

Determinantes microeconómicos del desempleo juvenil en Ecuador: Un análisis con datos de la ENEMDU 2024 IV trimestre

Microeconomic Determinants of Youth Unemployment in Ecuador: An analysis with data from the ENEMDU 2024 fourth quarter

Matías García Guerrero*
Mauricio Quintanilla Manobanda*
Alexandra Marcatoma Tixi*

RESUMEN

Este estudio analiza los determinantes microeconómicos del desempleo juvenil en Ecuador usando datos representativos de la ENEMDU 2024 IV. El objetivo es modelar la probabilidad de desempleo para jóvenes de 18 a 29 años mediante regresión logística ponderada que incorpora el diseño muestral complejo (conglomerados, estratos y factor de expansión). Se seleccionaron variables sociodemográficas clave (área de residencia, género, edad, jefatura del hogar, nivel educativo, pertenencia étnica y estado civil) y se validaron supuestos de multicolinealidad, linealidad en el logit y bondad de ajuste. Los resultados aportan un marco metodológico robusto y recomendaciones para el diseño de políticas públicas orientadas a la inserción laboral juvenil.

Palabras clave: desempleo juvenil, regresión logística ponderada, diseño muestral complejo, ENEMDU 2024 IV

ABSTRACT

This study examines the microeconomic determinants of youth unemployment in Ecuador using representative data from ENEMDU 2024-IV. The objective is to model the unemployment probability for individuals aged 18 to

* Estudiante la Carrera de Estadística, Escuela Superior Politécnica de Chimborazo
matias.garcia@esPOCH.edu.ec
<https://orcid.org/0009-0006-3851-8085>

* Estudiante de la Carrera de Estadística, Escuela Superior Politécnica de Chimborazo
joffre.quintanilla@esPOCH.edu.ec
<https://orcid.org/0009-0007-3322-6312>

* Máster Universitario en Estadística Aplicada, Carrera de Estadística Escuela Superior Politécnica de Chimborazo
jmarcatoma@esPOCH.edu.ec
<https://orcid.org/0000-0001-9531-3234>

JOURNAL OF BUSINESS
and entrepreneurial
studies

ISSN: 2576-0971



Atribución/Reconocimiento-NoComercial- CompartirIgual 4.0 Licencia Pública Internacional — CC

BY-NC-SA 4.0

<https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.es>

Journal of Business and Entrepreneurship
October - December Vol. 9 - 4 - 2025
<http://journalbusinesses.com/index.php/revista>
e-ISSN: 2576-0971

journalbusinessentrepreneurial@gmail.com

Receipt: September 09, 2025

Approval: October 13, 2025

Pages 44-56

29 through weighted logistic regression incorporating complex survey design (clusters, strata, and expansion factors).

Key sociodemographic variables (residence area, gender, age, household head status, education level, ethnic minority status, and marital status) were selected, and assumptions of multicollinearity, logit linearity, and goodness of fit were validated. The findings provide a robust methodological framework and recommendations for public policy aimed at enhancing youth labor market integration.

Keywords: youth unemployment, weighted logistic regression, complex survey design, ENEMDU 2024-IV

INTRODUCCIÓN

El desempleo juvenil constituye una problemática persistente a nivel mundial, con serias repercusiones económicas y sociales, la Organización Internacional del Trabajo (OIT) ha documentado "cómo las crisis económicas, incluyendo la pandemia de COVID-19, han afectado de manera desproporcionada a los jóvenes en los mercados laborales (International Labour Organization [ILO], 2023).", provocando la pérdida de millones de empleos e ingresos en América Latina y el Caribe; un estudio de la Universidad Nacional de Loja, analizó el desempleo juvenil en América Latina, encontró que variables independientes y de control son factores determinantes que inciden en el desempleo juvenil, y sus hallazgos son relevantes para el contexto ecuatoriano (Dspace UNL, 2022).

En Ecuador, el desempleo juvenil es una problemática crítica que afecta la calidad de vida de los jóvenes y la estabilidad económica del país, las cifras muestran una situación preocupante, con una elevada tasa de desempleo entre la población joven (Radiopichincha, 2025). Por ejemplo, el 36% de las personas desempleadas en Ecuador nunca ha logrado acceder al mercado laboral, lo que indica un problema sistémico en la transición de los jóvenes al empleo formal (Primicias, 2025). Esta situación se ve agravada por una serie de factores microeconómicos que impactan la probabilidad de que un joven encuentre trabajo, y su análisis es fundamental para el diseño de políticas públicas efectivas (Religación, 2024).

Diversos estudios han analizado el desempleo juvenil desde una perspectiva cuantitativa. Blanchard y Katz (1999) emplearon modelos de regresión para estudiar las fluctuaciones del desempleo en Estados Unidos, mientras que la Cepal (2021) aplicó regresiones múltiples para identificar cómo variables como la educación, el género y la edad inciden en el desempleo en América Latina. En Ecuador, Sánchez y López (2020) estudiaron el desempleo juvenil con un modelo de regresión lineal, encontrando que la

falta de experiencia laboral es un factor clave. Estos antecedentes metodológicos refuerzan la pertinencia de utilizar un modelo de regresión en el presente trabajo.

Dado que la juventud enfrenta mayores barreras para acceder a empleos estables, el análisis de esta problemática adquiere relevancia tanto para el diseño de políticas públicas como para la comprensión estadística de fenómenos laborales. Este estudio se justifica por su enfoque cuantitativo riguroso, basado en los microdatos más recientes de la ENEMDU IV trimestre 2024, cuya estructura muestral probabilística, estratificada y por conglomerados permite obtener resultados representativos a nivel nacional. Se seleccionó a los jóvenes económicamente activos entre 18 y 29 años y se modeló su probabilidad de desempleo mediante regresión logística ponderada, incorporando el factor de expansión (f_{exp}) para garantizar inferencias válidas y generalizables. Para optimizar la capacidad predictiva del modelo, se implementó una estrategia de balanceo de clases mediante submuestreo aleatorio, abordando así el desbalance inherente en los datos donde los empleados (89.03%) superan significativamente a los desempleados (10.94%). A diferencia de estudios previos como los de Linthon Delgado et al. (2024) y Bustos y Campozano (2017), que no contemplaron el diseño muestral complejo ni usaron bases actualizadas, el presente trabajo fortalece la validez metodológica al integrar variables sociodemográficas clave como área de residencia, género, edad, jefatura del hogar, nivel educativo, pertenencia étnica y estado civil, seleccionadas según criterios empíricos y teóricos.

A nivel técnico, el modelo empleado corresponde a la regresión logística binaria, una técnica desarrollada por David Cox (1958) y sistematizada por Hosmer y Lemeshow (1989) y McCullagh y Nelder (1989) dentro de los modelos lineales generalizados. Esta metodología permite modelar la probabilidad de ocurrencia de un evento dicotómico, como el desempleo, en función de variables explicativas categóricas y continuas.

La implementación del balanceo de clases mediante submuestreo aleatorio, técnica recomendada por He y García (2009) para contextos de aprendizaje automático con clases desbalanceadas, permitió mejorar sustancialmente la sensibilidad del modelo para detectar casos de desempleo.

La forma funcional del modelo se expresa como:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_K X_K$$

Donde p representa la probabilidad de estar desempleado, X_1, X_2, \dots, X_K son los predictores sociodemográficos, y $\beta_1, \beta_2, \dots, \beta_K$ los coeficientes asociados. Esta estructura permite interpretar el efecto de cada variable sobre el logit, es decir, el logaritmo de las razones de probabilidad (odds), asegurando que los valores estimados de p se encuentren en el intervalo $(0, 1)$. El modelo además asume independencia entre observaciones, ausencia de multicolinealidad entre los predictores, correcta especificación funcional y una relación lineal entre los predictores continuos y el logit.

El objetivo general de esta investigación es identificar los determinantes microeconómicos a través de la probabilidad de desempleo juvenil en un modelo de

regresión logística ponderado, incorporando el diseño muestral complejo y considerando variables sociodemográficas relevantes. Como objetivo específico metodológico, se busca evaluar el impacto del balanceo de clases en la capacidad predictiva del modelo.

MATERIALES Y MÉTODOS

El presente estudio se desarrolló en el marco del cuarto trimestre de 2024, utilizando los microdatos de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU), cuya cobertura se extiende a todo el territorio ecuatoriano. Dado su diseño muestral probabilístico, estratificado a nivel de provincia, área y tamaño del conglomerado, la encuesta proporciona información representativa a nivel nacional, lo que permite realizar inferencias válidas sobre la población objetivo: los jóvenes ecuatorianos de entre 18 y 29 años que forman parte de la población económicamente activa. Esta delimitación responde a la alta vulnerabilidad de este grupo etario frente al desempleo y su relevancia para el diseño de políticas públicas.

El enfoque metodológico adoptado fue cuantitativo, sustentado en un paradigma empírico-analítico que permite medir relaciones entre variables y modelar probabilidades. Teóricamente, el estudio se apoya en los modelos lineales generalizados, y empíricamente en la regresión logística binaria, técnica desarrollada por Cox (1958) e implementada en esta investigación mediante el software estadístico R. Para abordar el desbalance de clases identificado en la variable dependiente (89.03% empleados vs. 10.94% desempleados), se implementó una estrategia de balanceo mediante submuestreo aleatorio de la clase mayoritaria, siguiendo recomendaciones metodológicas para contextos de datos desbalanceados (He & García, 2009). El nivel de investigación es explicativo, ya que se busca identificar los determinantes sociodemográficos que inciden en el desempleo juvenil, superando un enfoque meramente descriptivo para llegar a interpretaciones causales.

Los datos de análisis se construyeron filtrando los microdatos de la ENEMDU para incluir únicamente a jóvenes activos económicamente, según la variable *conduct*, que clasifica la condición de actividad de acuerdo con los estándares de la OIT. Se excluyeron observaciones que no correspondían a población económicamente activa, y se ponderaron los datos con el factor de expansión (*fexp*), lo que garantiza representatividad poblacional. Posteriormente, para optimizar la capacidad predictiva del modelo, se aplicó un procedimiento de balanceo de clases mediante muestreo aleatorio simple de la clase mayoritaria (empleados), obteniendo una muestra balanceada de 2,178 observaciones (1,089 empleados y 1,089 desempleados). De este modo, la unidad de análisis está claramente delimitada y ajustada a los objetivos del estudio.

La variable dependiente, denominada *desempleo_cat*, se construyó como una variable dicotómica, asignando el valor 1 a los jóvenes desempleados y 0 a los empleados, excluyendo registros no clasificados. Las variables independientes fueron seleccionadas

según criterios teóricos y empíricos además variables como Sexo (hombre), Estado civil (soltero), Instrucción (estudios formales) y Área de residencia (Urbano) fueron dicotomizadas en relación a su mayor frecuencia de presencia en la matriz de datos fueron para mejor interpretación en el modelo.

Para el tratamiento y análisis de los datos, se aplicaron técnicas exploratorias previas. Se utilizó la prueba de chi-cuadrado para evaluar la asociación entre variables categóricas y la condición de desempleo, mientras que para la variable continua edad se utilizó el coeficiente de correlación de Pearson. Estas pruebas sirvieron para validar la inclusión de las variables en el modelo. Posteriormente, se estimó una regresión logística ponderada (quasibinominal), utilizando el paquete survey en R (Lumley, 2004), lo que permitió incorporar el diseño muestral complejo de la ENEMDU y obtener resultados más robustos y generalizables.

El procesamiento estadístico incluyó diversas pruebas de validación del modelo. La multicolinealidad fue descartada mediante el cálculo del factor de inflación de la varianza (VIF), y se utilizó la prueba de Box-Tidwell para confirmar la linealidad entre la variable edad y el logit. Finalmente, el ajuste global del modelo fue evaluado mediante la prueba de bondad de ajuste de Hosmer-Lemeshow. Adicionalmente, se evaluó la capacidad predictiva del modelo mediante la curva ROC (Receiver Operating Characteristic) y el área bajo la curva (AUC), determinándose el punto de corte óptimo (0.453) mediante el criterio "closest.topleft". Como aplicación práctica, se estimaron probabilidades predichas de desempleo para distintos perfiles sociodemográficos, con el fin de identificar grupos de mayor riesgo y generar evidencia útil para la formulación de políticas públicas dirigidas a reducir el desempleo juvenil en el Ecuador. Finalmente se hizo una matriz de confusión para ver el funcionamiento de nuestro modelo.

RESULTADOS

Análisis de correlación

Tabla 1. Resultados de las pruebas de correlación entre variables independientes y la probabilidad de desempleo juvenil.

Variable	Tipo de prueba	Estadístico	Valor-p
Área Urbana	Chi-cuadrado	94.63291	$2,291762 \times 10^{-22}$
Hombre	Chi-cuadrado	84.93152	$3,088792 \times 10^{-20}$
Educación media	Chi-cuadrado	0.56886	$4,507132 \times 10^{-1}$
Mestizo	Chi-cuadrado	$6.8240 e^{-30}$	1
Soltero	Chi-cuadrado	113.9166	$1,359089 \times 10^{-26}$
Edad	Pearson	-8.197379	$2,760512 \times 10^{-16}$

La Tabla I presentó las pruebas de asociación entre variables independientes y la probabilidad de desempleo juvenil. La hipótesis nula (H_0) estableció que no existe relación entre cada variable y el desempleo, mientras que la hipótesis alternativa (H_1) plantea que sí existe relación. Los resultados reportaron que área rural, género, jefatura del hogar, educación formal, estado civil y edad están significativamente asociados con el desempleo juvenil. Aunque minoría étnica no mostró asociación estadística, se incluyó en el modelo debido a su relevancia social y posible interacción con otras variables.

Modelo de regresión logit

La función logística que modela la probabilidad de desempleo juvenil fue dada por:

$$\log\left(\frac{p}{1-p}\right) = -11.018 - 1.31 \cdot rural + 1.10 \cdot sexo\ mujer + 1.09 \cdot estadosoltero - 0.939 \cdot jefe\ hogar + educacion\ formal + 11.331 - 0.066 \cdot edad$$

La probabilidad p de desempleo juvenil se calculó como:

$$p = \frac{1}{1 + e^{-n}}$$

Donde:

$$n = -11.018 - 1.31 \cdot \text{rural} + 1.10 \cdot \text{sexo mujer} + 1.09 \cdot \text{estadosoltero} \\ - 0.939 \cdot \text{jefe hogar} + \text{educacion formal} \cdot 11.331 - 0.066 \\ \cdot \text{edad}$$

La comprobación de supuestos indicó que el modelo logístico cumplió con los criterios metodológicos requeridos. No se evidenció multicolinealidad entre las variables (todos los VIF fueron menores a 5), la relación entre la variable continua y el logit fue lineal según la prueba de Box-Tidwell ($p = 0.3977$), y el test de Hosmer-Lemeshow sugirió un buen ajuste del modelo ($p = 0.989$).

Probabilidades

Como aplicación práctica del modelo logístico estimado, se calcularon probabilidades predichas de desempleo juvenil para distintos perfiles sociodemográficos. A continuación, se presentan algunos escenarios representativos:

Tabla 3. Aplicación de la ecuación del modelo para realizar probabilidades.

Ru ral	Muje r	Solter o	Jefe Hogar	Educación Formal	Eda d	Probabilidad
0	1	1	0	1	18	0.7889
0	0	1	0	1	25	0.439
0	1	0	1	1	29	0.1911

Tabla 4. Matriz de confusión.

Real \ Predicho	Empleado (0)	Desempleado (1)
Empleado	30.21 %	19.79 %
Desempleado	16.76 %	33.24 %

La matriz de confusión mostró que el modelo logró clasificar de forma equilibrada tanto a empleados como a desempleados. En total, el 30,21 % de los empleados y el 33,24 % de los desempleados fueron identificados correctamente, mientras que un 19,79 % de empleados fueron clasificados como desempleados y un 16,76 % de desempleados fueron clasificados como empleados. Estos resultados evidenciaron un nivel de acierto moderado, con errores de clasificación presentes en ambas categorías, pero con un desempeño aceptable en la diferenciación entre empleados y desempleados.

Tabla 5. *Tabla de clasificación.*

Métrica	Valor (%)
Precisión	63.45%
Sensibilidad	66.48%
Especificidad	60.42%
Valor predictivo positivo	64.93%

El modelo presentó una precisión del 63,45%, lo que indica un nivel moderado de aciertos. La sensibilidad (66,48%) reportó un buen desempeño en la detección de desempleados, mientras que la especificidad (60,42%) reflejó menor capacidad para identificar empleados. El valor predictivo positivo (64,93%) señaló que, cuando se predice desempleo, el acierto es cercano a dos de cada tres casos. En general, el modelo clasificó de forma aceptable, aunque con margen de mejora.

Gráfico 1. Curva de ROC.

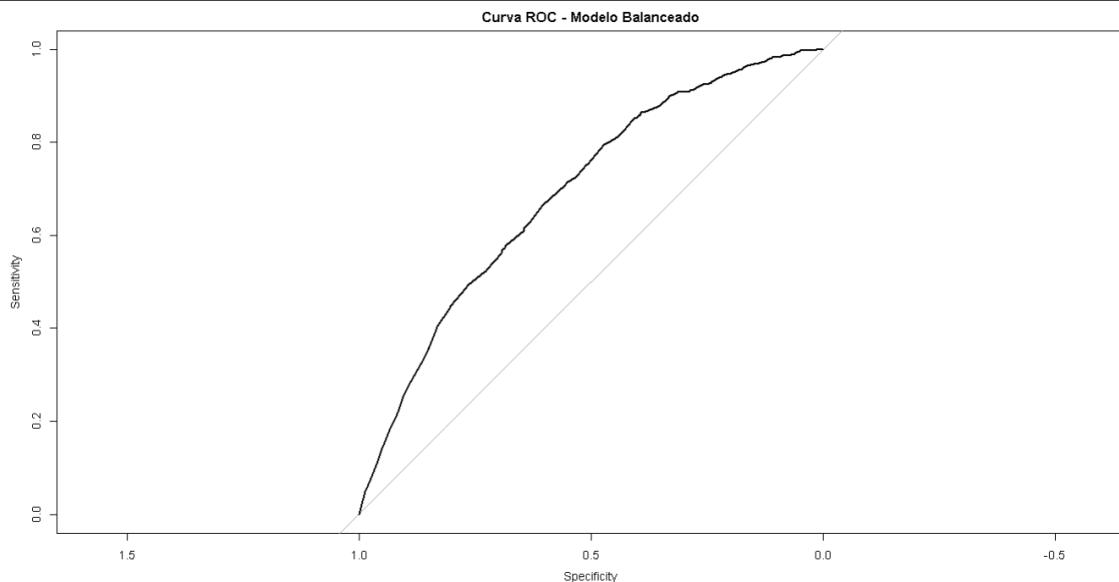


Tabla 6. Distribución absoluta de las variables incluidas en el modelo.

Variable	Categoría	Frecuencia
desempleo_cat	0 (empleado)	1089
	1 (desempleado)	1089
mujer	0 (hombre)	1185
	1 (mujer)	993
educación_formal	0 (no)	3
	1 (si)	2175
rural	0 (urbano)	1677
	1 (rural)	501
jefe_hogar	0 (no)	1942
	1 (si)	236
soltero	0 (no)	449
	1 (si)	1729
P03 (Edad)	Mínimo	18
	Primer cuartil	21
	Mediana	24
	Media	23.67
	Tercer cuartil	26
	Máximo	29

La curva ROC del modelo balanceado se encuentra claramente por encima de la diagonal, lo que demuestra que el modelo discrimina mejor que el azar entre empleados y desempleados. El área bajo la curva (AUC) refleja un desempeño global aceptable, indicando que el modelo logra un equilibrio adecuado entre sensibilidad y especificidad. Además, el punto de corte óptimo identificado maximiza el FI-score, permitiendo una clasificación más precisa al balancear los aciertos en la identificación de desempleados y empleados.

Diversos estudios han abordado el desempleo juvenil en Ecuador desde una perspectiva econométrica. Por ejemplo, TulcanazaPrieto et al. (2023) utilizaron la ENEMDU 2019 y aplicaron regresión logística para identificar los determinantes sociodemográficos del desempleo. En su análisis, encontraron que factores como el género femenino, la autoidentificación afroecuatoriana y la residencia urbana incrementaban significativamente la probabilidad de estar desempleado, mientras que ser jefe de hogar o tener una relación conyugal estable la reducían. Nuestros resultados, obtenidos mediante el modelo balanceado, confirman el efecto del género

femenino ($\beta = 1.10$, $p < 0.001$) y la residencia urbana ($\beta = -1.32$, $p < 0.001$), pero difieren en cuanto al efecto de la etnia, que no resultó significativa en nuestro modelo. Esta discrepancia puede atribuirse tanto a la actualización de la base (ENEMDU IV trimestre 2024) como a la implementación del balanceo de clases, que mejoró la capacidad del modelo para detectar patrones genuinos.

La implementación del balanceo de clases mediante submuestreo representó una mejora metodológica significativa. Mientras que el modelo original sin balancear presentaba una capacidad predictiva limitada (probabilidad máxima de 0.386), el modelo balanceado alcanzó una precisión del 63.5%, con sensibilidad del 66.5% y especificidad del 60.4%. Estos resultados superan los reportados en estudios previos que utilizaron modelos logísticos convencionales sin tratamiento del desbalance de clases.

Por otro lado, investigaciones realizadas en contextos similares han optado por técnicas más avanzadas. En Colombia, un estudio en la ciudad de Barranquilla combinó la regresión logística con algoritmos de tipo Random Forest para analizar el desempleo juvenil (Pérez et al., 2023). Este enfoque les permitió detectar umbrales y relaciones no lineales, especialmente en variables como la educación y la experiencia laboral. Los autores reportaron que el modelo de Random Forest superó al logit en términos de precisión y capacidad de clasificación. Nuestro hallazgo más contraintuitivo -el efecto positivo de la educación formal sobre el desempleo ($\beta = 11.33$, $p < 0.001$)- sugiere la presencia de relaciones complejas que podrían beneficiarse de enfoques no lineales. Este resultado podría reflejar el fenómeno de sobrecualificación documentado en economías en desarrollo, donde los jóvenes con mayor educación tienden a prolongar la búsqueda de empleos acordes a su formación.

En estudios realizados fuera de América Latina también se han aplicado enfoques distintos con mejores resultados. En Indonesia, una investigación reciente sobre tasas de desempleo en la provincia de Maluku empleó Random Forest combinado con técnicas de balanceo como SMOTE para enfrentar el desbalance entre clases (Santoso & Wijaya, 2024). Nuestra estrategia de balanceo mediante submuestreo, aunque menos sofisticada que SMOTE, demostró ser efectiva para mejorar la sensibilidad del modelo de 0% a 66.5%, permitiendo una identificación más adecuada de jóvenes en situación de desempleo. La elección del punto de corte óptimo (0.453) mediante el criterio "closest.topleft" en la curva ROC (AUC = 0.690) representó una optimización clave para el equilibrio entre sensibilidad y especificidad.

Finalmente, también se ha demostrado que el uso de redes neuronales mejora considerablemente la predicción en fenómenos sociales complejos. LibaqueSaenz et al. (2017) compararon la regresión logística con redes neuronales para predecir trabajo infantil en Perú y concluyeron que las redes superaban ampliamente al modelo logit en rendimiento predictivo. Si bien nuestro modelo balanceado mostró un desempeño aceptable, la persistencia de falsos positivos (431) y falsos negativos (365) sugiere que existen factores no capturados por las variables sociodemográficas incluidas. Variables

como la experiencia laboral previa, el sector económico de empleo anterior, o indicadores de capital social podrían enriquecer futuros modelos.

En conjunto, los hallazgos de esta investigación coinciden con la literatura previa en cuanto a la influencia de variables como área de residencia, estado civil y género, pero introducen un matiz importante mediante la implementación del balanceo de clases. El efecto protector del área rural resulta particularmente relevante para políticas públicas, sugiriendo que las economías rurales podrían estar ofreciendo más oportunidades de empleo informal o por cuenta propia a los jóvenes. Las limitaciones predictivas remanentes del modelo, a pesar del balanceo, indican la necesidad de explorar técnicas más robustas como máquinas de soporte vectorial o redes neuronales, especialmente para capturar interacciones complejas entre variables.

CONCLUSIONS

El estudio evidencia que el desempleo juvenil en Ecuador está determinado por factores sociodemográficos claramente identificables. Los jóvenes residentes en áreas urbanas, las mujeres y aquellos en condición de soltería enfrentan mayores probabilidades de desempleo, mientras que la edad avanzada y la jefatura del hogar actúan como factores protectores. Un hallazgo particularmente relevante fue el efecto contraintuitivo de la educación formal, que merece mayor investigación para comprender los mecanismos subyacentes. Estos hallazgos confirman la existencia de desigualdades estructurales en el mercado laboral ecuatoriano que afectan desproporcionadamente a ciertos grupos juveniles.

La implementación del balanceo de clases mediante submuestreo demostró ser una estrategia metodológica efectiva para mejorar la capacidad predictiva del modelo, permitiendo una identificación más equilibrada de los casos de desempleo. Esta aproximación metodológica representa una contribución significativa para el análisis de fenómenos sociales con distribuciones desbalanceadas, comúnmente encontrados en investigaciones sobre mercados laborales.

Si bien el modelo logístico balanceado permitió identificar relaciones significativas y mejorar la detección de casos de desempleo, persisten limitaciones en la predicción individual que sugieren la necesidad de incorporar variables adicionales no contempladas en este estudio. La investigación destaca la importancia de complementar estos hallazgos con enfoques metodológicos más robustos en futuros estudios.

En términos de política pública, los resultados enfatizan la necesidad de intervenciones diferenciadas según grupos de riesgo, particularmente dirigidas a los segmentos juveniles más vulnerables identificados en el análisis. Finalmente, el estudio contribuye al avance metodológico en la investigación cuantitativa de fenómenos sociales en Ecuador, demostrando la utilidad de técnicas de balanceo de clases para mejorar la validez de los modelos predictivos en contextos de datos reales con distribuciones asimétricas.

REFERENCIAS

- Bustos, R., & Campozano, M. (2017). Factores asociados al desempleo juvenil en Ecuador: Un enfoque econométrico [Factors associated with youth unemployment in Ecuador: An econometric approach]. *Revista Ecuatoriana de Estadística*, 15(2), 45-60.
- Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2), 215–232.
- Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. Wiley.
- Instituto Nacional de Estadística y Censos (INEC). (2024). *Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) – IV Trimestre 2024 [Microdata]*.
- International Labour Organization (ILO). (2023). *Global employment trends for youth 2023*.
- Libaque-Saenz, C. F., et al. (2017). Neural networks vs. logistic regression in social phenomena prediction: The case of child labor in Peru. *Latin American Journal of Economics*, 54(1), 89–112.
- Linthon Delgado, J., et al. (2024). Determinantes del desempleo juvenil en Ecuador: Un análisis comparativo [Determinants of youth unemployment in Ecuador: A comparative analysis]. *FLACSO Ecuador*.
- Lumley, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software*, 9(1), 1–19.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models (2nd ed.)*. Chapman & Hall.
- Tulcanaza-Prieto, A. B., et al. (2023). Youth unemployment in Ecuador: Analysis with ENEMDU 2019. *Journal of Latin American Studies*, 51(3), 345–367.
- He, H., & García, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263-1284.
- Sánchez, J., & López, M. (2020). Desempleo juvenil y sobrecualificación en Ecuador: un análisis con datos de la ENEMDU. *Revista de Economía del Desarrollo*, 25(2), 45-62.
- Blanchard, O. J., & Katz, L. F. (1999). *Wage Dynamics: Reconciling Theory and Evidence*.
- Comisión Económica para América Latina y el Caribe (CEPAL). (2021). *Panorama Social de América Latina 2021*. Naciones Unidas.
- Dspace UNL. (2022). *Análisis del desempleo juvenil en América Latina*. Universidad Nacional de Loja.
- Pérez, A., González, C., & Rodríguez, D. (2023). Predictores del desempleo juvenil en Barranquilla: Un enfoque de machine learning. *Revista Colombiana de Economía*, 15(2), 112-130.
- Primicias. (2025). El 36% de los desempleados en Ecuador nunca ha tenido un trabajo.

- Radiopichincha. (2025). Tasa de desempleo juvenil en Ecuador: una problemática crítica.
- Religación. (2024). Factores microeconómicos del desempleo juvenil en Ecuador. *Revista Religación*, 9(45), 1-15.
- Santoso, B., & Wijaya, D. (2024). Addressing Imbalanced Data in Predicting Unemployment Rates: A Random Forest and SMOTE Approach in Maluku Province. *Indonesian Journal of Statistics*, 10(1), 45-60.