleo. Su modelo Probit alcanzó un 99% de predicciones correctas, revelando la incidencia significativa de variables como área de residencia, edad, estado civil, etnia, ocupación e ingresos en la probabilidad de desempleo. En [9] investigaron la incidencia del desempleo y la pobreza en las remesas de migrantes, utilizando modelos de regresión lineal simple y técnicas de validación estadística. Un aspecto metodológico crucial fue la implementación de rigurosas técnicas de validación estadística. La mayoría de los estudios emplearon pruebas de estacionariedad como Dickey-Fuller Aumentada, pruebas de heterocedasticidad (White, Breusch-Pagan), análisis de autocorrelación (Breusch-Godfrey), pruebas de normalidad (Shapiro-Wilk) y evaluación de multicolinealidad mediante el Factor de Inflación de Varianza (VIF). En [10] ejemplificaron esta aproximación comprehensiva, implementando un modelo de regresión lineal simple con MCO y múltiples técnicas de validación estadística, revelando una capacidad predictiva del 43.30%.

Finalmente, estudios como [11] y [12] exploraron la influencia de variables exógenas, como precios del petróleo y shocks macroeconómicos, en la dinámica del desempleo. [11] utilizó modelos de rezagos distribuidos bajo Mínimos Cuadrados Ordinarios (MCO), demostrando que variaciones en el precio del petróleo y la cotización del dólar explicaban el 45.6% de la variación en la tasa de desempleo urbano. La evolución metodológica evidenciada en estos estudios refleja una transición gradual desde modelos econométricos tradicionales hacia enfoques más sofisticados que incorporan técnicas de machine learning, análisis multivariante avanzado y validación estadística rigurosa, proporcionando una comprensión cada vez más precisa de los factores que determinan el desempleo en Ecuador. A contribuir a la literatura existente, este estudio no solo busca mejorar la precisión en la predicción del desempleo, sino también proporcionar una herramienta analítica más robusta para la formulación de políticas económicas. En última instancia, esperamos que nuestros hallazgos en este trabajo ofrezcan nuevas perspectivas sobre cómo las dinámicas del desempleo interactúan con los factores macroeconómicos en Ecuador, ayudando a construir una base sólida para futuras investigaciones estadísticas, econométricas, y decisiones estratégicas en este campo.

2. Materiales y Métodos

Para realizar el artículo, han sido empleadas diversas bases de datos las cuales son bases en series de tiempo en [13], [14], [15] y [16], y se encuentran como registros de valores en índices porcentuales sobre 100%. El presente estudio es de tipo aplicado con un enfoque cuantitativo, con el propósito de analizar la relación entre variables económicas y sociales relacionadas con el desempleo mediante el uso de modelos econométricos y técnicas de Machine Learning. Este enfoque permite describir y prever patrones históricos del desempleo en Ecuador. Por ello, se clasifica como descriptivo y predictivo, pues además de identificar tendencias pasadas, se enfoca en desarrollar herramientas para la predicción de comportamientos futuros.

El diseño es no experimental dado que no se manipulan las variables, sino que se analizan datos observacionales obtenidos de fuentes secundarias. Es un diseño longitudinal, ya que se utilizan datos de series temporales que abarcan desde el año 1978 hasta 2023, y comparativo, debido a que se contrastan modelos econométricos tradicionales con técnicas modernas de Machine Learning. La metodología empleada en esta investigación es de tipo descriptiva, de acuerdo con [17], quienes señalan que la metodología descriptiva se enfoca en describir fenómenos sin intervención directa del investigador. En este caso, han sido recopilados, identificados y analizados antecedentes generales del desempleo en Ecuador, así como datos y estadísticas correspondientes al periodo antes mencionado. Se aplican métodos de investigación inductivo y deductivo, lo que permite interpretar los datos obtenidos de manera racional y objetiva, y analizar el impacto de las cifras de desempleo en la población ecuatoriana.

El fundamento teórico de esta investigación se sustenta en la investigación bibliográfica, consistente con lo señalado por [17], quienes destacan que la recolección de información bibliográfica es una de las etapas más relevantes de cualquier investigación, garantizando la pertinencia y calidad de los datos utilizados. Para ello, se consultaron diversas bases de datos científicos, tales como: EBSCO, Redalyc, Scielo y Research Gate. Además, se recurrió a informes y estadísticas proporcionadas por organizaciones e instituciones nacionales e internacionales, tales como: Instituto Ecuatoriano de Estadísticas y Censos (INEC), Banco Central del Ecuador (BCE), Organización Internacional del Trabajo (OIT), Datos macroeconómicos del Ecuador y de América Latina y el Caribe (CEPAL). Estas fuentes aseguraron la validez y confiabilidad de los datos empleados en el análisis del desempleo en Ecuador, así como en el desarrollo de modelos



Figura 1: Indicadores nacionales en % respecto a la PEA

econométricos tradicionales y modernos.

Población. Incluye todos los datos macroeconómicos y sociales relevantes del Ecuador entre los años 1978 y 2023. Los indicadores clave, de [9], [13], [14] y [15] incluyen la variable dependiente es el Desempleo (y) que es el Porcentaje de la población activa desempleada, y variables independientes son PIB (x_1) que es el Producto Interno Bruto en términos reales, la Inflación (x_2) que es el Medida como el cambio porcentual en los precios al consumidor, las Exportaciones e Importaciones (x_3) y (x_4) que son valorizadas en millones de dólares estadounidenses, se mide en porcentaje sobre 100 %, el Salario Mínimo (x_5) que esta En dólares, ajustado por inflación cuando sea necesario, el Gasto Público (x_6) que es Expresado como porcentaje del PIB, y la Tasa de interés (x_7) que es el valor de impuesto agregado a un préstamo.

Muestra. El estudio abarcó el periodo 1978-2023, recopilando observaciones anuales que totalizan 45 datos por variable. Los criterios de selección incluyeron la disponibilidad completa y consistente de datos macroeconómicos y la representatividad de variables relacionadas directamente con el desempleo. El área de intervención comprendió el análisis de datos macroeconómicos históricos obtenidos de fuentes confiables, con un levantamiento de datos basado en registros oficiales disponibles en línea. La muestra ha sido seleccionada mediante un enfoque de muestreo no probabilístico, garantizando la inclusión de datos completos y consistentes para todas las variables consideradas.

La variable principal de estudio es la tasa de desempleo, la cual se calcula utilizando la fórmula que relaciona el número de personas desempleadas con la población económicamente activa (PEA), expresada como porcentaje. Esta fórmula se define como:

$$Tasa\ de\ Desempleo\ (\%) = \left(\frac{Número\ de\ personas\ desempleadas}{Población\ total\ activa} \right) \times 100. \quad (1)$$

En este contexto, en base a 1, las personas desempleadas incluyen a aquellos que no tienen empleo, pero están disponibles para trabajar y han realizado acciones para buscar empleo en un período específico. Por ejemplo, se puede apreciar una muestra ligera de la tasa de desempleo entre el año del 2019 al 2024, ver Figura 1.

La Población Económicamente Activa (PEA) comprende a todas las personas empleadas y desempleadas que conforman la fuerza laboral activa. Esta métrica permite expresar el nivel de desempleo como un porcentaje del total de la fuerza laboral, lo cual facilita su interpretación y el análisis de su impacto en la economía nacional. La tasa de desempleo, en este contexto, se modela en función de variables independientes previamente definidas, con el objetivo de realizar pronósticos y generar nuevas contribuciones al análisis estadístico del desempleo en el Ecuador. El estudio se basa en 46 observaciones correspondientes a los datos disponibles hasta la fecha, constituyendo la población total utilizada para el modelado y pronóstico. Asimismo, se aplicaron criterios de exclusión a través de instrumentos metodológicos que permitieron eliminar variables con registros incompletos, inconsistencias o series temporales interrumpidas. Para el análisis, se emplearon herramientas estadísticas y de programación: R fue utilizado para el preprocesamiento de datos, análisis econométrico y pruebas estadísticas; mientras que Python se aplicó en la implementación y evaluación de modelos de aprendizaje automático (Machine Learning). Entre las técnicas estadísticas aplicadas se incluyen el Análisis Exploratorio de Datos (EDA), pruebas de multicolinealidad mediante el Índice de Inflación de la Varianza (VIF), y la prueba de Breusch-Pagan para la detección de heterocedasticidad, con el fin de asegurar la solidez y consistencia de los modelos propuestos.

Se emplean herramientas ya expuestas en la literatura como modelos econométricos, redes neuronales, series temporales y regresión lineal. Lo cual, se define posteriormente: Con los Modelos Econométricos que Se utilizan para analizar la relación entre variables económicas y prever el comportamiento de indicadores como el desempleo, basándose en métodos estadísticos clásicos y avanzados, como en el contexto de [1]. Las Redes Neuronales que Son modelos computacionales que imitan el cerebro humano, capaces de aprender patrones complejos en grandes volúmenes de datos, como la predicción de series temporales, esto se puede ver en [7]. Las Series Temporales que son una Técnica estadística utilizada para analizar datos recogidos en intervalos regulares de tiempo, permitiendo prever el comportamiento futuro de variables como el desempleo, [18]. Y la Regresión Lineal que es un Método estadístico utilizado para modelar la relación lineal entre una variable dependiente y una o más variables independientes, comúnmente aplicado en economía para análisis predictivos, [18].

2.1. Tratamiento de datos

Distancias de Mahalanobis La construcción de bases de datos en cualquier estudio puede incluir registros incompletos o valores atípicos, lo que hace imprescindible aplicar un protocolo de análisis de datos antes de realizar cualquier procedimiento estadístico. Entre las técnicas más utilizadas para analizar muestras multivariadas se encuentra la distancia de Mahalanobis, que mide cuántas desviaciones estándar separa una observación \mathbf{X} de la media (μ) en una distribución. Esta métrica es especialmente útil para identificar datos atípicos, ya que estos suelen diferir significativamente de los patrones comunes observados en los datos [19].

Desde una perspectiva geométrica, la distancia Euclidiana define la mínima separación entre dos puntos, pero no tiene en cuenta la correlación entre variables. Por el contrario, la distancia de Mahalanobis incorpora estas correlaciones, convirtiéndola en una herramienta más adecuada para muestras con variables relacionadas [19].

Esta métrica invariante a cambios de escala mide la distancia entre un punto $x \in \mathbb{R}^p$, generado por una distribución $f_{\mathbf{X}(\cdot)}$, y la media de dicha distribución ($\mu = E(X)$), utilizando la matriz de covarianza $\Sigma = E(X - \mu)$. Se define como:

$$D(X, \mu) = \sqrt{(X - \mu)^T \Sigma^{-1} (X - \mu)}. \quad (2)$$

Análisis de Componentes Principales (PCA) El PCA es una técnica utilizada para reducir la dimensionalidad de un conjunto de datos, transformando variables correlacionadas en componentes principales no correlacionadas. Es empleado para explorar relaciones entre variables y simplificar datos, además de ser útil para determinar el número de capas ocultas en redes neuronales [20].

Dado un conjunto de datos *dataset* $x^{(1)}, x^{(2)}, \dots, x^{(m)}$ con observaciones n -dimensionales, se pretende reducir el *dataset* a observaciones k -dimensionales (cuando $k < n$). Para esto, el proceso inicia con la estandarización de los datos:

$$x_j^i = \frac{x_j^i - \bar{x}_j}{\sigma_j}, \quad (3)$$

luego, se calcula la matriz de covarianza mediante:

$$\Sigma = \frac{1}{r} \sum_i (x_i)(x_i)^T, \quad \Sigma \in \mathbb{R}^{n \times n} \quad (4)$$

A continuación, se obtiene el *eigenvector* y *eigenvalue* de la matriz de covarianza mediante la ecuación:

$$U^T \Sigma = \lambda \mu, U = \begin{bmatrix} | & | & \cdots & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & \cdots & | \end{bmatrix}, \quad u_i \in \mathbb{R}^n, \quad (5)$$

de esta manera, los datos originales son proyectados a un subespacio de k dimensiones, de manera que los principales *eigenvectors* de la matriz de covarianza son seleccionados. Estas nuevas variables son capaces de representar a los datos originales y su varianza. Cada uno de estos nuevos vectores se pueden obtener mediante la expresión:

$$x_i^{\text{new}} = \begin{bmatrix} u_1^T x_i \\ u_2^T x_i \\ \vdots \\ u_k^T x_i \end{bmatrix} \in \mathbb{R}^k \quad (6)$$

Lo previo se utiliza en específico en conjunto con las técnicas de modelaje que se observan en lo posterior. Esto es en:

Modelos Econométricos Tradicionales

En este caso se aplican los Modelos de Mínimos Cuadrados Ordinarios (OLS), como en el contexto de [3].

Mínimos Cuadrados Ordinarios (OLS)

La fórmula de OLS es:

$$Y = B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n + \varepsilon. \quad (7)$$

En la cual: Y es la variable dependiente (en este caso, la tasa de desempleo); B_0 es el término constante o intercepto; B_1, B_2, \dots, B_n son los coeficientes de las variables independientes; X_1, X_2, \dots, X_n que son las variables independientes (PIB, inflación, salario mínimo, etc.); y ε es el término de error (variación no explicada por el modelo).

Se empleó modelos de *Machine Learning*, en los cuales se utilizó un modelo secuencial de red neuronal recurrente, para capturar patrones temporales complejos y no lineales en los datos, como se encuentra en [7] y [21].

Redes neuronales: Las redes neuronales son un enfoque de clasificación que imita el funcionamiento de las neuronas biológicas. Cada neurona artificial procesa señales de entrada ponderadas por un conjunto de pesos, activándose únicamente si la suma alcanza un umbral determinado [22]. Existen diversas funciones de activación que definen el comportamiento de las neuronas; en este caso, se emplearon **ReLU** = $\max(0, y)$, en las capas ocultas y en la salida la función lineal x , con x en cierto intervalo continuo.

En las redes neuronales, se crean enlaces entre las neuronas, a cada uno de los cuales se le asigna un peso $V_{(j,k)}$, el cual es ajustado durante el proceso de aprendizaje. Estos pesos permiten que las combinaciones de entradas activen o no las neuronas de las capas ocultas y de salida, de modo que el modelo aprenda comportamientos no lineales a partir de los datos. En este proceso, la propagación de señales en cada capa se describe por las ecuaciones $X_j = R_{ij} \cdot I$ y $O_j = \text{activation}(X_j)$, donde X_j es la matriz de señales de entrada a la capa j , R_{ij} son los pesos entre las capas, I es la matriz de señales de entrada y O_j es la matriz de salida de la capa.

Para que la red neuronal aprenda de manera efectiva, se calcula el error $e_{(out,k)} = t_k - o_k$ en la capa de salida, comparando los valores obtenidos con los esperados. Este error debe retroalimentarse a través de la red para ajustar los pesos. La retro propagación de los errores se realiza mediante la fórmula $C_i = V_{ij}^T \cdot C_j$, donde C_i es el error que se retropropaga a la capa anterior y C_j es el error proveniente de la capa siguiente. Al retropropagar los errores, los pesos se ajustan para que la red neuronal pueda retener información de ejemplos previos y actualizarse con nuevas observaciones.

Uno de los procesos más ampliamente empleados para este propósito es el descenso de gradiente, que es formulado en [19], como se muestra en las ecuaciones (8-10).

$$\frac{\partial C}{\partial V_{jk}} = \frac{\partial \sum_n (t_n - o_n)}{\partial V_{jk}} = \frac{\partial C}{\partial O_k} \cdot \frac{\partial O_k}{\partial V_{jk}} = -2(t_n - O_n) \cdot \frac{\partial O_k}{\partial V_{jk}} \quad (8)$$

$$\frac{\partial C}{\partial V_{jk}} = -2(t_n - o_n) \cdot \frac{\partial}{\partial V_{jk}} \left[\text{activation} \left(\sum_j V_{jk} \cdot O_j \right) \right] \quad (9)$$

$$V_{jk}^{(r+1)} = V_{jk}^{(r)} - \alpha \cdot \frac{\partial C}{\partial V_{jk}} \quad (10)$$

Donde $V_{jk}^{(r+1)}$ representa el nuevo peso actualizado para un enlace jk , que es actualizado a partir de su valor anterior $V_{jk}^{(r)}$ en el gradiente $\frac{\partial C}{\partial V_{jk}}$ que aporta una porción de nueva energía transmitida mediante el hiper-parámetro *learning rate* α [22].

Deep Learning. Las redes neuronales artificiales que presentan 2 o más capas ocultas con funciones de activación no lineales consecutivas se denominan modelos de Deep Learning como se ve en [23]. La adición desmesurada de capas ocultas y un mayor número de neuronas, no siempre es la mejor alternativa ya que esto puede llevar al modelo a presentar problemas de sobreajuste. Además, el cálculo de los parámetros implicados en el modelo puede representar una tarea desafiante ya que el cálculo de la actualización de parámetros involucrará un mayor número de derivadas. Este problema puede ser abordado mediante la regla de la cadena que se establece como:

$$\frac{df_3}{du}(x) = \frac{df_3}{du} [f_2(f_1(x))] \times \frac{df_2}{du} [f_1(x)] \times \frac{df_1}{du}(x), \quad (11)$$

Por ejemplo, en [21] para un modelo de Deep Learning de dos capas ocultas, a más de la matriz de pesos \mathbf{V}_k involucrada en cada capa, se puede añadir un término Bias a manera de intercepto \mathbf{B}_k . El concepto de un modelo de dos capas ocultas se presenta en la Figura 2, en la cual \mathbf{V} es denotada por \mathbf{W} .

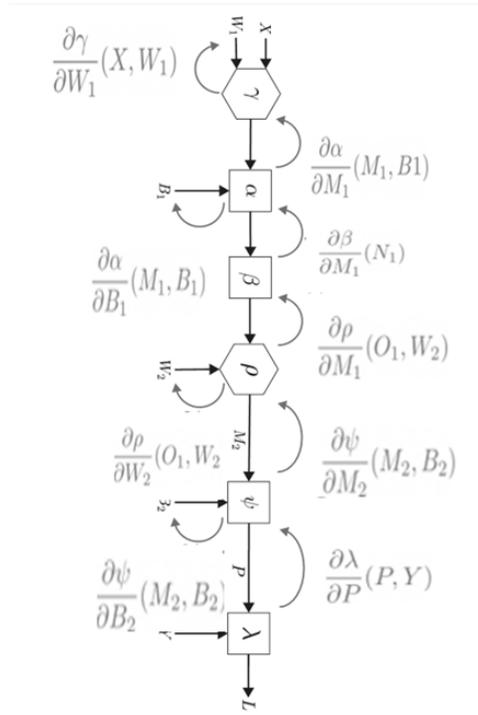


Figura 2: Formulación de un modelo de Deep Learning con dos capas ocultas

Siguiendo la formulación propuesta los gradientes, empleados para la actualización de los pesos en las conexiones de la red neuronal, se pueden calcular mediante las expresiones, como en [24]:

$$\frac{\partial L}{\partial B_2} = \frac{\partial \lambda}{\partial P}(P, Y) \times \frac{\partial \Psi}{\partial B_2}(M_2, B_2), \tag{12}$$

$$\frac{\partial L}{\partial W_2} = \frac{\partial \lambda}{\partial P}(P, Y) \times \frac{\partial \psi}{\partial M_2}(M_2, B_2) \times \frac{\partial \rho}{\partial M_2}(O_1, W_2), \tag{13}$$

$$\frac{\partial L}{\partial B_1} = \frac{\partial \lambda}{\partial P}(P, Y) \times \frac{\partial \psi}{\partial M_2}(M_2, B_2) \times \frac{\partial \rho}{\partial M_1}(O_1, W_2) \times \frac{\partial \beta}{\partial M_1}(N_1) \times \frac{\partial \alpha}{\partial B_1}(M_1, B_1), \tag{14}$$

$$\frac{\partial L}{\partial W_1} = \frac{\partial \lambda}{\partial P}(P, Y) \times \frac{\partial \psi}{\partial M_2}(M_2, B_2) \times \frac{\partial \rho}{\partial M_1}(O_1, W_2) \times \frac{\partial \beta}{\partial M_1}(N_1) \times \frac{\partial \alpha}{\partial M_1}(M_1, B_1) \times \frac{\partial \gamma}{\partial W_1}(X, W_1). \tag{15}$$

Estos son un tipo de red neuronal diseñado para trabajar con datos secuenciales, como series temporales, texto o cualquier tipo de información ordenada en el tiempo. Su principal característica es que tienen conexiones con capas de la red apiladas una sobre otra, lo que les permite recordar información previa y capturar dependencias temporales en los datos.

En este trabajo se emplean los modelos secuenciales de redes neuronales, los cuales son diseñados para manejar secuencias de datos con dependencias a largo plazo. En este tipo de redes, cada célula de la NN contiene puertas de entrada, salida y olvido, las cuales controlan el flujo de información a través de la red. Estas puertas permiten decidir qué información conservar, actualizar o descartar, asegurando que la red pueda mantener información relevante durante varias iteraciones [7, 11, 22].

Así mismo, por otro lado, como las Redes Neuronales Recurrentes (RNN) tradicionales, las redes con memoria de corto y largo plazo (LSTM) son capaces de aprender relaciones más complejas y mitigar problemas como el desvanecimiento de gradientes, lo que las hace ideales para tareas de predicción en series temporales y modelado de datos secuenciales.

La estructura básica de una RNN con LSTM, según [24, 21], consiste en varias capas interconectadas. La primera capa es la capa de entrada, que recibe los datos de entrada secuenciales X como una secuencia de características, donde cada observación tiene n características. Esta capa transforma la entrada en una secuencia que será procesada por la capa oculta LSTM. En este tipo de redes, la propagación de la señal se realiza mediante la combinación de pesos, activaciones y funciones de activación, las cuales ayudan a determinar qué información se mantiene o se descarta a medida que avanza la secuencia temporal. A lo largo de este proceso, cada neurona oculta ajusta sus pesos en función del aprendizaje para activar las siguientes neuronas o capas.

En la capa oculta LSTM, se emplean puertas que controlan la información que pasa a lo largo de la red. La puerta de olvido f_t decide qué información del estado anterior se descarta, calculándose mediante la fórmula: $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$, donde σ es la función sigmoide, W_f son los pesos, h_{t-1} es el estado oculto anterior y x_t es la entrada actual.

La puerta de entrada i_t determina qué nueva información debe almacenarse en el estado de memoria, usando la fórmula: $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$. El estado de memoria actualizado C_t se calcula como: $C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$, donde \odot denota el producto de qHadamard (elemento a elemento).

El estado oculto h_t , una representación comprimida de la secuencia temporal que pasa a la siguiente capa de la red, se calcula como: $h_t = o_t \odot \tanh(C_t)$, donde o_t es la puerta de salida. La capa de salida utiliza el estado oculto h_t para generar la predicción final. Si la tarea es de regresión, la salida es directamente h_t .

Además, una red neuronal realiza los cálculos considerando las entradas x_1, x_2, \dots , propagando los valores desde las neuronas de entrada a las de salida Y , y usando los pesos w como parámetros intermedios, con b siendo el sesgo. El aprendizaje se basa en el ajuste de los pesos que conectan las neuronas. La base de datos proporciona la información sobre la exactitud de los pesos al comparar qué tan bien coincide la salida pronosticada para una entrada con la etiqueta anotada en los datos de entrenamiento [22].

Por ejemplo, en una neurona artificial, la salida se calcula como: $Y = x_1 \cdot w_1 + x_2 \cdot w_2 + b_y$, donde w_i son los pesos, x_i las entradas y b_y el sesgo. La red ajusta sus pesos mediante una función de pérdida. Para tareas de regresión, esta función es el error cuadrático medio (MSE), definido como:

$$\text{MSE} = \frac{1}{r} \sum_{i=1}^r (Y_i - \hat{Y}_i)^2,$$

y luego optimiza estos pesos usando algoritmos como Adam, con el fin de minimizar el error y mejorar el rendimiento del modelo. Posterior se enfatizan las clases de métricas para medir los errores.

2.2. Análisis Estadístico con Métricas de Evaluación

Las métricas de precisión predictiva, según [7], incluyen diversas medidas para evaluar el rendimiento de los modelos predictivos:

El Error Cuadrático Medio (MSE):

$$\text{MSE} = \frac{1}{n} \sum (Y_i - \hat{Y}_i)^2$$

evalúa la desviación promedio cuadrática entre los valores reales y predichos, mientras que el Error Porcentual Absoluto Medio (MAPE):

$$\text{MAPE} = \frac{100}{n} \sum \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|$$

mide el error como porcentaje relativo sobre los valores reales. El Error Medio Absoluto (MAE):

$$\text{MAE} = \frac{1}{n} \sum |Y_i - \hat{Y}_i|$$

representa la magnitud promedio del error absoluto, y La Raíz del Error Cuadrático Medio (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum (Y_i - \hat{Y}_i)^2}$$

ofrece una medida del error en las mismas unidades que los datos originales. Adicionalmente, se utilizan métricas establecidas en la literatura como El Criterio de Información de Akaike (AIC):

$$\text{AIC} = 2k - 2 \ln(L)$$

, que penaliza modelos complejos que no mejoran significativamente el ajuste, El R^2 ajustado:

$$R_{\text{adj}}^2 = 1 - \left(\frac{1 - R^2}{n - 1} \cdot \frac{n - 1}{n - k - 1} \right)$$

que ajusta el R^2 estándar para reflejar el número de predictores en el modelo, y El Factor de Inflación de la Varianza (VIF):

$$\text{VIF} = \frac{1}{1 - R_j^2}$$

que evalúa la multicolinealidad entre las variables independientes. Los intervalos de confianza (IC) estiman el rango donde podría estar la media poblacional, mientras que los intervalos de predicción (IP) estiman el rango donde caerá un nuevo valor individual observado.

Estas métricas permiten comparar el desempeño entre enfoques tradicionales y modernos.

3. Resultados

Se lograron los siguientes resultados para el análisis del desempleo en el Ecuador. La base de datos consta de 46 observaciones obtenidas del instrumento aplicado para modelar el desempleo en función de 7 variables predictoras. El instrumento es multivariado continuo, por lo cual las técnicas estadísticas convencionales fueron suficientes para abordar el problema planteado. Sin embargo, también fue necesario el uso de redes neuronales (NN) y técnicas de *Deep Learning*, herramientas ampliamente utilizadas en los últimos años para resolver problemas complejos [18].

La base de datos fue procesada con el lenguaje de programación estadística R y también con Python, a través de su distribución Anaconda, para manejar los paquetes `TensorFlow` y `Keras` desde el entorno `RStudio`, utilizando la librería `reticulate`. El análisis comenzó con la imputación de datos mediante la técnica *K-Nearest Neighbors* (KNN), empleando las librerías `VIM` y `skimr`. Los resultados muestran que no existen datos faltantes (véase Figura 5).

Posteriormente, se aplicó un tratamiento a los datos para detectar observaciones atípicas mediante el uso de las distancias de Mahalanobis. Se utilizó un punto de corte basado en la distribución χ^2 con un nivel de confianza del 99.9%, excluyendo así el 0.1% de las observaciones consideradas como atípicas. Los resultados mostraron que no se excluyó ninguna observación, por lo tanto, la muestra final permanece con 46 registros.

Antes de analizar el instrumento, fue necesaria su validación para garantizar la obtención de inferencias y resultados confiables. Como técnica paramétrica, se verificaron cinco supuestos fundamentales: normalidad, independencia, linealidad, homocedasticidad y multicolinealidad. Además, se construyó una matriz de correlación multivariada con las siete variables predictoras para observar las relaciones entre ellas (véase Figura 4).

3.1. Análisis De Los Datos.

Se realizó la visualización de los datos mediante diagramas de caja (boxplots) para cada una de las variables explicativas, lo que permitió identificar y remover posibles datos atípicos del conjunto. También se evaluaron supuestos e hipótesis estadísticas relevantes:

1. **Normalidad:** para este apartado se aplica la prueba de Lilliefors (Kolmogorov-Smirnov for normality test), en la cual se observa que H_0 : Los datos sí cumplen la cualidad de normalidad, con un $p < 0,05$, y H_1 : Los datos no muestran normalidad, con $p > 0,05$. En este caso, el p -value es de 0.0004214, por tanto no se rechaza la H_0 .
2. **Independencia:** se utiliza la Independencia de los predictores Durbin-Watson, con H_0 : los residuos no presentan problemas de autocorrelación y H_1 : los residuos presentan problemas de autocorrelación. En este caso, se observa que p -value < 0 . Por tanto se rechaza la H_0 , en este caso, las variables explicativas no son independientes entre sí.
3. **Linealidad:** del data set, se emplea la Linealidad Q-Q plot, prueba de Box-Tidwel, con H_0 : Los datos tienen linealidad, y H_1 : Los datos no tienen linealidad. En este caso el p -value es menor que 0.05, por esto se rechaza la H_0 , ver Figura 3.

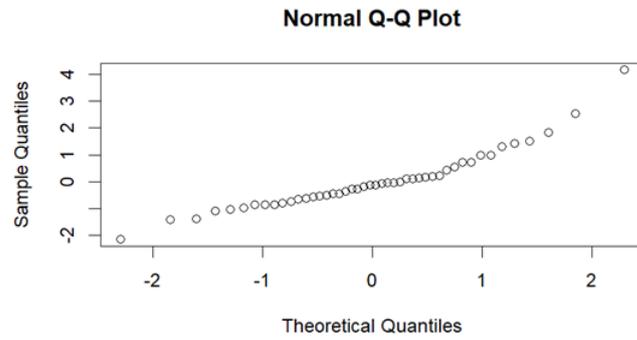


Figura 3: Linealidad

4. **Homocedasticidad:** Se aplica Homoscedasticidad de Breusch Pagan, (Non Constant Variance Score NCVS), en este caso la H_0 : los datos presentan homocedasticidad, y H_1 : los datos no presentan homocedasticidad, en este caso $p\text{-value} = 0.042 < 0.05$. Por tanto, no se rechaza la H_0 , es decir, los residuos muestran varianza constante, como en la Figura 4.

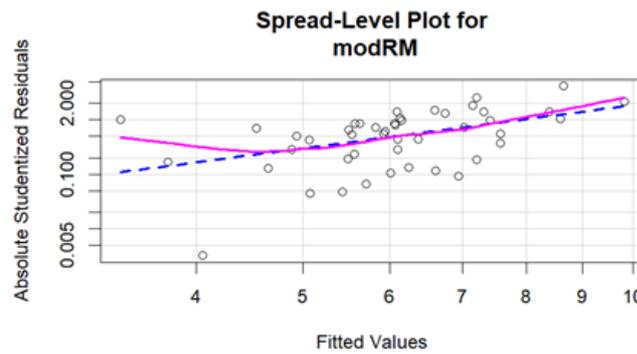


Figura 4: Homoscedasticidad de los datos

5. Se emplea la Multicolinealidad VIF (Variance inflation factor), con un valor menor que 10 para cada variable, lo que significa que cada variable presenta una varianza moderada, lo cual sirve para dar una buena calidad de ajuste del modelo, ver Figura 5.

Además, en la figura 6 podemos apreciar los histogramas de cada una de las variables.

3.1.1. Regresión Lineal Múltiple.

Con la Regresión Lineal se modela el desempleo en función de las variables predictoras. Se utilizó *Stepwise* (ir eliminando las variables que no son significantes), para descubrir el modelo (la combinación de variables apropiada). Se empezó con el modelo:

$$\text{Desempleo} \sim \text{Exportaciones} + \text{Pib} + \text{Importaciones} + \text{GastoPublico} + \text{SalarioMinimo} + \text{TasaInteres} + \text{Inflacion}. \quad (16)$$

A medida que se utilizó el *Stepwise*, para remover variables que no eran altamente significativas, es decir superaban el $p\text{-value}$ del 5%. Obteniendo, así los modelos posteriores:

$$\text{Desempleo} \sim \text{Exportaciones} + \text{Pib} + \text{Importaciones} + \text{TasaInteres} + \text{GastoPublico} + \text{SalarioMinimo} + \text{Inflacion}, \quad (17)$$

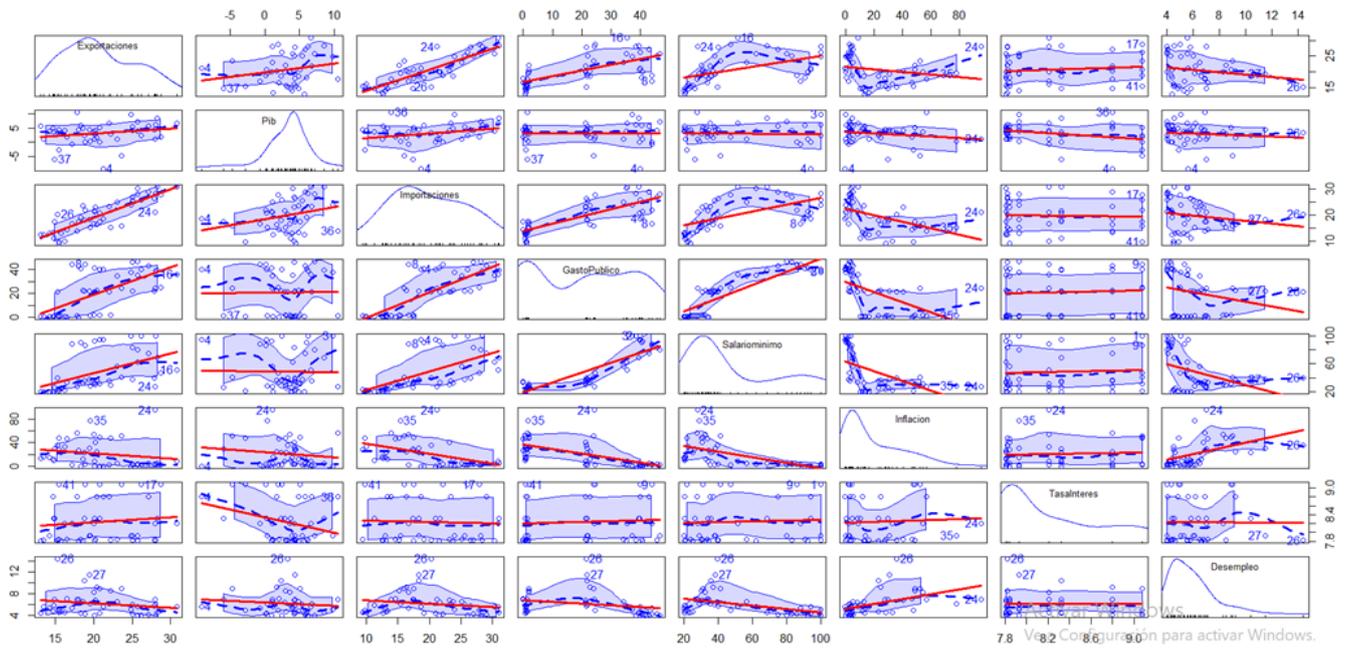


Figura 5: Correlación de las variables

Group variables	None									
— Variable type: numeric		complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
1 Exportaciones	0	1	20.4	4.85	13.1	16.3	20.1	23.6	30.8	
2 Pib	0	1	3.03	3.85	-9.2	1.2	3.7	5.1	10.5	
3 Importaciones	0	1	19.5	6.14	9.38	14.6	18.7	23.8	30.9	
4 GastoPublico	0	1	20.1	16.2	0	1.4	22.0	35.6	44.1	
5 SalarioMinimo	0	1	46.6	26.6	20	27.5	35.8	62.1	100	
6 Inflacion	0	1	20.7	21.8	-0.34	3.46	12.6	32.8	95.5	
7 TasaInteres	0	1	8.25	0.459	7.81	7.84	8.21	8.8	9.08	
8 Desempleo	0	1	6.33	2.21	4.05	4.64	5.62	7	14.4	

Figura 6: Estadísticos de las variables de estudio.

en el cual, la variable tasa de interés, no ha sido altamente significativa, con un *p-value* de 0.92881 y ha sido excluido, posterior está el siguiente modelo:

$$Desempleo \sim Exportaciones + Pib + Importaciones + GastoPublico + SalarioMinimo + Inflacion, \tag{18}$$

en el cual, la variable PIB, no ha sido altamente significativa, con un *p-value* de 0.67 y ha sido excluido, posterior está el modelo:

$$Desempleo \sim Exportaciones + Importaciones + GastoPublico + SalarioMinimo + Inflacion, \tag{19}$$

en el cual, la variable Importaciones, no ha sido altamente significativa, con un *p-value* de 0.2468 y ha sido excluido, posterior está el modelo:

$$Desempleo \sim Exportaciones + GastoPublico + SalarioMinimo + Inflacion, \tag{20}$$

en el cual, la variable Exportaciones, no ha sido altamente significativa, con un *p-value* de 0.18737, en el cual el modelo presenta un AIC=194.17 y ha sido excluido, posterior está el modelo:

$$Desempleo \sim GastoPublico + SalarioMinimo + Inflacion. \tag{21}$$

Y en función al *p-value* <0.05 para cada variable, por último está el modelo que se ajusta medianamente con un $R^2 = 29,73\%$ de calidad, el cual es:

$$Desempleo \sim GastoPublico + SalarioMinimo + Inflacion. \tag{22}$$

Con la regresión lineal se observa que la tasa de desempleo se ajusta de una manera eficaz con las variables explicativas Gasto Público, Salario Mínimo e Inflación, las cuales muestran un p-value altamente significativo, ver tabla 1.

Tabla 1: Estadísticos descriptivos del mejor modelo de regresión lineal múltiple

Componente	Detalles				
Fórmula del modelo	$Desempleo \sim GastoPúblico + SalarioMínimo + Inflación$				
Residuales	Mín: -3.31	10 %: -1.043	Mediana: -0.1946	30 %: 0.6305	Máx: 7.157
Coefficientes					
- (Intercepto)	Estimación: 6.33614 / Error estándar: 0.87516 / t: 7.240 / p-valor: <0.001 (***)				
- <i>GastoPúblico</i> x_6	Estimación: 0.06453 / Error estándar: 0.03511 / t: 1.838 / p-valor: 0.0731 (.)				
- <i>SalarioMínimo</i> x_5	Estimación: -0.04742 / Error estándar: 0.02187 / t: -2.169 / p-valor: 0.0358 (*)				
- <i>Inflación</i> x_2	Estimación: 0.03900 / Error estándar: 0.01517 / t: 2.572 / p-valor: 0.0137 (*)				
Estadísticos globales					
- Error estándar residual	1.874 (42 <i>gl</i>)				
- R^2	0.2973				
- R^2 ajustado	0.2471				
- Estadístico F	5.922 (3 y 42 <i>gl</i>) / p-valor: 0.001829				
- AIC	194.1479				
Notas	***: $p \leq 0.001$; **: $p \leq 0.01$; *: $p \leq 0.05$; .: $p \leq 0.1$ <i>gl</i> = grados de libertad				

Por tanto, el modelo en base a la Tabla 1, para el modelo (22), sería:

$$y = 6,33614 + 0,039 \cdot x_2 - 0,04 \cdot x_5 + 0,06 \cdot x_6 + error. \quad (11)$$

Además, el modelo se ajusta a través de regresión lineal con un R^2 cerca al 29.73 %, lo cual muestra que la calidad de ajuste es baja, y las variables explicativas no se ajustan a los datos de una manera eficaz. También, se observa que el modelo presenta un AIC=194.14, que es medianamente bueno y menor con respecto al AIC del modelo 20, el cual muestra un valor de 194.17.

3.1.2. Modelos secuenciales de Redes Neuronales (NN).

Se emplearon modelos secuenciales de Redes Neuronales con una capa de entrada de 8 neuronas, 2 capas ocultas (layer dense) y una de salida. En la capa de entrada hay 8 neuronas, en la primera capa oculta entre 8 a 20 neuronas, y en la capa oculta 2, entre 9 a 38 neuronas, esto se hizo con validación cruzada, es decir, con la idea de hiperparámetros sujetos a la idea de función de pérdida y el error cuadrático medio. Además, las capas ocultas ocupan normalización. El modelo se entrenó con 50 epochs de aprendizaje mediante la optimización por el modelo Adam el cual utiliza una tasa de aprendizaje (rate decay) igual a 0.1, el cual decrece con una razón de 0.1/50 para cada etapa de aprendizaje nueva. El valor de 50 epochs de aprendizaje viene por la rule of thumb de [16] y [25], que permite el triple del número de variables para las etapas de aprendizaje. Los resultados de esta etapa de aprendizaje se muestran en la figura 6 y su arquitectura para el clasificador en la figura 7.

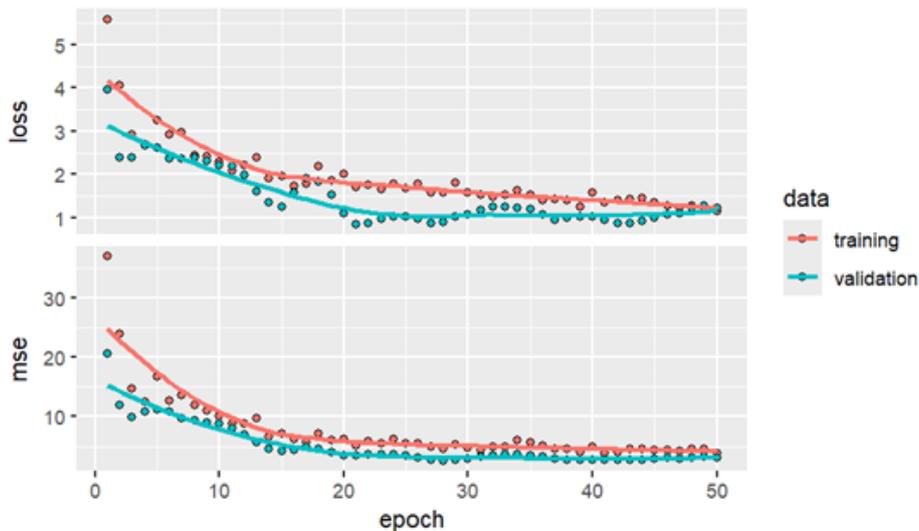


Figura 7: Proceso de entrenamiento de la red recurrente con una capa oculta

El modelo de red neuronal recurrente diseñado con 8 neuronas de entrada, con una capa oculta, fue evaluado en el conjunto de prueba que corresponde al 20% de las observaciones totales (15 observaciones que la red no había visto previamente), ver figura 8. Al evaluar el desempeño de la red con estas nuevas observaciones, se obtuvo un Loss de 0.2846069, y un MSE de 0.2063614, con un número exacto de 16 neuronas de la primera capa oculta. Este valor se tomó puesto es el menor en comparación a los otros valores entre las 8 hasta las 20 neuronas con respecto al Loss y al MSE, ver la Tabla 2. Además, se empleó una segunda capa oculta, entre 9 a 38 neuronas, ver Tabla 3, en la cual se puede apreciar que el MSE es 0.4178960 y el Loss es 0.5044433 para cada una de las neuronas, respectivamente.

Al implementar una segunda capa oculta en la red neuronal, se observó un cambio en el desempeño del modelo en comparación con la arquitectura que contenía una sola capa oculta. Específicamente, se analizó la variación del error cuadrático medio (MSE), identificándose mejoras o deterioros en función de la configuración aplicada. La determinación de la topología óptima de la red reveló que el uso de Análisis de Componentes Principales (PCA) fue eficaz para establecer el número adecuado de capas ocultas [26]. Asimismo, la evaluación exhaustiva de distintas combinaciones en el número de neuronas por capa, siguiendo criterios previamente establecidos en la literatura, permitió afinar la arquitectura final del modelo, contribuyendo a una mejora en la precisión del pronóstico del desempleo [27].

Model: "sequential_57"

Layer (type)	Output Shape	Param #	Trainable
normalization_5 (Normalization)	(None, 7)	15	Y
dropout_52 (Dropout)	(None, 7)	0	Y
dense_154 (Dense)	(None, 16)	128	Y
dense_153 (Dense)	(None, 1)	17	Y

Total params: 160
 Trainable params: 145
 Non-trainable params: 15

Figura 8: Arquitectura de la red neuronal recurrente con una capa oculta

3.1.3. Análisis de Sensibilidad con PCA.

El modelo secuencial de NN con 8 entradas y dos capas ocultas (16 neuronas en la primera capa, óptima tras pruebas entre 8-20 neuronas), logró un Loss de 0.284 y MSE de 0.206 en pruebas con datos no vistos. El análisis reveló que la primera capa es crítica es decir, pequeños cambios en sus pesos afectan más el rendimiento que en la segunda. Algunas entradas son sensibles, es decir, ciertos datos influyen desproporcionadamente en el error, sugiriendo dependencias temporales clave. No hay problemas de gradientes ya que pese a la profundidad, el flujo de errores se

mantuvo estable (tanh/ReLU ayudaron). La arquitectura es robusta, pero conviene monitorear entradas críticas y ajustar regularización. Ideal para series temporales con patrones complejos El PCA como técnica de reducción de dimensiones, permitió calcular el número de variables mediante las cuales se pudo determinar progresivamente la varianza de un conjunto de variables. El PCA se ejecutó mediante la función princomp de R y los resultados se muestran en la Tabla 2 y figura 9.

Tabla 2: Resultados del análisis de componentes principales ejecutado sobre la base de datos

	Comp.1	Comp.2	Comp.3	Comp.4
Standard deviation	35.1015223	16.4474154	8.0623244	4.47688338
Proportion of Variance	0.7687486	0.1687826	0.0405558	0.01250501
Cumulative Proportion	0.7687486	0.9375312	0.9780870	0.99059200
	Comp.5	Comp.6	Comp.7	Comp.8
Standard deviation	3.020677813	1.900832634	1.472044510	0.417319785
Proportion of Variance	0.005693001	0.002253444	0.001351993	0.0001086602
Cumulative Proportion	0.996285003	0.998539347	0.999891340	1.000000000

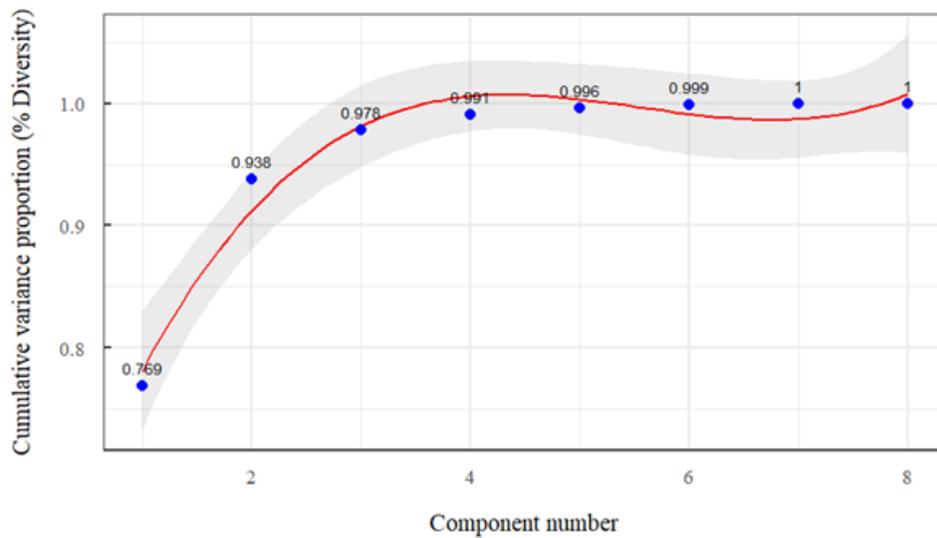


Figura 9: Proporción acumulada de la varianza para cada número de componentes obtenido mediante el PCA

Como se observó en la Figura 8, se necesitó más de 8 componentes principales en el modelo para poder explicar más del 75 % de la varianza presente en los datos. Por lo cual y de acuerdo con [17], se probaron modelos de hasta 2 capas ocultas para configurar la topología óptima de la red neuronal.

3.1.4. Modelo de Deep Learning con dos capas ocultas.

Para continuar explorando el modelo, se añadió una segunda capa oculta con el fin de evaluar si esto genera mejoras en el rendimiento en comparación con la configuración previa. En la primera capa oculta se determinó que 8 neuronas ofrecían el mejor desempeño. Por lo tanto, en esta nueva capa oculta se probó combinaciones que iban desde 8 hasta 20 neuronas, es decir, desde la mitad hasta el doble de neuronas que en la primera capa. Se mantuvo el uso de la función de activación ReLU, con tasa de aprendizaje de 0.1 y una Layer dropout de 0.2. Los resultados de las métricas de rendimiento para las distintas configuraciones de la segunda capa oculta se presentan en la Tabla 3.

Tabla 3: Métricas de desempeño para diferentes configuraciones de redes neuronales con una capa oculta

Number of neurons	Loss	MSE
8	0.3253975	0.2586036
9	0.3584613	0.3152414
10	0.2977118	0.2471480
11	0.3277122	0.2732615
12	0.3142872	0.2550138
13	0.2971343	0.2419511
14	0.3522452	0.3275869
15	0.3093087	0.2364802
16	0.2846069	0.2063614
17	0.2998758	0.2382888
18	0.3691668	0.3620637
19	0.3027082	0.2160653
20	0.3027082	0.9700000
21	0.3027082	1.6400000

En la tabla 3 se apreció varias configuraciones de número de neuronas para el modelo con una sola capa, en este caso las configuraciones que presentaron un mayor desempeño son con 16 y 19 neuronas las cuales alcanzaron las mejores métricas del bucle iterativo que se realizó para encontrar las mejores configuraciones.

3.1.5. Modelo de Deep Learning con dos capas ocultas.

Se volvió a correr el proceso anterior, se añadió una segunda capa oculta con el fin de evaluar si esto genera mejoras en el rendimiento en comparación con la configuración previa. En la primera capa se determinó 16 neuronas para el mejor desempeño. En la segunda capa oculta se evaluó cada modelo posible desde 8 hasta 32 neuronas, es decir, desde la mitad hasta el doble de neuronas de la capa anterior. Se mantuvo el uso de la función de activación ReLU, así como la optimización mediante el modelo Adam con tasa de aprendizaje de 0.1. Los resultados obtenidos para las distintas configuraciones de la segunda capa oculta se presentan en la Tabla 4.

Tabla 4: Métricas de desempeño para diferentes configuraciones de redes neuronales con dos capas ocultas.

Número de Neuronas (Capa 2)	Loss	MSE
9	0.5044433	0.4178960
10	0.5044433	0.4178960
11	0.5044433	0.4178960
12	0.5044433	0.4178960
13	0.5044433	0.4178960
14	0.5044433	0.4178960
15	0.5044433	0.4178960
16	0.5044433	0.4178960
17	0.5044433	0.4178960
18	0.5044433	0.4178960
19	0.5044433	0.4178960
20	0.5044433	0.4178960
21	0.5044433	0.4178960
22	0.5044433	0.4178960
23	0.5044433	0.4178960
24	0.5044433	0.4178960
25	0.5044433	0.4178960
26	0.5044433	0.4178960
27	0.5044433	0.4178960
28	0.5044433	0.4178960
29	0.5044433	0.4178960
30	0.5044433	0.4178960
31	0.5044433	0.4178960
32	0.5044433	0.4178960
33	0.5044433	0.4178960
34	0.5044433	0.4178960
35	0.5044433	0.4178960
36	0.5044433	0.4178960
37	0.5044433	0.4178960
38	0.5044433	0.4178960

En la tabla 4 se muestra varias configuraciones de número de neuronas para el modelo con dos capas ocultas, no existen algunas configuraciones que presenten un mayor desempeño. Es decir, el loss y MSE no representan un cambio menor con respecto a los valores de la tabla 3. El proceso de entrenamiento y arquitectura de la red neuronal con dos capas ocultas se observan en las figuras 10 y 11, a continuación.

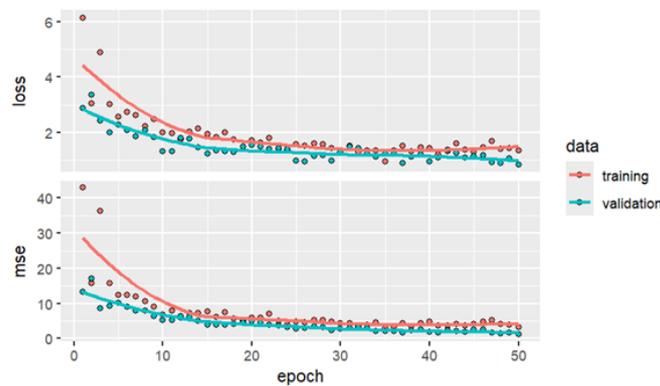


Figura 10: Proceso de entrenamiento de la red neuronal propuesta con dos capas ocultas

```

Model: "sequential_58"
-----
Layer (type)                Output Shape         Param #   Trainable
-----
normalization_5 (Normalizat (None, 7)           15        Y
ion)
dropout_53 (Dropout)        (None, 7)           0          Y
dense_157 (Dense)           (None, 16)          128        Y
dense_156 (Dense)           (None, 17)          289        Y
dense_155 (Dense)           (None, 1)           18         Y
-----
Total params: 450
Trainable params: 435
Non-trainable params: 15
    
```

Figura 11: Arquitectura de la red neuronal propuesta con dos capas ocultas

3.1.6. Modelo de red con pronóstico

El modelo con una sola capa oculta presenta mayor desempeño con respecto a la red con dos capas ocultas en función al MSE igual a 0.2063614, con 16 neuronas para estas capas respectivamente. Por último, con la red de una sola capa oculta presenta un accuracy de 0.75% con 16 neuronas, esto permite pronosticar el desempleo para determinados intervalos de tiempo a futuro con 75% de precisión como esta en la Tabla 5.

Tabla 5: Métricas de desempeño para diferentes configuraciones de redes neuronales con una capa oculta y pronóstico

Número de Neuronas	Loss	MSE	Accuracy
8	0.3253975	0.2586036	0.6000000
9	0.3584613	0.3152414	0.7000000
10	0.2977118	0.2471480	0.6000000
11	0.3277122	0.2732615	0.6000000
12	0.3142872	0.2550138	0.6000000
13	0.2971343	0.2419511	0.5000000
14	0.3522452	0.3275869	0.6500000
15	0.3093087	0.2364802	0.7000000
16	0.2846069	0.2063614	0.7500000
17	0.2998758	0.2382888	0.6500000
18	0.3691668	0.3620637	0.6500000
19	0.3027082	0.2160653	0.6500000
20	0.3027082	0.9700000	0.6500000
21	0.3027082	1.6400000	0.6500000

En la Tabla 6, se observa la red con dos capas ocultas y su accuracy.

Tabla 6: Métricas de desempeño para diferentes configuraciones de redes neuronales con dos capas ocultas y *accuracy*

Número de Neuronas	Loss	MSE	Accuracy
9	0.5044433	0.4178960	0.4100000
10	0.5044433	0.4178960	0.4100000
11	0.5044433	0.4178960	0.4100000
12	0.5044433	0.4178960	0.4100000
13	0.5044433	0.4178960	0.4100000
14	0.5044433	0.4178960	0.4100000
15	0.5044433	0.4178960	0.4100000
16	0.5044433	0.4178960	0.4100000
17	0.5044433	0.4178960	0.4100000
18	0.5044433	0.4178960	0.4100000
19	0.5044433	0.4178960	0.4100000
20	0.5044433	0.4178960	0.4100000
21	0.5044433	0.4178960	0.4100000
22	0.5044433	0.4178960	0.4100000
23	0.5044433	0.4178960	0.4100000
24	0.5044433	0.4178960	0.4100000
25	0.5044433	0.4178960	0.4100000
26	0.5044433	0.4178960	0.4100000
27	0.5044433	0.4178960	0.4100000
28	0.5044433	0.4178960	0.4100000
29	0.5044433	0.4178960	0.4100000
30	0.5044433	0.4178960	0.4100000
31	0.5044433	0.4178960	0.4100000
32	0.5044433	0.4178960	0.4100000
33	0.5044433	0.4178960	0.4100000
34	0.5044433	0.4178960	0.4100000
35	0.5044433	0.4178960	0.4100000
36	0.5044433	0.4178960	0.4100000
37	0.5044433	0.4178960	0.4100000
38	0.5044433	0.4178960	0.6500000

Finalmente, se verificó cada modelo con su desempeño con los datos de prueba que no se observaron en cada modelo durante el proceso de entrenamiento. Este conjunto estuvo conformado por 15 observaciones y se midió su desempeño con las métricas Loss y MSE. Los resultados se muestran en la siguiente tabla 7.

Tabla 7: Comparación de los modelos implementados, evaluados sobre la base de datos Test

Métricas	Modelo secuencial de una capa oculta	Modelo secuencial de dos capas ocultas
Loss	0.2846069	0.5044433
Accuracy	0.7500000	0.4100000
MSE	0.2063614	0.4178960

Como se observó en la Tabla 7, el mejor modelo fue con una capa oculta y 16 neuronas, logrando un desempeño con las métricas de Loss = 0.2846069 y MSE = 0.2063614, es decir, el modelo más efectivo para abordar el modelo y pronóstico del desempleo.

4. Discusion

Los resultados muestran que existen diversas técnicas multivariantes propuestas en varios textos para analizar variables continuas. Al analizar el desempleo en función de las variables predictoras

$$[x_1, \dots, x_7]$$

se observó que técnicas convencionales como las regresiones múltiples no son lo suficientemente robustas para obtener buenos resultados con los distintos modelos, dado el conjunto de datos propuesto.

Por el contrario, las redes neuronales son ensambles de neuronas artificiales y modelos secuenciales, en donde cada una presenta la posibilidad de aprender los comportamientos no lineales partiendo de los datos. Como se pudo visualizar en la Tabla 3, los modelos de Deep Learning alcanzan niveles de desempeño y precisión excelentes en la tarea de encontrar los valores óptimos para el *loss* y el *MSE*. Esto demuestra la capacidad de las redes neuronales (NN) de modelar el desempleo con menor error en la calidad de ajuste del modelo en comparación con la regresión lineal múltiple.

Si relacionamos la información del estudio previo de [7], en donde la investigación se realizó de forma semejante para modelar y predecir el desempleo en función de varias variables, entre ellas el producto interno bruto, se puede apreciar que el uso de las redes neuronales es óptimo con respecto a otros modelos como los obtenidos por regresión lineal, árboles de decisión, etc.

Los hallazgos de este estudio resaltan la superioridad de las NN frente a métodos econométricos tradicionales para modelar y predecir el desempleo en Ecuador. Mientras que investigaciones previas, como las basadas en la Ley de Okun [1, 2, 3, 4], identificaron correlaciones lineales entre variables macroeconómicas y desempleo, sus enfoques estáticos (ej. Mínimos Cuadrados Ordinarios - MCO - o modelos logarítmicos) mostraron limitaciones para capturar dinámicas no lineales y dependencias temporales inherentes a fenómenos socioeconómicos. Por ejemplo, los coeficientes de Okun reportados (entre $-0,84$ y $-0,266$) reflejan relaciones inversas entre PIB y desempleo, pero su capacidad predictiva se ve afectada ante shocks externos o cambios estructurales, comunes en economías en desarrollo como la ecuatoriana.

En contraste, las NN demostraron adaptabilidad al aprender patrones complejos y secuenciales en los datos, incluso con un conjunto limitado de variables (8 predictores). La configuración óptima de una capa oculta con 16 neuronas, que alcanzó un MSE de 0.2064 y precisión del 75 %, sugiere que arquitecturas simples pero especializadas pueden lograr un equilibrio entre rendimiento y eficiencia computacional. Esto contrasta con modelos lineales multivariantes, que subestimaron el impacto no lineal de variables como la inflación (x_2) o la tasa de interés (x_7) en el desempleo.

Adicionalmente, la capacidad de las NN para procesar interdependencias temporales como el efecto retardado del PIB (x_1) o el gasto público (x_6) explica su ventaja en escenarios donde el desempleo responde asincrónicamente a cambios macroeconómicos.

Cabe destacar que variables como el salario mínimo real (x_5) y las exportaciones (x_3) mostraron influencias ambiguas en modelos tradicionales, pero las NN lograron contextualizar su rol dentro de interacciones multivariadas. Esto coincide con hallazgos de [7], donde redes neuronales superaron a árboles de decisión y regresiones lineales al incorporar efectos indirectos, como el impacto de las importaciones (x_4) en la competitividad laboral. No obstante, la escasez de datos históricos de alta frecuencia en Ecuador, especialmente en períodos de crisis, plantea desafíos para generalizar los resultados, un problema común en estudios de economías emergentes [4].

Conclusiones

El presente estudio, enfocado en modelar y predecir el desempleo en Ecuador mediante el uso de redes neuronales, demostró ser óptimo, ayudando a contrastar con otros trabajos similares que el desempleo puede verse influenciado por variables como el gasto público, el salario mínimo y la inflación. El instrumento constó de 8 variables que consideran los factores anteriormente mencionados. Todos los datos obtenidos de cada variable fueron estructurados como observaciones, las cuales ayudaron a determinar las variables automáticas obtenidas por técnicas multivariantes.

En el presente estudio se emplearon redes neuronales con una y dos capas ocultas, y varios modelos de Deep Learning. De todas las técnicas implementadas, se determinó que el modelo de una capa oculta con la configuración de 16 neuronas fue el que logró el mejor desempeño para la detección automática del modelamiento del desempleo, alcanzando un desempeño con las métricas de $Loss = 0.2846069$, $Accuracy = 0.7500000$ y $MSE = 0.2063614$, considerando la existencia de una base de datos de prueba que no fue observada durante el proceso de entrenamiento en las redes neuronales. En contraste, la regresión lineal mostró una calidad de ajuste de 29.73 %.

Se concluye entonces que, para modelar y predecir el desempleo, el uso de redes neuronales (NN) refleja que se pueden obtener resultados óptimos con dichas métricas, por ende, una mejor calidad de ajuste de la variable desempleo en función de las variables predictoras. Este estudio evidencia que las NN son herramientas efectivas para modelar y predecir el desempleo en Ecuador, superando métodos tradicionales como la Ley de Okun o regresiones múltiples. La configuración óptima, una red con una capa oculta de 16 neuronas, logró métricas robustas ($MSE = 0.2064$, precisión del 75 %) al capturar relaciones no lineales y secuenciales entre variables macroeconómicas clave: PIB, inflación, gasto público y tasa de interés. Estos resultados subrayan que el desempleo en Ecuador no solo responde a fluctuaciones económicas directas, sino también a interacciones complejas entre políticas fiscales, variables externas

(exportaciones/importaciones) y factores laborales (salario mínimo).

En concreto, pese al ajuste bajo que refleja el método de mínimos cuadrados ordinarios (OLS) para modelar el desempleo, se puede respaldar con los aportes de la NN con una sola capa oculta y 16 neuronas, concluyendo que el modelo está significativamente condicionado al gasto público, salario mínimo e inflación, con $MSE = 0.20$ para el ajuste del modelo.

La superioridad de las NN radica en su capacidad para adaptarse a la volatilidad estructural de economías en desarrollo, donde shocks externos (ej. crisis globales, fluctuaciones de commodities) alteran abruptamente tendencias históricas. Esto las posiciona como herramientas valiosas para diseñar políticas públicas proactivas, permitiendo simular escenarios bajo diferentes condiciones macroeconómicas.

Sin embargo, se recomienda ampliar el conjunto de variables predictoras — incorporando, por ejemplo, indicadores de informalidad laboral o inversión extranjera — y validar los modelos con datos en tiempo real para mejorar su aplicabilidad. Futuras investigaciones podrían explorar arquitecturas híbridas (ej. NN con atención mecanista) o comparar desempeños con otros algoritmos de inteligencia artificial (ej. Transformers) [7], especialmente en contextos de alta incertidumbre.

Referencias

- [1] M. B. Castro and K. Beltrán, “Análisis correlación del crecimiento económico y la tasa de desempleo en el ecuador (período 2012-2021),” *Ciencia Latina Revista Científica Multidisciplinar*, vol. 7, no. 4, pp. 4020–4035, 2023.
- [2] A. D. C. Vega-Granda, Y. L. Bermeo-Pineda, and F. Y. Vega-Jaramillo, “Solange del cisne gonzález-toro,” vol. 7, pp. 661–680, 2021.
- [3] L. Flores-Cevallos and C. Flores-Tapia, “Análisis econométrico del desempleo y crecimiento económico: aplicación de la ley de okun para la economía ecuatoriana en el período 1990 - 2010,” *Ekotemas*, vol. 3, no. 2, pp. 97–118, 2017. [Online]. Available: <https://www.ekotemas.cu/index.php/ekotemas/article/view/226/146>
- [4] R. Patiño, L. Molero, J. Ruperti, and V. Fernández, “Efectos del pib sobre el desempleo en mujeres y jóvenes en el ecuador (2007-2019),” *Avances en Investigación Científica*, vol. 2, pp. 741–754, 2020. [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=7958752>
- [5] A. Campoverde, C. Ortiz, and V. Sánchez, “Relación entre la inflación y el desempleo: una aplicación de la curva de phillips para ecuador, latinoamérica y el mundo,” *Revista Vista Económica*, vol. 1, no. 1, pp. 22–34, 2016. [Online]. Available: <https://revistas.unl.edu.ec/index.php/economica/article/view/200>
- [6] C. A. Abad-Basantes, J. H. Abad-Basantes, J. C. Ávila Armijos, and L. F. Albuja-Loachamin, “La dinámica de la inflación y el desempleo: Un estudio econométrico basado en la curva de phillips,” *Journal of Economic and Social Science Research*, vol. 4, no. 2, pp. 1–16, 2024.
- [7] K. Mero, N. Salgado, J. Meza, J. Pacheco-Delgado, and S. Ventura, “Unemployment rate prediction using a hybrid model of recurrent neural networks and genetic algorithms,” *Applied Sciences (Switzerland)*, vol. 14, no. 8, 2024.
- [8] I. Bonet Cruz, S. Salazar Martínez, A. Rodríguez Abed, G. Ábalo, M. García Lorenzo, and M. Matilde, “Redes neuronales recurrentes para el análisis de secuencias,” *IEEE*, 2007.
- [9] “Datosmacro,” <https://datosmacro.expansion.com/>, 2024.
- [10] R. T. Yela Burgos, J. L. Rivera Velasco, Y. M. Leturne Macías, and S. M. Moreira Macías, “Efecto de la variación de la producción en el desempleo del ecuador basado en la ley de okun,” *Espergesia*, vol. 11, no. 1, p. e110105, 2024.
- [11] G. Avellán-Solines, “Efectos del dólar y petróleo en el desempleo urbano en ecuador,” *Economía y Negocios*, vol. 9, no. 1, pp. 113–119, 2018.
- [12] L. A. Lopez Nuñez, “Facultad de contabilidad y auditoría,” <https://repositorio.uta.edu.ec/bitstream/123456789/27090/1/T4158e.pdf>, 2021, new England Journal of Medicine, pp. 1–93.
- [13] Banco Central del Ecuador (BCE), “Datos económicos,” 2024.
- [14] “Variables económicas del ecuador,” <https://www.indexmundi.com/g/g.aspx?v=39&c=ec&l=es>, 2024, indexmundi.
- [15] INEC, “Desempleo en el ecuador,” <https://www.ecuadorencifras.gob.ec/estadisticas/>, 2024.
- [16] M. T. Hagan, H. B. Demuth, M. H. Beale, and O. De Jesús, *Neural Network Design 2nd Edition*, 2nd ed., 2014.
- [17] M. I. C. Rachmatullah, J. Santoso, and K. Surendro, “Determining the number of hidden layer and hidden neuron of neural network for wind speed prediction,” *PeerJ Comput Sci*, vol. 7, pp. 1–19, 2021.
- [18] S. Indolia, A. K. Goswami, S. P. Mishra, and P. Asopa, “Conceptual understanding of convolutional neural network- a deep learning approach,” in *Procedia Computer Science*. Elsevier B.V., 2018, pp. 679–688.
- [19] E. P. Herrera-Granda and M. J. Aza-Espinosa, “Assessment of learning achievements in higher education using moodle virtual classrooms and h5p: A case study,” *Preprints*, feb 2025, licencia: CC BY 4.0.
- [20] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, pp. 1735–1780, 1997.
- [21] I. Quinteros, *Modelos econométricos y estadísticos*. Universidad Politécnica Estatal del Carchi-UPEC, 2024.

- [22] C. C. Aggarwal, *Neural Networks and Deep Learning*. Cham: Springer International Publishing, 2018.
- [23] S. Weidman, *Deep Learning from Scratch*, 1st ed. O'Reilly, 2019.
- [24] Y. LeCun and Y. Bengio, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, mayo 2015.
- [25] R. S. Pindyck and D. L. Rubinfeld, *Econometría: Modelos y pronósticos*, 4th ed. McGraw-Hill, 2001.
- [26] E. M. Quindemil Torrijo and F. Rumbaut León, "Análisis de componentes principales para obtener indicadores reducidos de medición en la búsqueda de información," 2019. [Online]. Available: <https://orcid.org/0000-0003-3705-6297>
- [27] M. T. Hagan, H. B. Demuth, M. H. Beale, and O. De Jesús, *Neural Network Design 2nd Edition*, 2nd ed., 2014.