

Identificación de variables determinantes en la situación ocupacional en Ecuador durante el año 2023

Identification of determining variables in the occupational situation in Ecuador during the year 2023

Christian Xavier Bueno Lliguín

Ingeniero en Estadística
Investigador externo

<https://orcid.org/0009-0009-9585-7771>
diseniarte2018@gmail.com

Melanie Nicole Galarza Reyes

Ingeniera en Estadística
Investigadora externa

<https://orcid.org/0009-0007-3810-4744>
melani.galarza@esepoch.edu.ec

Johanna Enith Aguilar Reyes

Ingeniera en Estadística Informática
Docente de la Escuela Superior Politécnica de Chimborazo
Investigadora del grupo CIED

<https://orcid.org/0000-0002-1230-2503>
johannae.aguilar@esepoch.edu.ec

Tania Paulina Morocho Barrionuevo

Biofísica
Docente de la Escuela Superior Politécnica de Chimborazo
<https://orcid.org/0000-0002-1019-6049>
tpaulina.morocho@esepoch.edu.ec

Edwin Fernando Mejía Peñafiel

Ingeniero en Sistemas
Docente de la Escuela Superior Politécnica de Chimborazo
<https://orcid.org/0009-0007-3810-4740>
efmejia@esepoch.edu.ec



Imaginario Social
Entidad editora
REDICME (reg-red-18-0061)

e-ISSN: 2737-6362
enero-junio 2025 Vol. 8-1-2025
<http://revista-imaginariosocial.com/index.php/es/index>

Recepción: 16 de agosto de 2025
Aceptación: 20 de septiembre de 2025

38-61

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>

Resumen

Ecuador no presenta factores establecidos que influyan en la situación ocupacional o laboral de los ecuatorianos, haciendo uso de modelos de regresión logística y árboles de clasificación para identificar y comparar los determinantes decisivos, por lo tanto, el objetivo de la presente investigación es comparar la eficacia de los modelos para determinar los factores influyentes en la situación laboral en Ecuador. La metodología implementada tuvo un enfoque cuantitativo, usando un diseño no experimental de tipo transversal sin la manipulación de variables establecidas dentro de un determinado periodo de tiempo; la población considerada para el estudio son las personas económicamente activas en el Ecuador.

Esta metodología permitió identificar y comprender de manera profunda los diversos factores que se puedan considerar como influyente en la situación laboral de los ecuatorianos considerados como limitantes al encontrar empleo.

En este contexto se concluye que al evaluar la eficacia de dos modelos estadísticos, regresión logística y árboles de clasificación presentan resultados muestran similar en cuanto al error de predicción y la exactitud, pero una diferencia en cuanto a la eficacia del modelo de regresión del 73% considerado regular pero no el ideal y árboles de clasificación 70%, donde el modelo de regresión logística identifica variables significativas, tales como sexo, área, nivel de instrucción, estado civil y edad el cual no considera provincia como un factor influyente respecto a que el otro modelo si lo considera. Esta investigación es de gran aporte para el estado ecuatoriano debido a su crecimiento poblacional que es constante por lo que es necesario identificar factores influyentes en cuanto a la situación laboral de los y las ecuatorianas con la finalidad de brindarles mejores oportunidades de acuerdo con sus cualidades y características representativas.

Palabras clave: árboles de clasificación, regresión logística, estadística, eficacia, modelos.

Abstract

Ecuador does not present established factors that influence the occupational or labor situation of Ecuadorians. Using logistic regression models and classification trees to identify and compare decisive determinants, the objective of this research is to

compare the effectiveness of these models in determining the factors that influence employment conditions in Ecuador.

The methodology implemented had a quantitative approach, employing a non-experimental cross-sectional design without manipulation of established variables within a specific period of time. The population considered for the study consists of economically active individuals in Ecuador.

This methodology made it possible to identify and gain a deep understanding of the various factors that may be considered as influential in the employment situation of Ecuadorians, regarded as limitations when seeking employment.

In this context, it is concluded that when evaluating the effectiveness of two statistical models logistic regression and classification trees the results show similar performance in terms of prediction error and accuracy. However, there is a difference in model effectiveness: logistic regression achieves 73%, considered fair but not ideal, while classification trees reach 70%. The logistic regression model identifies significant variables such as sex, area, level of education, marital status, and age, while excluding province as an influencing factor, which is considered relevant by the classification tree model.

This research provides valuable contributions for the Ecuadorian state due to its constant population growth, making it essential to identify factors that influence the employment situation of Ecuadorian men and women in order to provide them with better opportunities aligned with their qualities and representative characteristics.

Keywords: classification trees, logistic regression, statistics, effectiveness, models.

Introducción

Para comprender mejor el empleo en la realidad ecuatoriana, resulta pertinente identificar los elementos determinantes en la situación laboral de los ecuatorianos, donde los modelos de regresión logística y árboles de clasificación podrían ayudar a comprender mejor la interrelación entre los principales desafíos del mercado laboral ecuatoriano al identificar los factores clave que influyen en cada problemática y prever las estrategias más efectivas para abordarlas, ofreciendo un enfoque más preciso y cuantitativo en la formulación de políticas y soluciones (Arellano Estrada y Ayaviri Nina 2021).

Con base en el artículo Decision Sciences se compara la precisión predictiva de pacientes con riesgo de reingreso con los modelos: regresión logística, árboles de regresión, modelos aditivos generalizados (GAM) y Splines de regresión adaptativa multivariante (MARS). La precisión predictiva se evaluó utilizando las Curvas ROC en una muestra de validación. Los resultados concluyen que tanto los modelos de regresión logística como los árboles de clasificación tuvieron un rendimiento comparable a los modelos más flexibles basados en GAM y MARS (Austin 2007).

De acuerdo con una comparativa "Classification Trees vs. Logistic Regression in the Generic Skill Development in Engineering". Se realiza una comparación experimental entre los dos modelos en el contexto de competencias genéricas en ingeniería, específicamente en razonamiento cuantitativo y comprensión lectora. Con la finalidad de mejorar el rendimiento de estos modelos. Con resultados que cuando se usa exclusivamente constructos bajo regresión logística, ni el análisis de varianza (ANOVA) ni las curvas de característica operativa del receptor (ROC) con área bajo la curva (AUC) reflejan diferencias en el rendimiento de los métodos (Pérez Rave y González Echavarría 2018).

Sin embargo, hasta el momento no se ha evaluado adecuadamente qué método analítico, entre la regresión logística y los árboles de decisión, es más efectivo para identificar con precisión los factores determinantes en la situación laboral ecuatoriana. Esto limita la capacidad para comprender y predecir con exactitud los aspectos clave que influyen en el ámbito laboral del país (Congacha Ortega 2020).

Citando a (Flores y Salto 2018) deduce que realiza una comparativa entre tres métodos de aprendizaje automático ampliamente utilizados, árboles de clasificación, bosques aleatorios y aumento de gradiente, aplicados a la predicción de la satisfacción laboral en Ecuador. Los resultados revelaron que el método de aumento de gradiente superó a los árboles de clasificación y los bosques aleatorios en términos de precisión en la predicción de la satisfacción laboral. Sin embargo, se observó que todos los métodos proporcionaron información valiosa sobre los factores que afectan la satisfacción laboral en Ecuador, lo que puede ser útil para implementar intervenciones y políticas laborales efectivas.

Antecedentes

Según el (INEC, ENEMDU 2023), señala que, en los últimos años la situación laboral de los ecuatorianos ha mejorado en algunos aspectos, pero aún existen desafíos importantes. Según la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) de 2023, la tasa de desempleo se situó en el 4,8%, la tasa de subempleo en el 63,6% y la tasa de informalidad en el 47,9%. La comprensión de los factores que influyen en esta situación es fundamental para el desarrollo de políticas públicas que permitan mejorarla. Los factores que influyen en la situación laboral de los ecuatorianos son diversos y complejos. Entre los factores más importantes se encuentran: educación, experiencia laboral, sexo, edad y ubicación.

Modelos Estadísticos en estudio

Con base en el artículo Decision Sciences se compara la precisión predictiva de pacientes con riesgo de reingreso con los modelos: regresión logística, árboles de regresión, modelos aditivos generalizados (GAM) y Splines de regresión adaptativa multivariante (MARS). La precisión predictiva se evaluó utilizando las Curvas ROC en una muestra de validación. Los resultados concluyen que tanto los modelos de regresión logística como los árboles de clasificación tuvieron un rendimiento comparable a los modelos más flexibles basados en GAM y MARS (Austin 2007).

De acuerdo con una comparativa "Classification Trees vs. Logistic Regression in the Generic Skill Development in Engineering". Se realiza una comparación experimental entre los dos modelos en el contexto de competencias genéricas en ingeniería, específicamente en razonamiento cuantitativo y comprensión lectora. Con la finalidad de mejorar el rendimiento de estos modelos. Con resultados que cuando se usa exclusivamente constructos bajo regresión logística, ni el análisis de varianza (ANOVA) ni las curvas de característica operativa del receptor (ROC) con área bajo la curva (AUC) reflejan diferencias en el rendimiento de los métodos (Pérez Rave y González Echavarría 2018).

Investigaciones Similares

Citando a (Flores y Salto 2018) deduce que realiza una comparativa entre tres métodos de aprendizaje automático ampliamente utilizados, árboles de clasificación, bosques

aleatorios y aumento de gradiente, aplicados a la predicción de la satisfacción laboral en Ecuador. Los resultados revelaron que el método de aumento de gradiente superó a los árboles de clasificación y los bosques aleatorios en términos de precisión en la predicción de la satisfacción laboral. Sin embargo, se observó que todos los métodos proporcionaron información valiosa sobre los factores que afectan la satisfacción laboral en Ecuador, lo que puede ser útil para implementar intervenciones y políticas laborales efectivas.

De acuerdo con (Pérez Rave y González Echavarría 2018) menciona que, en el campo del desarrollo de competencias genéricas en ingeniería, es crucial identificar la metodología más adecuada para analizar y predecir el rendimiento de los estudiantes. En este estudio, se compara el uso de árboles de clasificación y regresión logística con el fin de determinar cuál de estas técnicas es más efectiva. Se analizan datos recopilados de estudiantes de ingeniería en varias áreas para evaluar la relación entre las variables predictoras y el éxito académico. Los resultados revelan las ventajas y limitaciones de cada método, proporcionando información valiosa para futuras investigaciones y prácticas pedagógicas en el ámbito de la ingeniería.

En un estudio realizado en Corea se señala que la satisfacción con las condiciones de trabajo se vincula con la salud y productividad de los trabajadores, y se demostró que, para los trabajadores coreanos, los principales factores que afectan a la satisfacción con las condiciones de trabajo han cambiado con el tiempo. Los aspectos que tienen mayor impacto en la satisfacción con las condiciones laborales son los siguientes: el ambiente físico, interacciones sociales negativas, puesto ocupacional y calidad de la gestión (Lee y Park 2021).

En base al estudio referente a la satisfacción laboral señalan que el nivel educativo tiene un impacto continuo en la satisfacción laboral, mientras que el estilo de liderazgo y la competencia son factores de profunda influencia. Además, se valida la confiabilidad del marco de integración de las metodologías del Laboratorio de Ensayos y Evaluación para la Toma de Decisiones (DEMATEL) y el Modelado Estructural Interpretativo (ISM) utilizadas en el estudio (Ni Guodong et al. 2022).

Hay que mencionar que la OIL en su informe 2023 estima que existió un déficit global de empleo que afectará a alrededor de 453 millones de personas, representando un 11,7% de la población. Esta cifra es más del doble del índice de desempleo registrado.

El indicador de la OIL sobre el déficit de empleo, que engloba a todas las personas deseadas de trabajar, pero sin empleo, ilustra la verdadera dimensión de los desafíos en el ámbito laboral. Es importante destacar que este déficit es significativamente mayor entre las mujeres, alcanzando un 14,5%, en contraste con el 9,8% observado en los hombres (International Labour Organizational 2023).

Metodología

Tipo de investigación

El presente estudio es de carácter aplicativo, el cual es una técnica de investigación cuantitativa en la que nos permite comparar la eficacia de modelos de regresión logística y arboles de clasificación, con un diseño no experimental de tipo transversal pues se centra en el análisis de datos históricos de variables sociodemográficas, observadas durante el 2023, registradas por el Instituto Nacional de Estadísticas y Censos (INEC).

Población del estudio

Conformada por los datos históricos considerando a las Personas Económicamente Activas según el INEC, de acuerdo con variables sociodemográficas (Área, Estado Civil, Edad, Sexo, Nivel de Instrucción, Provincia, Ocupación y Situación Laboral o Ocupacional), del año 2023.

Identificación de Variables

Variable Dependiente

- Situación Laboral u Ocupacional

Variables Independientes

- Área.
- Estado civil.
- Edad
- Sexo
- Nivel de Instrucción
- Provincia
- Ocupación

Método

Se emplearon dos modelos estadísticos con sus indicadores de eficacia como son la matriz de confusión y Curva ROC e Índice AUC para identificar que modelo es el más eficaz al momento de identificar las variables sociodemográficas influyentes respecto a la variable respuesta del estudio, se ha analizado el comportamiento de dichas variables con la finalidad de identificar patrones predominantes de acuerdo a las categorías que se presentasen de ser cualitativos y estadísticas descriptivas en cuanto a la edad.

Resultados

Análisis general de los datos

Con la finalidad de identificar concentraciones representativas, patrones que permitan identificar de manera detallada la distribución de cada variable acorde a la categoría se presenta la siguiente tabla:

Tabla 1: Análisis univariado variables independientes cualitativas

Variable	Categoría	Porcentaje (%)
Sexo	Hombre	55.7
	Mujer	44.3
Área	Urbana	71.6
	Rural	28.4
Nivel de instrucción	Educación media/bachillerato	35.1
	Educación básica	34.0
	Superior	28.6
	Ninguno	2.1
	Centro de alfabetización	0.2
Estado civil	Casado/a	34.1
	Soltero/a	32.2
	Unión libre	19.2
	Separado/a	7.0
	Divorciado/a	4.4
Ocupación	Viudo/a	3.2
	Trabajadores no calificados, ocupaciones elementales	21.2
	Trabajad. de los servicios y comerciantes	21.0
	Trabajad. calificados agropecuarios y pesqueros	21.0
	Oficiales operarios y artesanos	15.3
	Profesionales científicos e intelectuales	0.2
	Operadores de instalaciones. máquinas y montad.	10.8
	Técnicos y profesionales de nivel medio	9.9
Sin actividad	6.3	
Empleados de oficina	6.0	

	Personal direct./admin. pública y empresas	4.7
	Fuerzas Armadas	3.3
Provincia	Pichincha	13.8
	Guayas	11.6
	El Oro	10.6
	Azuay	9.2
	Tunguragua	8.9
	Esmeraldas	5.2
	Imbabura	4.8
	Loja	3.6
	Santa Elena	3.4
	Sucumbíos	2.4
	Orellana	2.4
	Cotopaxi	2.2
	Pastaza	2.2
	Carchi	2.1
	Chimborazo	2.1
	Zamora Chinchipe	2.0
	Napo	2.0
	Manabí	2.0
	Santo Domingo	1.9
	Cañar	1.9
Los Rios	1.6	
Bolivar	1.6	
Morona Santiago	1.4	
Galapagos	1.1	

Elaborado por: Autores, 2025.

Al observar la Tabla 1 se presentan las características predominantes de los individuos en estudio, donde se puede indicar que el sexo predominante es el masculino representado con el 55.7% respecto al sexo femenino que presenta el 43.3%, por lo que se concluye que los hombres son mayoría en el estudio.

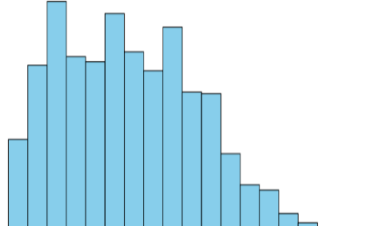
El área juega un rol importante en el estudio se presenta con el 71.6% ciudadanos ecuatorianos viviendo en área urbana mientras que el 28.4% pertenecen al área rural es decir que los ciudadanos tienen mayor oportunidad laboral en las grandes ciudades que en sus residencias rurales puede ser resultado a que el mayor movimiento económico está en las principales ciudades del Ecuador.

De acuerdo con el nivel de instrucción, la educación media o bachillerato es la que predomina en los y las ecuatorianos con el 35.4%, seguida de educación básica con el 34% y por último superior con el 28.6% siendo estas las más predominantes en cuanto al nivel de instrucción que tienen los ecuatorianos. Por otra parte, los ecuatorianos que no tienen ningún nivel de instrucción representan el 2.1% y los que tienen algún reconocimiento por el centro de alfabetización es de 0.2%.

De acuerdo con la ocupación que presentan los ecuatorianos el 21.2% son trabajadores no calificados, ocupaciones elementales, es decir realizan tareas sencillas y rutinarias que requieren principalmente la utilización de herramientas manuales y, a menudo cierto esfuerzo físico. El 21.0% representa a los trabajadores en servicios y comerciantes, el 15.3% representa los trabajadores calificados agropecuarios y pesqueros quienes son un fuerte en Ecuador.

Los ecuatorianos con empleo y desempleo se presentan gran mayoría en la provincia de Pichincha capital de la república del Ecuador con 13.8%, Guayas 11.6% provincia costera con un alto impacto económico, el Oro 10.6% provincia, considerada como la "capital bananera del mundo" quien dinamiza la economía del país con sus importaciones de productos.

Tabla 1: Estadísticas descriptivas de la variable Edad

Variable		Estadísticas		
Edad	Persona económicamente activa que inicia desde 15 años hasta los 98 años, donde al menos realice una actividad.	Media	42.22	
		Mediana	41	
		Desviación Estándar	15.54	
		Curtosis	0.36	
		Asimetría	2.40	
		Coficiente Variación	0.37	

En cuanto a la variable edad el promedio de los ecuatorianos que están dentro de las personas que tienen una actividad remunerada es de 42 años, se puede indicar que es una edad adulta mayor la que predomina en cuanto al estudio tomando en cuenta que la edad mediana es de 41 años. Se presenta una dispersión de la edad respecto a la media del 15.54 lo cual es moderada. La edad como tal no está concentrada en ciertos años pues está distribuida similar a una curva normal tal como lo refleja la curtosis de 0.36. La variabilidad de los datos es moderada pues representa el 37.5 de la edad promedio.

Prueba de independencia

En el estudio se realizó la prueba de independencia mediante la prueba de Chi Cuadrado (χ^2) donde el índice del p-value será usado para probar o no probar

dependencia de la variable dependiente. Mediante el siguiente proceso para finalmente concluir en una decisión estadística.

- Planteamiento de Hipótesis

H_0 : No existe relación entre las variables independientes y la dependiente (Situación Laboral).

H_1 : Existe relación entre las variables independientes y la dependiente (Situación Laboral).

- Nivel de significancia

En el estudio para todos los casos, se usa un nivel de significancia $\alpha = 0.05$.

- Región de rechazo

La decisión se toma comparando el p-valor con α . Si el p-valor es menor que α , rechazamos

Tabla 2: Prueba de independencia mediante Chi Cuadrado

Variables	p-valor		Nivel significancia	Independiente	Decisión
Sexo	0.001	<	$\alpha = 0.05$	Situación Laboral	Dependiente
Nivel de instrucción	0.001	<			Dependiente
Estado Civil	0.000	<			Dependiente
Provincia	0.000	<			Dependiente
Área	0.001	<			Dependiente
Ocupación	0.000	<			Dependiente

De acuerdo con el estudio de independencia existe suficiente evidencia estadística para concluir que las variables (sexo, nivel de instrucción, estado civil, provincia, área, ocupación) presentan relación significativa con respecto a la situación laboral, es decir son dependientes.

Modelo de regresión logística

Tabla 3: Índice de significancia de acuerdo con el modelo

	Estimación	Error Estándar	Valor de Z	p-valor	Significancia
Intercepto	2.090	0.111	18.814	2e-116	***
Sexo	0.329	0.023	14.167	2e-116	***

Área	0.759	0.032	23.193	2e-116	***
Nivel de instrucción	-0.188	0.015	-11.953	2e-116	***
Estado civil	-0.165	0.007	-23.368	2e-116	***
Provincia	-0.000741	0.001	-0.403	0.687	
Edad	0.0345	0.001	32.382	2e-116	***

Se observa en la Tabla 3 el modelo de regresión logística donde la significancia va de acuerdo con el índice de significancia, el sexo, área, nivel de instrucción, estado civil y edad aportan significativamente a la variable independiente como es la situación laboral. Por otro lado, la variable provincias no influye dentro de la situación laboral.

Indicadores de eficacia

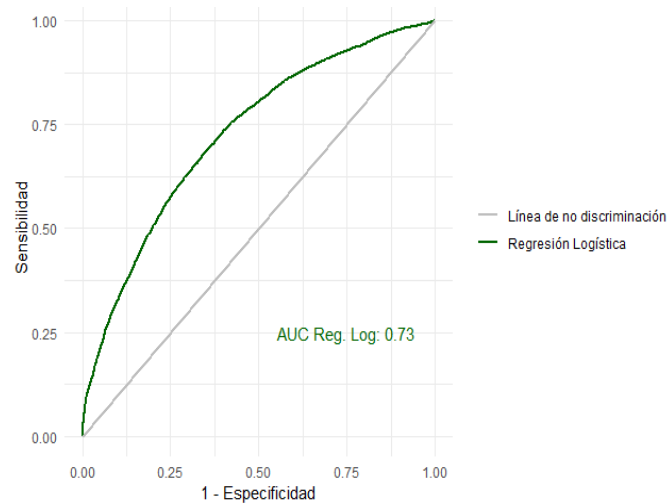
$$\text{Error de predicción} = \frac{8166 + 0}{174096} * 100 = 4.7\%$$

$$\text{Exactitud} = \frac{165930 + 0}{165930 + 8166 + 0 + 0} * 100 = 95.3\%$$

Según la matriz de confusión se presenta una exactitud del 95.3% por lo que se concluye que las predicciones mediante el modelo de regresión logística son correctas, se tiene un error de predicción del 4.7 %, en cuanto a la predicción de la situación laboral que presenten los ecuatorianos el cual recae dentro de un rango aceptable.

Curva Roc- Análisis de Regresión

Representación gráfica mediante la curva ROC del modelo de regresión logística se presenta en la Figura 1;

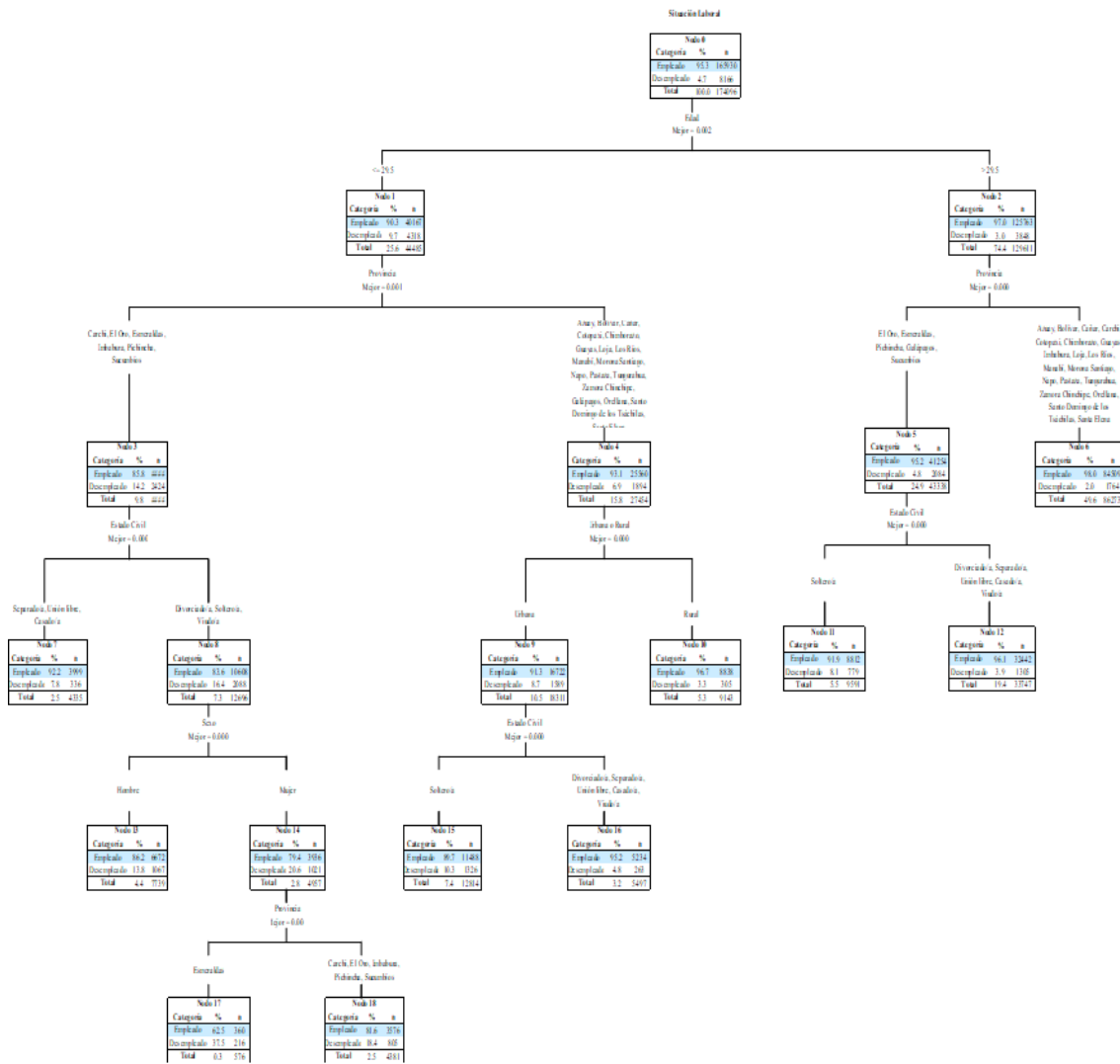
Figura 1: Curva Área ROC de la regresión logística

De acuerdo con la figura 1, el modelo de regresión logística utilizado tiene una capacidad decente para predecir la situación laboral basada en los factores considerados. Esta curva ROC sugiere que el modelo tiene un rendimiento decente para distinguir entre las clases positivas y negativas, pero no es perfecto. El AUC representa el 0.73 el cual está dentro de la escala regular entre los intervalos [0.6 - 0.75] que es medianamente buena, pero no perfecta.

Modelo de aboles de clasificación

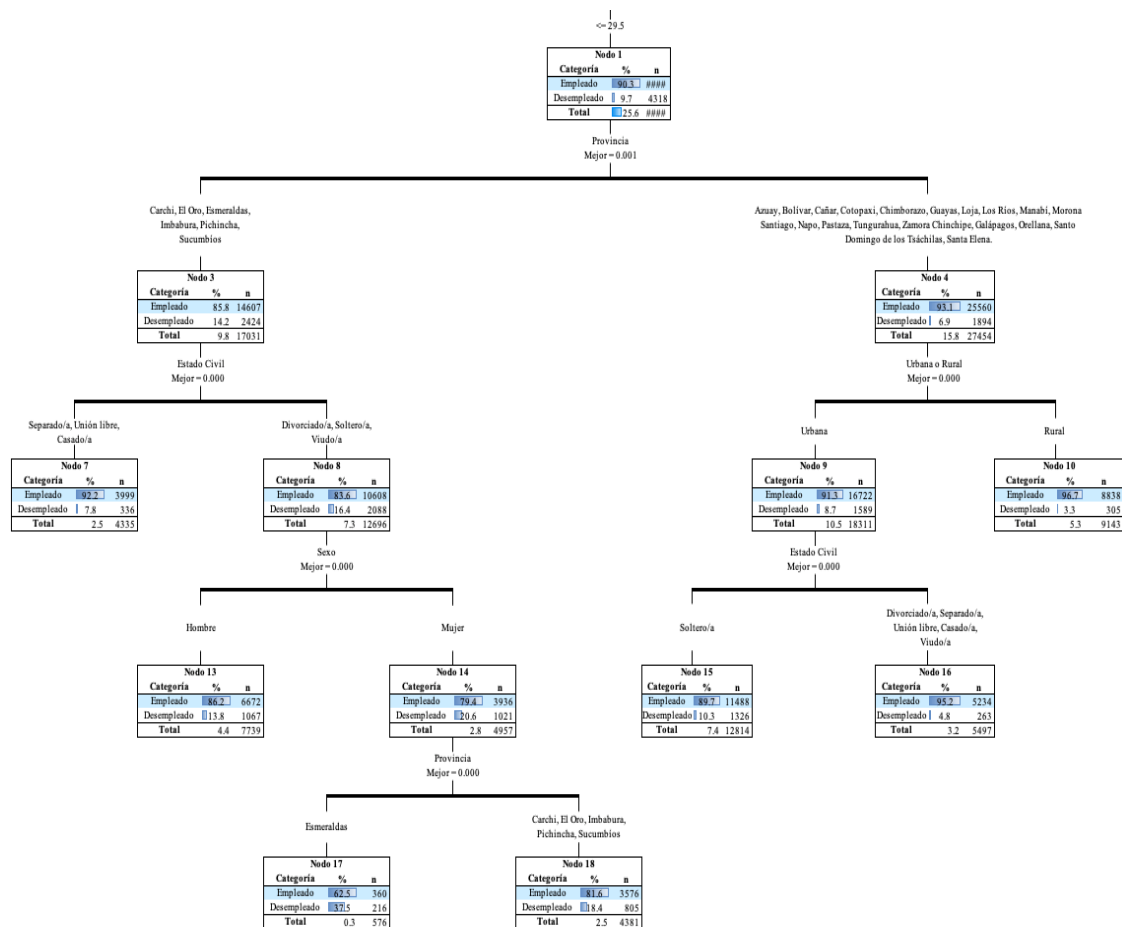
De acuerdo con el estudio la matriz de confusión en cuanto a arboles de decisión permite determinar el estado desde que se encuentra desempleado a empleado despreciando el estado empleado y el estado de desempleado. De acuerdo con el árbol existen 18 nodos, de un total de 174096 observaciones utilizadas, 165930 (95.3%) pertenecen a personas empleadas y 8166 (4.7%) a desempleadas. Interpretación del árbol de clasificación desglosada de acuerdo con dos ramificaciones como se lo muestra a continuación:

Figura 2: Árboles de clasificación utilizando el método CRT



Ramificación de acuerdo con la edad menor o igual a 29 años y 6 meses edad.

Figura 3: Árboles de clasificación parte I



Interpretación

En el nodo 0, describe la variable dependiente como es la situación laboral se aprecia que el 95.3% presentarán un empleo y el 4.3% presentará desempleo, en donde la variable importante es la edad donde se ramifica en dos nodos importantes, menor o igual a 29.5 años que representa el 25.6% y otro nodo mayor o igual a 29.5 años representado con el 74.4%, pertenecientes a la misma, indicando que esta es la variable principal predictora de la situación laboral.

El Nodo 1, si los ecuatorianos menores o de igual edad a 29.5 años se clasifican un 90.3% conseguir un empleo, y un 9.7% se clasificarán como desempleados. Es decir, existe una disminución en proporción de desempleo en esta edad debido a que por lo general el nivel de instrucción de bachillerato y superior se culminan en la etapa joven.

En cuanto al nodo 3, se tomará en cuenta de acuerdo con la provincia en donde se toman en cuenta el Carchi, El Oro, Esmeraldas, Imbabura, Pichincha, Sucumbíos se presenta una proporción de 85.8% conseguir un empleo, y un 14.2% se clasificarán como desempleados, en este caso existe un aumento considerable de desempleados en estas provincias.

El nodo 7, se representa de acuerdo con las seis provincias antes mencionadas, sin embargo, se toma en cuenta el estado civil con las categorías; separado/a, Unión libre y Casado/a lo que indica que una persona que presenta responsabilidades económicas como sentimentales es más probable presente empleo con un 92.2% frente a un 7.8% que este desempleado siendo este un nodo terminal.

En cuanto al nodo 8, sigue el estado dentro del nodo 3 pero exceptuando en cuanto a las categorías de la variable estado civil donde ahora se toman en consideración divorciado/a, soltero/a y viudo/a, por lo que se concluye que una persona al estar sin obligaciones o vínculos se reflejará un menor índice de empleo con el 83.6% frente al 16.4 % de desempleados el cual ya es alarmante este indicador.

El nodo 13 y 14 se vuelve a replicar el nodo 8 pero ahora se explicará de acuerdo con la variable sexo, en cuando al nodo 13 el hombre presentará una participación de empleo con el 83.2% hombres con empleo frente al 13.8 hombres sin empleo siendo este nodo terminal. En cuanto al nodo 14 la participación de mujeres con el empleo es del 79.4% respecto al 20.6% mujeres desempleadas.

En el nodo 17 y 18 se toma en cuenta el sexo, pero específicamente las mujeres de acuerdo con las provincias, en donde de acuerdo con las mujeres que residan en esmeraldas presentaran un empleo 62.55 frente a un 37.5% de mujeres desempleadas. Mientras que en las provincias del: Carchi, El Oro, Imbabura, Pichincha, Sucumbíos el 81.6% de mujeres presenta un empleo frente al 18.4% que no tendrán un empleo, siendo estos dos nodos terminales.

El nodo 4, hace referencia a las provincias, pero son en mayor proporción, 18 provincias como son: Azuay, Bolívar, Cañar, Cotopaxi, Chimborazo, Guayas, Loja, Los Ríos, Manabí, Morona Santiago, Napo, Pastaza, Tungurahua, Zamora Chinchipe, Galápagos, Orellana, Santo Domingo de los Tsáchilas, Santa Elena las cuales al ser

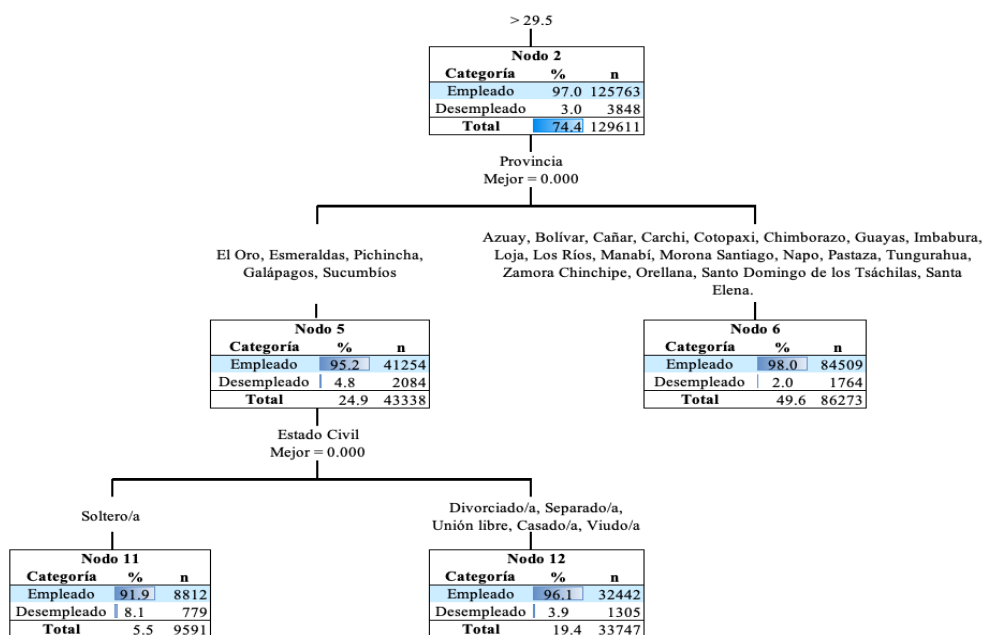
mayoría y ser unas de las principales ciudades en aportes económicos presentaran un empleo 93.1% con un desempleo 6.9%.

El nodo 9 y 10, se deriva de acuerdo las provincias en mayor proporción, pero de acuerdo con la variable área, en cuanto al zona urbana con el nodo 9 el empleo se presentará con el 91.3% frente a 8.7 que se presenta como desempleados. Mientras el nodo 10 representa la zona rural presenta un empleo del 96.7% frente a un 3.3 este último puede ser resultado que las mejores oportunidades se para generar ingresos y fuentes de empleo se encuentren fuera de esta zona por ende se vea reflejado salir en búsqueda de un empleo, cabe mencionar este nodo es terminal.

El nodo 15 y 16 se replica con el nodo 4 de acuerdo con el área con la categoría urbana explicando con la variable estado civil, la categoría soltera/o presenta un empleo 89.7% frente al 10.3% de desempleados solteros. Mientras que los Divorciado/a, Separado/a, Unión libre, Casado/a y Viudo/a se presentan con empleo 95.2% de ciudadanos respecto al 4.8% de desempleados siendo estos nodos terminales.

Ramificación de acuerdo con la edad mayor a 29 años y 6 meses edad.

Figura 3: Árboles de clasificación parte II



Interpretación

El Nodo 2, de acuerdo con la segunda ramificación importante se toma en cuenta la edad de los ecuatorianos mayores o de igual edad a 29.5 años representada por el 74.4%, clasificando un 97.0% a conseguir un empleo, y un 3.0 % se clasificarán como desempleados. Es decir, existe una disminución mínima en proporción de desempleo en esta edad.

El cuanto al nodo 5 y 6 se ramifica en dos grupos de acuerdo con la provincia las cuales; El Oro, Esmeraldas, Pichincha, Galápagos y Sucumbíos presentaran empleo con el 95.2% frente al 4.8% con desempleo. Por otra en el nodo 6, las provincias Azuay, Bolívar, Cañar, Carchi, Cotopaxi, Chimborazo, Guayas, Imbabura, Loja, Los Ríos, Manabí, Morona Santiago, Napo, Pastaza, Tungurahua, Zamora Chinchipe, Orellana, Santo Domingo de los Tsáchilas y Santa Elena, el empleo se presenta con el 98% frente al 2% siendo este no muy significativo en cuanto a la situación laboral, dando como resultado nodo terminal.

Finalmente, la ramificación que da como resultado al nodo 11 y 12 es ligada de acuerdo con el estado civil. El nodo 11 toma la categoría soltero/a, en donde las personas solteras que presenten empleo serán del 91.9% frente a las desempleadas con el 8.1%. mientras que para el nodo 12 se toma las categorías Divorciado/a, Separado/a, Unión libre, Casado/a y Viudo/a pero la cual estas personas presentan una mayor proporción en cuanto a empleados con el 96.1% frente 3.9% de desempleados, siendo estos nodos terminales.

Indicadores de eficacia

$$\begin{aligned} \text{Error de predicción} &= \frac{8166 + 0}{174096} * 100 = 4.7\% \\ \text{Exactitud} &= \frac{165930 + 0}{165930 + 8166 + 0 + 0} * 100 = 95.3\% \end{aligned}$$

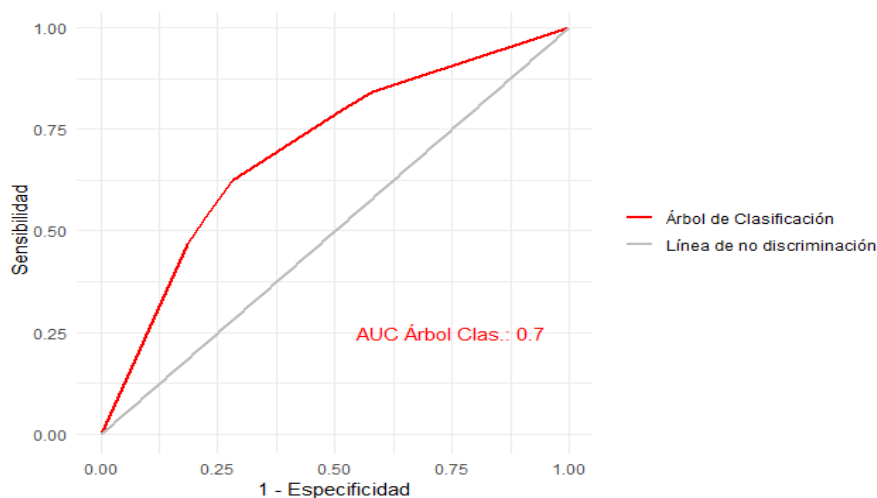
Según la matriz de confusión se presenta una exactitud del 95.3% por lo que se concluye que las predicciones mediante el modelo de árboles de clasificación son correctas, se tiene un error de predicción del 4.7 %, en cuanto a la predicción de la

situación laboral que presenten los ecuatorianos el cual recae dentro de un rango aceptable de la misma forma que el modelo anterior.

Curva Roc-Árboles de clasificación

Representación gráfica mediante la curva ROC del modelo de árboles clasificación se presenta en la Figura 2;

Figura 4: Curva Área ROC de árboles de clasificación



De acuerdo con la Figura, esta curva el modelo de árbol de clasificación tiene un rendimiento regular para predecir la situación laboral basada en los factores considerados. Es decir, la curva ROC indica que el modelo es útil y ofrece una capacidad de discriminación algo razonable, aunque no perfecta. El área bajo la curva (AUC) representa el 0.70 el cual está dentro de la escala [0.6 - 0.75] que está dentro de la escala regular, en cuanto a presentar los factores influyentes.

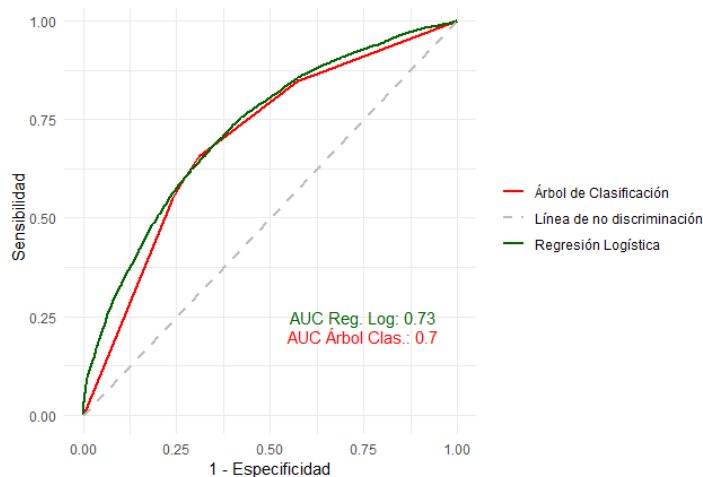
Evaluación y comparación

De acuerdo con el estudio mediante la comparación de los modelos: Regresión Logística y Árboles de Clasificación usando las técnicas como es la matriz de confusión y curva ROC con su AUC, el modelo de árbol de clasificación está dentro de la escala regular para predecir la situación laboral basada en los factores considerados.

En cuanto al modelo de regresión logística pese a presentar similares matrices de confusión con el modelo de árboles de clasificación se presenta una diferencia en

cuanto a la curva ROC y el AUC estando dentro de una escala regular pero superior respecto a la otra, por lo cual se toma en consideración que es el mejor modelo que explica los factores considerados que influyen en la situación laboral.

Figura 5: Representación conjunta de curvas de cada modelo



De acuerdo con la Figura 6, se evidencia que la curva ROC que mejor exactitud presenta es el de regresión logística, donde existe una curva evidentemente no perfecta pero esta técnica permite identificar factores que influyan en cuanto a la variable predictora como es la situación laboral, mencionar que es un modelo intuitivo y sencillo de realizar para la toma de decisiones. Dejando por detrás al modelo de árboles de clasificación con una diferencia con 0.3 puntos en la escala de clasificación AUC, pero ubicándole una escala menor.

Una vez analizado las ilustraciones ROC se procede a clasificar de acuerdo tasa de error y la exactitud de cada modelo comparando el área bajo la curva (AUC) con los resultados siguientes:

Tabla 4: Clasificadores

	Exactitud	Error de predicción	AUC
Regresión Logística	95.3%	4.7%	0.73 = 73%
Arboles de clasificación	95.3%	4.7%	0.70 = 70%

Como se observa en la Tabla 4, el modelo que presenta la regresión logística es la que mejor ayuda a predecir la variable predictora en el estudio como es la situación laboral de las/os ecuatorianos basado en los datos presentado durante el 2023.

Hay que mencionar que el modelo de regresión logística la variable provincias no se consideró como factor influyente en la situación laboral mientras que en arboles de clasificación si se consideró y ocupaciones se utilizó como factor de información más no como factor influyente en los modelos. Factores que influyeron de acuerdo con el modelo presentado en cada modelo;

Tabla 5:Comparativa entre factores resultante a cada modelo

Factor considerado en regresión logística	No considerado	Factor considerado en arboles de clasificación
Sexo ✓	Ocupación	Sexo ✓
Área ✓	Provincia	Área ✓
Nivel de instrucción ✓		Nivel de instrucción ✓
Estado civil ✓		Estado civil ✓
Edad ✓		Edad ✓
		Provincia ✓

Las variables con el símbolo de visto se consideran como factores tanto para regresión logística y arboles de clasificación, por otra parte, la ocupación se considera como factor informativo más no se consideró como un posible factor influyente dentro de los modelos.

Discusión

general se demostró que resultados parecidos en cuanto al modelo de regresión y arboles de clasificación exceptuando la curva la curva ROC y su AUC, en donde la regresión logística fue el modelo más efectivo para predecir la variable predictora que relaciona la situación laboral que presentan cada ecuatoriano. El modelo de regresión logística es estricto en cuanto a la significancia que pueda presentar cada variable en estudio sin embargo es de fácil comprensión e intuitivo para llegar a la toma de decisiones. El modelo de árboles de clasificación presenta varios métodos de

crecimiento para identificar nodos significativos como son el CHAID, CHAID EXHAUSTIVO, QUEST y CRT, este último siendo el considerado para el estudio debido a su uso en común para árboles de clasificación o regresión, clasificando por tratamiento en grupos o segmentos para conseguir que los datos sean más homogéneos respecto a la variable dependiente. Los árboles de clasificación son comúnmente usados para identificar interacciones no lineales las cuales son algo complejas en la interpretación y en el tiempo de respuesta a al momento de resolver un problema y llegar a una correcta toma de decisiones. Si bien la regresión logística mostró ser más precisa en este escenario particular, es fundamental evaluar las fortalezas y limitaciones de cada método en relación con las necesidades y características específicas del problema de predicción pues es más intuitivo de interpreta el modelo de árboles de clasificación con sus limitantes, pero es muy útil a la hora de toma de decisiones.

Conclusiones

En el marco de análisis exploratorio y graficas conjuntas de acuerdo con la situación laboral se observó ciertas características o rasgos predominantes en grupos de categorías de acuerdo con el sexo, nivel de instrucción, estado civil, provincias, áreas, edad todas estas juegan un rol importante en cada ecuatoriano/a pues se depende mucho de las ciudades importantes la cuales mueven la economía del país y la educación a la que tengamos acceso para presentar mejores oportunidades laborales. Al modelar las tendencias representativas en cuanto a los factores influyentes con relación a la situación laboral en el Ecuador utilizando el modelo de regresión logística se presenta que la provincia no era significativa, mientras en árboles de clasificación si se considera dentro de su ramificación.

La capacidad predictiva de los modelos de regresión logística e árboles de clasificación presenta un valor de 95.3% exactitud y error de predicción del 4.7% fue igual para los dos sin embargo existe una diferencia en la curva ROC y el área bajo la curva (AUC) con el 73% de eficacia para la regresión logística frente al 70% de árboles de clasificación para predecir situación laboral en el Ecuador.

Es importante destacar la importancia relativa de cada variable predictor en los modelos de regresión logística y árboles de clasificación. Las variables representativas

son; nivel de instrucción la educación básica es la que tiene un índice mayor de empleo con un 33.2%, y el bachillerato presenta un 33.0%; el área representada con el 71.6% pertenecientes a la zona urbana y el 28.4% rural existiendo una brecha considerable; y el estado civil es de gran importancia considéralo con factor influyente debido a que son de gran importancia en el estudio donde las personas casadas representan el 34.1%, solteras 32.2% y unión libre 19.2%.

Por último, se considera una edad promedio de 42 años, este indicador evidencia que las personas económicamente activas quienes aportan al desarrollo económico del país están dentro de la categoría de la adultez más no en la categoría juventud plena, lo que es un limitante a la hora de conseguir un empleo.

Referencias Bibliográficas

- Arellano Estrada, P., & Ayaviri Nina, D. (2021). *Los determinantes del desempleo en el Ecuador. Revista Perspectivas*, (48). Recuperado el 8 de abril de 2024, de http://www.scielo.org.bo/scielo.php?script=sci_abstract&pid=S1994-37332021000200009&lng=es&nrm=iso&tlng=es
- Austin, P. C. (2007). A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting AMI mortality. *Statistics in Medicine*, 26(15). <https://doi.org/10.1002/sim.2770>
- Congacha Ortega, G. N. (2020). *Comparación de modelos logísticos y árboles de decisión para identificar y predecir factores asociados a la desnutrición crónica infantil basados en la Encuesta Nacional de Salud y Nutrición (ENSANUT) 2018-2019* [Tesis de pregrado, Escuela Superior Politécnica de Chimborazo]. Recuperado el 31 de enero de 2024, de <http://dspace.esPOCH.edu.ec/handle/123456789/14551>
- Flores, P., & Salto, V. (2018). Comparativa entre classification trees, random forest y gradient boosting en la predicción de la satisfacción laboral en Ecuador. *Ciencia Digital*, 2(4.1). <https://doi.org/10.33262/cienciadigital.v2i4.1..189>
- Instituto Nacional de Estadística y Censos (INEC). (2023). *ENEMDU histórico empleo 2023*. Recuperado el 8 de abril de 2024, de <https://www.ecuadorencifras.gob.ec/enemdu-historico-empleo-2023/>

- International Labour Organization (OIT). (2023). *ILO monitor on the world of work* (11th ed.). Suiza: Nota informativa. Recuperado el 6 de enero de 2024, de http://www.ilo.org/global/publications/books/WCMS_883344/lang--es/index.htm
- Lee, C., & Park, S. (2021). Changing factors of employee satisfaction with working conditions: An analysis of the Korean Working Conditions Survey. *Safety and Health at Work*, 12(3). <https://doi.org/10.1016/j.shaw.2021.04.003>
- Ni, G., Li, G., Jin, T., Hu, H., & Zhang, Z. (2022). Analysis of factors influencing the job satisfaction of new generation of construction workers in China: A study based on DEMATEL and ISM. *Buildings*, 12(5), 609. Recuperado el 8 de abril de 2024, de <https://www.mdpi.com/2075-5309/12/5/609>
- Pérez Rave, J., & González Echavarría, F. (2018). Classification trees vs. logistic regression in the generic skill development in engineering. *Computación y Sistemas*, 22(4). <https://doi.org/10.13053/cys-22-4-2804>