



Doi: <https://doi.org/10.70577/asce.v5i2.827>

Recibido: 2026-04-11

Aceptado: 2026-04-25

Publicado: 2026-05-13

Análisis de factores asociados al rendimiento académico en estudiantes de educación básica en Ecuador: un estudio comparativo mediante árboles de clasificación (CHAID)

Analysis of factors associated with academic performance in basic education students in Ecuador: a comparative study using classification trees (CHAID)

Autores

Josselyn A. Caiza Fuel¹

Maestría en Estadística Aplicada

josselyna.caiza@upec.edu.ec

<https://orcid.org/0009-0004-7567-7128>

Universidad Politécnica Estatal del Carchi

Tulcán – Ecuador

Luis F. Basantes Moreano²

luisfbasantes@gmail.com

<https://orcid.org/0000-0001-7336-7549>

Universidad Politécnica Estatal del Carchi

Tulcán – Ecuador

Como Citar

Caiza Fuel, J. A. &, Basantes Moreano, L. F. (2026) Análisis de factores asociados al rendimiento académico en estudiantes de educación básica en Ecuador: un estudio comparativo mediante árboles de clasificación (CHAID). 1329-1351



Resumen

La presente investigación es de tipo cuantitativo, transversal y asociativo-explicativo, se trabaja con datos obtenidos de la encuesta “Ser Estudiante” (INEVAL) aplicada a nivel nacional. El propósito es analizar los factores asociados al rendimiento académico de estudiantes de tercero de bachillerato en Ecuador, mediante la comparación de los periodos de pandemia y postpandemia. Utilizando árboles de clasificación CHAID, se estimaron modelos independientes para cada periodo y se validó su desempeño mediante matriz de confusión, exactitud global, exactitud balanceada y coeficiente Kappa ponderado, además se complementó con validación cruzada. Los principales resultados arrojan que la interacción entre el docente y el estudiante y el tipo de financiamiento institucional se relacionan de manera significativa con el rendimiento académico. En la pandemia se encontró relación entre el área de asentamiento de la institución y el uso del computador con el rendimiento; mientras que, en postpandemia, la situación laboral del estudiante y la autonomía de estudio son factores relevantes. Los modelos muestran una exactitud global superior al 84% y Kappa ponderado que muestra concordancia sustancial; con lo cual se puede concluir estabilidad predictiva.

Palabras clave: Rendimiento académico; Árboles de Clasificación; CHAID; Pandemia; Postpandemia; Evaluación Educativa.



Abstract

This research follows a quantitative, cross-sectional, and associative-explanatory design, utilizing national data from the “Ser Estudiante” survey (INEVAL) in Ecuador. The study aims to analyze the factors associated with the academic performance of twelfth-grade students (tercero de bachillerato), comparing the pandemic and post-pandemic periods. Using CHAID (Chi-square Automatic Interaction Detector) classification trees, independent models were estimated for each period. Model performance was validated using a confusion matrix, global accuracy, balanced accuracy, and weighted Kappa coefficient, further complemented by cross-validation. The main results indicate that teacher-student interaction and the type of institutional funding are significantly related to academic performance. During the pandemic, the school’s geographical location and computer usage were found to be key factors; whereas, in the post-pandemic period, the student’s employment status and study autonomy emerged as relevant variables. The models demonstrate a global accuracy exceeding 84% and a substantial agreement according to the weighted Kappa, concluding predictive stability across the analyzed periods.

Keywords: Academic performance; Multivariate methods; Classification tree; Pandemic; Post-pandemic; High school.



Introducción

La educación es un importante agente de cambio y promotora de desarrollo social y económico de los países, ya que influye directamente en la formación de capital humano y la reducción de desigualdades estructurales (Díaz & Alemán, 2008). En el marco de los Objetivos de Desarrollo Sostenible, el Objetivo 4 plantea la necesidad de garantizar una educación de calidad, inclusiva y equitativa promoviendo oportunidades de aprendizaje a lo largo de la vida (Naciones Unidas, s.f.). En este sentido, el estudio del rendimiento académico se convierte en una guía importante al evaluar la garantía de los sistemas educativos y con base en la evidencia permite direccionar políticas públicas.

El rendimiento académico es considerado un constructo susceptible de adoptar valores cuantitativos y cualitativos (Navarro, 2003), al ser multicausal, envuelve una gran capacidad explicativa de los distintos factores y espacios temporales que intervienen en el proceso de aprendizaje, entre ellos intervienen componentes tanto internos como externos al estudiante y estos factores son de orden social, cognitivo y emocional, que se clasifican en tres categorías: personales, sociales e institucionales (Pazmiño et al., 2023). El rendimiento académico es una parte fundamental en el proceso de enseñanza y aprendizaje porque nos permite identificar si el estudiante cumple con los estándares de aprendizaje que dispone el currículo de educación para ser promovido de nivel (Estrada, 2018).

Diversas investigaciones empíricas han demostrado que diferentes variables como el género, el aprendizaje autorregulado y la formación de docentes tienen asociaciones significativas con el rendimiento académico (Larruzea & Ramírez, 2020). Asimismo, las limitaciones económicas y la necesidad de trabajar representan barreras significativas en el rendimiento académico (Guaman et al., 2024) La pandemia ocasionada por la COVID-19 representó un cambio drástico sin precedentes en los sistemas educativos a nivel mundial; en más de 190 países a mediados de mayo de 2020, más de 1.200 millones de estudiantes de todos los niveles de enseñanza, habían dejado de tener clases presenciales en la escuela. De ellos, más de 160 millones eran estudiantes de América Latina y el Caribe, en la mayoría de países se establecieron formas de continuidad de los estudios en diversas modalidades a distancia (Naciones Unidas y CEPAL, 2020) Sin embargo, las medidas adoptadas generaron brechas asociadas al acceso a la tecnología, competencias tecnológicas y



condiciones de estudio en el hogar, las cuales se profundizaron en países en vías de desarrollo incrementando la desigualdad existente.

La modalidad a distancia o virtual fue predominante, con el objetivo de disminuir la propagación de la enfermedad y continuar con el aprendizaje, en consecuencia, la educación migro de un entorno presencial a una enseñanza virtual, que involucra una planificación detallada, el uso de recursos tecnológicos e innovadores y una preparación adecuada de docentes y estudiantes (Montaño-Blacio et al., 2023). En el caso ecuatoriano, la pandemia resaltó las desigualdades preexistentes, especialmente en la falta de acceso a tecnología y conectividad que afecta especialmente a los estudiantes en áreas rurales y sectores económicamente desfavorecidos (Guapulema Ocampo et al., 2024).

Una vez frenada la propagación del virus, las instituciones educativas regresaron a la presencialidad, la cual es conocida como educación tradicional porque es dictada por el docente a sus estudiantes dentro de un espacio físico llamado aula, sin embargo, esta modalidad no ha sido estudiada con mayor profundidad, sobre todo en la época postpandemia (Cobos Reina & Ledesma Silva, 2024). Además, el retorno a la presencialidad planteó ciertos desafíos en los estudiantes con retos educativos relacionados al soporte socioemocional, el aprendizaje, la enseñanza, el apoyo familiar y la capacitación, considerando una educación que debe ser renovada, accesible y que beneficie a todo el estudiantado (Correa et al., 2023).

En cuanto a la metodología, los métodos multivariantes son una herramienta de gran ayuda cuando el objetivo es analizar y caracterizar factores que se interrelacionan entre sí, además permiten reducir y clasificar datos, así como también crear constructos o variables compuestas (Basantes & Useche, 2022). En particular, el algoritmo CHAID se fundamenta en la prueba de independencia de Chi-cuadrado para detectar interacciones significativas entre la variable dependiente y los predictores; constituyendo a los árboles de clasificación, como una técnica estadística que sirve para modelar la relación entre una variable dependiente con más variables predictoras, es un método no paramétrico que puede manejar variables numéricas como categóricas y puede ser utilizado para clasificación y predicción (Carrillo et al., 2018). El análisis de árbol de clasificación



es fácil de interpretar y proporciona una explicación de la relación entre las variables predictoras y la variable de respuesta (Rodríguez-Garcés et al., 2021).

Al contexto de la investigación, el objetivo busca analizar los factores asociados al rendimiento académico de estudiantes de educación básica en Ecuador, comparando los periodos de pandemia y postpandemia, mediante el uso de árboles de clasificación CHAID, utilizando datos de la encuesta “Ser Estudiante” aplicada a nivel nacional por el INEVAL. La comparación entre los periodos que se encuentran diferenciados por un contexto sanitario, permite explorar cómo las condiciones de estudio (virtualidad y presencialidad) se relacionan con el desempeño académico.

Este enfoque no intenta establecer relaciones causales, sino identificar asociaciones estadísticamente significativas que aporten información para comprender los niveles de segmentación del rendimiento académico bajo distintos escenarios de aprendizaje. Los resultados aportan evidencia empírica relevante con el fin de diseñar estrategias orientadas a fortalecer la interacción entre docentes y estudiantes, mejorar el tiempo de estudio autónomo, reducir diferencias institucionales y optimizar el uso recursos tecnológicos a nivel educativo, especialmente en contextos de vulnerabilidad socioeconómica.

Material y métodos

Diseño de la investigación

El presente estudio se enmarca en un enfoque cuantitativo, con diseño no experimental, transversal y de alcance explicativo–asociativo. La comparación descriptiva comprende dos periodos académicos diferenciados por el contexto sanitario global ocasionado por la COVID-19: el periodo de pandemia (2020-2021) que se desarrolló bajo la modalidad virtual, y el periodo de postpandemia (2023-2024) desarrollado bajo la modalidad presencial.

El objetivo metodológico es analizar los factores asociados al rendimiento académico de estudiantes de tercero de bachillerato en Ecuador mediante la aplicación de árboles de clasificación CHAID, diferenciando los resultados entre ambos contextos.



La unidad de análisis corresponde a los estudiantes evaluados de tercero de bachillerato de las Instituciones de Educación Secundaria del Ecuador; la información utilizada proviene de las bases de datos de la encuesta “Ser Estudiante” del Instituto Nacional de Evaluación Educativa (Ineval), la cual recoge información demográfica, socioeconómica, académica y emocional de los estudiantes evaluados a nivel nacional. El tamaño muestral inicial fue de 3.893 observaciones para el periodo de pandemia y 12.112 para el periodo postpandemia. La mayor representatividad del periodo postpandemia se debe a que se tuvo mayor cobertura de las instituciones educativas a nivel nacional tras el levantamiento de las restricciones de movilidad, lo cual en el periodo de pandemia se vio frenado, para evitar el contagio de la enfermedad.

Tratamiento de datos

Al realizar un análisis exploratorio de los valores faltantes, se identificó el porcentaje de ausencia por variable y su distribución respecto a la variable dependiente. En el periodo de pandemia se eliminaron 914 registros con datos faltantes en la variable dependiente y 9 observaciones sin información en variables explicativas. Adicionalmente, se excluyó la variable correspondiente al quintil económico debido a que presentaba un 24,2% de datos faltantes, proporción que podría afectar la estabilidad del modelo. En el periodo postpandemia se eliminaron 778 registros con valores perdidos en la variable dependiente y 167 en variables explicativas.

Con el fin de trabajar con una base de datos completamente depurada y de esta manera asegurar la integridad del análisis, se aplicó el método de eliminación por lista (Listwise Deletion), el cual es usado cuando se tiene un gran conjunto de datos y se puede eliminar la fila o columna donde se encuentra la data perdida, para obtener una base, aunque más pequeña que esta sea completa (Useche & Mesa, 2006). A pesar que esta estrategia puede reducir eficiencia estadística y potencialmente introducir sesgo si la ausencia no es completamente aleatoria, el análisis exploratorio no muestra patrones sistemáticos claros asociados a una categoría específica de la variable dependiente. Se consideraron técnicas de imputación; sin embargo, dado que el algoritmo CHAID se fundamenta en particiones jerárquicas basadas en frecuencias observadas, la imputación de categorías podría alterar la estructura del árbol. Por lo que, el tamaño muestral final se consideró



suficiente para garantizar robustez analítica (n=2.970 en pandemia y n=11.167 en postpandemia). Esta decisión se reconoce como una limitación metodológica.

Variables del modelo

Variable dependiente

La variable de respuesta en ambos periodos es el promedio global, que hace referencia a la nota global obtenida por el estudiante, la cual originalmente es cuantitativa. Para cumplir con los requerimientos del algoritmo CHAID, se procedió a su categorización en niveles ordinales de logro conforme la clasificación de la ficha metodológica de la encuesta:

- Insuficiente (400–599)
- Elemental (600–699)
- Satisfactorio (700–799)
- Excelente (800–1000)

Si bien esta categorización permite adecuar la variable al modelo a usar, esto implica pérdida de información respecto a la escala continua original, en consecuencia, los resultados se interpretan en términos de asociaciones entre categorías de logro y no como variaciones marginales del puntaje numérico.

Variables independientes

Se incluyeron variables demográficas (sexo, autoidentificación étnica), socioeconómicas (tipo de financiamiento institucional, participación laboral), académicas (retroalimentación docente, frecuencia de estudio independiente), tecnológicas (uso del computador, acceso a internet) y emocionales (satisfacción familiar, dificultades para dormir, entre otras).

Se realizó la selección inicial de variables en función a criterios teóricos y disponibilidad en la base de datos, la selección definitiva de variables predictoras se efectuó por el algoritmo CHAID mediante pruebas de significancia estadística; finalmente, no se añadieron componentes sintéticos en el modelo final.

Materiales y métodos

Árbol de clasificación CHAID

Se aplicó el algoritmo CHAID (Chi-square Automatic Interaction Detector), el cual es un método no paramétrico que permite ver la asociación o dependencia entre la variable rendimiento académico con el resto de variables explicativas mediante pruebas de independencia Chi-cuadrado, considerando únicamente a las variables que resultaron estadísticamente significativas. En cada paso, CHAID elige la variable independiente (predictora) que presenta la interacción más fuerte con la variable dependiente. Las categorías de cada predictor se funden si no son significativamente distintas respecto a la variable dependiente (IBM Corp., 2017).

Los parámetros de construcción fueron:

- Nivel de significancia para división de nodos: $\alpha = 0,05$
- Nivel de significancia para fusión de categorías: $\alpha = 0,05$
- Corrección de Bonferroni para comparaciones múltiples
- Profundidad máxima del árbol: 3 niveles
- Casos mínimos en nodo padre: 100
- Casos mínimos en nodo hijo: 50

Estos criterios fueron establecidos con el fin de controlar el crecimiento excesivo del árbol y prevenir sobreajuste.

Se estimó un modelo independiente para cada periodo: pandemia y postpandemia.

Evaluación del modelo y control de sobreajuste

Dado que la variable dependiente es categórica ordinal, la evaluación del desempeño se realizó mediante métricas propias de clasificación.



Se construyó la matriz de confusión para cada modelo y se calcularon:

- Exactitud global (accuracy)
- Exactitud balanceada
- Coeficiente Kappa ponderado
- Sensibilidad por clase

Debido a la desproporción entre categorías, principalmente en el periodo postpandemia, la exactitud balanceada y el Kappa ponderado fueron considerados indicadores centrales de desempeño.

Como complemento para evaluar la estabilidad fuera de muestra, se consideró validación cruzada de 10 particiones, este procedimiento permitió obtener una medida más robusta del rendimiento predictivo y reducir el riesgo de sobreajuste.

Criterios de comparabilidad entre periodos

La comparación entre pandemia y postpandemia se realiza bajo un enfoque descriptivo–asociativo. Dado que existen diferencias en tamaño muestral, cobertura institucional y contexto de aplicación de la encuesta, es importante indicar que esto puede influir en la distribución de los niveles de logro observados. En consecuencia, los resultados comparativos deben interpretarse como patrones observacionales y no como estimaciones causales del efecto de la modalidad de estudio sobre el rendimiento académico.

Resultados

Análisis descriptivo del rendimiento académico

La tabla 1 muestra las categorías de la variable dependiente y su estadística descriptiva en los periodos de pandemia y postpandemia.

Durante la pandemia se observa una distribución en la muestra relativamente equilibrada entre las categorías Elemental (50,5%) y Satisfactorio (48,6%) y proporciones marginales en las categorías extremas Insuficiente (0,6%) y Excelente (0,3%); por otro lado, en el periodo postpandemia existe

información únicamente en las categorías Elemental y Satisfactoria, teniendo mayor concentración la categoría Elemental (64,5%) y una reducción relativa en la categoría Satisfactorio (35,5%).

Se debe tener precaución al interpretar estos resultados, debido a que los periodos muestran diferencias en el tamaño muestral, cobertura institucional y contexto de aplicación de la encuesta. Por lo que, la comparación es de tipo descriptivo y observacional, sin implicación de inferencias causales que dependan de la modalidad de estudio.

Tabla 1.

Promedio global durante los periodos de pandemia y postpandemia

Periodo	Categorías	Frecuencia absoluta	Frecuencia Relativa
Postpandemia	Elemental	7.198	64,5
	Satisfactorio	3.969	35,5
Pandemia	Insuficiente	17	0,6
	Elemental	1500	50,5
	Satisfactorio	1443	48,6
	Excelente	10	0,3

Elaboración propia

Resultados del árbol de clasificación CHAID – Periodo Pandemia

La técnica de árbol de clasificación CHAID determina que la variable principal para obtener un promedio satisfactorio es el número de días que utiliza el computador. Esta variable es considerada el nodo raíz del árbol, indicando una mayor asociación estadística con la variable dependiente.

En los otros niveles del árbol surgen los siguientes factores relevantes:

- Tipo de financiamiento institucional
- Retroalimentación docente de las evaluaciones



En este sentido, los estudiantes con mayor probabilidad de obtener notas satisfactorias son aquellos que utilizan el computador con fines académicos 6 o más días a la semana, cursan sus estudios en instituciones privadas y reciben retroalimentación referente a sus evaluaciones por parte del docente.

Por el contrario, la mayor probabilidad de obtener notas de categoría Elemental se asocia con:

- Área de asentamiento de la institución a la que asisten

Los estudiantes con mayor probabilidad de obtener una nota elemental son aquellos que utilizan el computador un día a la semana y asisten a instituciones en zonas rurales.

Evaluación del modelo – Periodo Pandemia

Matriz de confusión

La matriz de confusión (tabla 2) evalúa la capacidad del modelo para clasificar de manera correcta las categorías observadas del promedio global, arrojando una exactitud de 90,3%. Referente a la validación cruzada de 10 particiones se obtuvo que 9,7% de observaciones de clasifican de forma errónea, con un error estándar de 0,09 lo que evidencia estabilidad predictiva y bajo riesgo de sobreajuste.

Tabla 2.

Matriz de confusión – Periodo Pandemia

Observado/Pronosticado	Insuficiente	Elemental	Satisfactorio	Excelente	Total
Insuficiente	7	6	4	0	41,2%
Elemental	5	1.374	121	0	91,6%
Satisfactorio	0	138	1.301	4	90,2%



Excelente	0	0	10	0	0,0%
Total	0,0%	51,1%	48,4%	0,0%	90,3%

Elaboración propia.

De acuerdo con la tabla 3, el modelo clasifica correctamente aproximadamente 90% de los casos. El coeficiente Kappa ponderado sugiere un nivel de concordancia sustancial entre clasificación observada y predicha, ajustando por acuerdo esperado al azar. La sensibilidad por clase muestra que el modelo tiene una buena capacidad de detectar una clase en particular.

Tabla 3.

Métricas de clasificación para el periodo Pandemia

Métrica	Valor
Exactitud global	0,903
Exactitud balanceada	0,557
Kappa ponderado	0,800
Sensibilidad – Elemental	0,916
Sensibilidad – Satisfactorio	0,902

Elaboración propia

Resultados del árbol de clasificación CHAID – Periodo Postpandemia

En el escenario de postpandemia, el árbol de clasificación revela que, el nodo raíz es el financiamiento de la institución educativa. Los estudiantes que asisten a instituciones particulares tienen mayor probabilidad de obtener notas Satisfactorias.



En los próximos niveles del árbol surgen los siguientes factores relevantes:

- Realizar tareas o repasar materias en casa durante 6 días a la semana
- Condición laboral del estudiante

Al respecto, los estudiantes con mayor probabilidad de tener nivel satisfactorio son quienes estudian en instituciones particulares, no realizar actividades laborales y realizan tareas o repasan la materia por 6 días a la semana en casa.

Mientras que, el nivel elemental se asocia con los siguientes factores:

- Corrección de tareas en clase
- Retroalimentación del docente

Los estudiantes que tienen mayor probabilidad de obtener notas entre 600 y 699 puntos, es decir que obtienen un promedio Elemental, son aquellos que asisten a instituciones fiscales, no reciben o reciben poca retroalimentación por parte del docente.

Evaluación del modelo – Periodo Postpandemia

Matriz de confusión

La matriz de confusión (tabla 4) muestra una exactitud de 84,7%, indicando que este es el porcentaje de observaciones correctamente clasificadas. Sobre la validación cruzada de 10 particiones se obtuvo que 15,3% de observaciones de clasifican de forma errónea, con un error estándar de 0,04 lo que evidencia estabilidad predictiva y bajo riesgo de sobreajuste.

Tabla 4.

Matriz de confusión – Periodo Postpandemia

Observado/Pronosticado	Elemental	Satisfactorio	Total
Elemental	5.950	1.248	82,7%



Satisfactorio	466	3.503	88,3%
---------------	-----	-------	-------

Total	62,2%	37,8%	84,7%
-------	-------	-------	-------

Elaboración propia.

De acuerdo con la tabla 5, la exactitud balanceada del modelo confirma un desempeño adecuado pese al desbalance entre categorías. El Kappa ponderado evidencia concordancia sustancial, es decir gran parte de observaciones bien clasificadas son reales y no fruto del azar. Similar al periodo de pandemia, la sensibilidad por clase muestra que el modelo tiene una buena capacidad de detectar una clase en específico.

Tabla 5.

Métricas de clasificación para el periodo Postpandemia

Métrica	Valor
Exactitud global	0,847

Exactitud balanceada	0,855
----------------------	-------

Kappa ponderado	0,679
-----------------	-------

Sensibilidad – Elemental	0,827
--------------------------	-------

Sensibilidad – Satisfactorio	0,883
------------------------------	-------

Elaboración propia

Comparación estructural entre periodos

En ambos periodos se evidencia la importancia de la interacción docente y el financiamiento institucional. El periodo pandemia revela que el uso del computador y el asentamiento de la



institución educativa son factores que influyen en el rendimiento académico, mientras que, en el periodo de postpandemia, existen otros factores relevantes relacionados a la condición laboral del estudiante y al estudio autónomo sistemático.

Debe reiterarse que estas diferencias reflejan patrones observacionales dentro de contextos diferenciados por un contexto sanitario y no deben interpretarse como efectos causales directos de la modalidad educativa adoptada en cada periodo.

Discusión

Los resultados de la presente investigación permiten identificar asociaciones consistentes entre variables académicas, institucionales y contextuales de los estudiantes de educación básica en Ecuador con su rendimiento académico, considerando que existe una diferenciación entre los periodos de pandemia y postpandemia. Desde un enfoque metodológico robusto, los hallazgos son relaciones asociativas y no causales, debido al diseño transversal y el uso de modelos de clasificación.

Factores estructurales estables en ambos periodos

La relevancia de la interacción docente–estudiante, particularmente la retroalimentación sobre tareas y evaluaciones es uno de los resultados más importantes. En este sentido, el hallazgo sugiere que, independientemente de la modalidad educativa (virtual o presencial), la calidad de la participación del docente mantiene una asociación significativa con el desempeño académico del estudiante. Este resultado es coherente con la literatura sobre eficacia escolar, donde la retroalimentación formativa destaca como un mecanismo central de mejora del aprendizaje, ya que es considerada una herramienta pedagógica fundamental para potenciar el rendimiento académico, al ofrecer a los estudiantes información concreta, oportuna y procesable sobre su desempeño (Moya-Muñoz et al., 2025). Es así que los resultados indican que en la pandemia una retroalimentación adecuada por parte del docente incrementa la probabilidad de obtener notas satisfactorias; por otro lado, en postpandemia la poca o nula retroalimentación de tareas influye directamente a obtener notas de nivel elemental.



De igual manera, el tipo de financiamiento institucional se relaciona a niveles superiores de logro, lo cual debe interpretarse con precaución, debido a que el modelo no controla de manera explícita otras variables socioeconómicas omitidas. Sin embargo, este resultado podría indicar diferencias estructurales en recursos educativos, infraestructura tecnológica y capital cultural. En el periodo de postpandemia, esta variable constituye el nodo raíz del árbol CHAID, lo cual evidencia su mayor poder discriminante respecto al nivel de logro académico. Aunque con menor jerarquía estructural, este resultado coincide en el periodo pandemia.

Cambios estructurales entre pandemia y postpandemia

Durante el periodo de pandemia surge como factor relevante el número de días de uso del computador, mientras que, en el periodo posterior adquiere mayor relevancia el tipo de financiamiento de la institución educativa. El análisis del periodo postpandemia revela que la condición laboral tiene una fuerte relación con el rendimiento académico, indicando que los estudiantes que no trabajan tienen mayor probabilidad de obtener promedios satisfactorios. Este cambio puede darse dada la transición a la presencialidad, en la cual la carga laboral limita la organización del tiempo, que se vera reflejado con mayor intensidad en el rendimiento académico. Además, se coincide con Barragán Velasco et al. (2025), quienes concluyen que el trabajo fuera del horario escolar limita el tiempo de estudio, afecta además a la concentración y reduce la participación en actividades académicas y sociales del estudiante.

Un aporte significativo surge del análisis del periodo de pandemia, donde el nivel Elemental está determinado por el asentamiento de la institución. Este resultado sugiere que el asistir a instituciones ubicadas en zonas rurales aumenta la probabilidad de obtener promedios elementales, indicando que el contexto sanitario profundizo la diferencia entre los estudiantes de zonas urbanas y rurales.

Por otro lado, el estudio autónomo sistemático (repasar materias seis días por semana) se considera un factor relevante en el periodo postpandemia, lo cual puede reflejar un proceso de adaptación académica a la presencialidad, sin embargo, es importante tomar en consideración, que estudiar de forma más estratégica de lo habitual, implica menor tiempo de aprendizaje, por lo que, el uso eficaz del tiempo también puede reducir la probabilidad de notas bajas cuando se estudia mucho tiempo



sin concentración, lo cual subraya la clara necesidad de enseñar a los estudiantes estrategias de estudio eficaces (Theobald, 2025). En este sentido resulta relevante expandir el estudio hacia el tiempo que se destina al estudio y las estrategias de aprendizaje.

Evaluación del desempeño predictivo

El modelo CHAID mostró niveles elevados de exactitud global 0,903 (pandemia) y 0,847 (postpandemia) y los coeficientes Kappa ponderados indican una concordancia sustancial entre clasificación observada y predicha en ambos periodos. La validación cruzada confirmó estabilidad fuera de muestra, reduciendo la probabilidad de sobreajuste. Estos cálculos demuestran que el modelo elegido destaca ya que tiene gran capacidad para capturar interacciones jerárquicas entre variables y relaciones no lineales.

Al agregar métricas propias de clasificación derivadas de la matriz de confusión de cada periodo, se pudo evaluar la metodología de manera coherente tomando en consideración la naturaleza ordinal de la variable dependiente.

No obstante, cabe destacar que la concentración de casos en las categorías Elemental y Satisfactorio influye en la elevada exactitud observada, especialmente en el periodo postpandemia, lo que podría limitar la sensibilidad para detectar diferencias en las categorías extremas.

Limitaciones metodológicas

Es necesario reconocer ciertas limitaciones que presenta el estudio:

1. **Categorización de la variable dependiente:** La transformación del promedio, la cual originalmente es continua, a una variable ordinal implica pérdida de información y reducción de variabilidad, sin embargo, este cambio fue necesario para aplicar el algoritmo CHAID.
2. **Tratamiento de datos perdidos:** La eliminación por lista completa pudo haber reducido eficiencia estadística y potencialmente introducir sesgo si los datos no eran completamente aleatorios.
3. **Comparabilidad entre periodos:** Las diferencias en tamaño muestral, cobertura institucional y contexto de aplicación de la encuesta limitan la interpretación longitudinal estricta.



4. **Diseño transversal:** No permite establecer relaciones causales ni efectos dinámicos derivados del contexto de pandemia.

Implicaciones sustantivas

Para obtener promedios más altos, los resultados de la presente investigación sugieren que las políticas educativas deben estar orientadas a fortalecer la retroalimentación pedagógica, promover el estudio autónomo, fortalecer las instituciones de financiamiento fiscal y con asentamiento rural; y reducir las barreras laborales para estudiantes en contextos vulnerables.

Además, la estabilidad estructural de factores como el financiamiento institucional y la retroalimentación por parte del docente en ambos periodos indica que, más allá de la coyuntura sanitaria, existen determinantes que se encuentran asociados al rendimiento académico de manera persistente, lo cuales deben ser abordados desde una perspectiva sistémica.

Síntesis interpretativa

En conjunto, los resultados evidencian que el rendimiento académico responde a una combinación de factores pedagógicos, institucionales y contextuales cuya relevancia relativa varía según el entorno educativo. El uso de árboles de clasificación CHAID permitió identificar estructuras jerárquicas de asociación.

Por tanto, el estudio aporta evidencia empírica robusta sobre los factores asociados al rendimiento académico en Ecuador, diferenciando dos contextos educativos recientes y metodológicamente contrastados.

Conclusiones

Este artículo busca identificar los factores asociados al rendimiento académico de los estudiantes de educación básica a nivel nacional durante los periodos de pandemia y postpandemia, mediante el uso de árboles de clasificación CHAID. A partir de los resultados se extraen conclusiones sustantivas y metodológicas relevantes.



Primero, se identificó que obtener un mejor promedio global, tanto en periodos de pandemia como de postpandemia, está estrechamente ligado a la interacción entre el docente con el estudiante. Este resultado sugiere que la retroalimentación por parte del docente, conservan su importancia independientemente de la modalidad educativa.

Segundo, el tipo de financiamiento muestra una asociación persistente con el rendimiento académico en ambos periodos, Durante postpandemia el financiamiento institucional presento mayor poder discriminante, mientras que en pandemia tuvo una asociación estadísticamente significativa, pero con menor jerarquía. Aunque el modelo no permite establecer relaciones causales, los resultados muestran la existencia de brechas estructurales que pueden estar ligadas a recursos, capital cultural y apoyo académico.

En tercer lugar, se observa una reconfiguración de factores asociados a cada periodo. En la pandemia los días de uso del computador y el asentamiento de la institución emergen como variables significativas, indicando que en este periodo fue determinante el acceso a un computador para adaptarse a la virtualidad y la diferencia entre instituciones de zonas urbanas y rurales, asistir a instituciones en zonas rurales se relaciona significativamente con obtener promedios de nivel elemental. Mientras que, en postpandemia el estudio autónomo y no realizar actividades laborales incrementan la probabilidad de ubicarse en niveles superiores de rendimiento, esto sugiere que la disponibilidad de tiempo y la carga laboral son factores relevantes en el contexto de retorno a la presencialidad.

Desde la parte metodológica, el uso del algoritmo CHAID facilitó la identificación de estructuras jerárquicas de asociación y permitió la caracterización de perfiles de estudiantiles con distinto nivel académico. La evaluación del desempeño mediante métricas propias de clasificación (exactitud global, exactitud balanceada y Kappa ponderado), junto con la validación cruzada, establece estabilidad predictiva y bajo riesgo de sobreajuste.

No obstante, el estudio muestra limitaciones que deben ser consideradas. La categorización de la variable dependiente en niveles ordinales implicó pérdida de información respecto a la escala continua original. Asimismo, el método de Listwise Deletion en el tratamiento de datos perdidos pudo reducir eficiencia estadística e introducir sesgo si las observaciones faltantes no fueron completamente aleatorias. Finalmente, las diferencias en tamaño muestral, cobertura institucional



y contexto de aplicación de la encuesta limitan la comparabilidad estrictamente longitudinal entre periodos.

De acuerdo a las implicaciones, los resultados sugieren que las políticas de índole educativo se deben orientar al fortalecimiento de la interacción entre el docente y el estudiante, principalmente en la retroalimentación de tareas y evaluaciones; además se debe promover el estudio autónomo y para los estudiantes en vulnerabilidad laboral resulta importante reducir las barreras laborales con el propósito de mejorar el desempeño académico. Además, el hecho de que ciertos factores persistan en ambos periodos indica la necesidad de intervenciones permanentes más allá de contingencias sanitarias específicas.

Como líneas futuras de investigación, se recomienda: estimar modelos con el promedio global en escala continua para evitar pérdida de información, incorporar técnicas de imputación múltiple para evaluar la sensibilidad frente a datos perdidos y desarrollar estudios que posibiliten establecer relaciones causales.

En síntesis, el estudio contribuye con evidencia empírica relevante sobre los factores asociados al rendimiento académico en educación básica en el Ecuador, integrando técnicas de segmentación no paramétrica bajo un marco metodológico consistente. Estos resultados contribuyen tanto al debate académico sobre determinantes del logro educativo como al diseño de políticas públicas basadas en evidencia.

Referencias bibliográficas

- Barragán Velasco, G. B., Casquete Velásquez, K. C., Guingla Quincha, G. R., & Irma Melania, A. T. (2025). Trabajo infantil y rendimiento académico en estudiantes de básica superior: Caso de la Unidad Educativa Corina Parral, Bolívar, Ecuador. *Ciencia Latina Revista Científica Multidisciplinar*, 9(3), 5022–5058. https://doi.org/10.37811/cl_rcm.v9i3.18129
- Basantes, L., & Useche, L. (2022). *Revisión de métodos estadísticos multivariantes aplicados en los resultados del examen ser bachiller en el Ecuador*.



- Carrillo, S., Sanabria, B., Bermudez, V., & Espinosa, J. (2018). *Actores en la educación: una mirada desde la psicología educativa*.
- Cobos Reina, R., & Ledesma Silva, Y. (2024). Study modality and academic performance in higher education. *European Public and Social Innovation Review*, 9. <https://doi.org/10.31637/epsir-2024-1487>
- Correa-Alzate, J. I., Suárez-Vallejo, J. P., & Restrepo-Restrepo, N. (2023). Teachers' Experience in Times of Pandemic with Students with Educational Challenges: Challenges in Returning to Face-To-Face Education with an Inclusive Approach in Medellín, Colombia. *Revista Electronica Educare*, 27(2). <https://doi.org/10.15359/ree.27-2.15856>
- Díaz, T., & Alemán, P. (2008). *La educación como factor de desarrollo Education as a Development Factor*.
- Estrada, A. (2018). Estilos de aprendizaje y rendimiento académico. *Revista Boletín Redipe*, 7(7), 218–228.
- Guaman Gualan, J. A., Lujé Caisa, K. A., Caiza Ortiz, D. E., & Varguillas Carmona, C. S. (2024). Los Factores Sociales Relacionados con el Rendimiento Académico: Un Análisis en la Educación Superior. *Reincisol.*, 3(6), 7069–7089. [https://doi.org/10.59282/reincisol.v3\(6\)7069-7089](https://doi.org/10.59282/reincisol.v3(6)7069-7089)
- Guapulema Ocampo, K. J., Alvarado Guapulema, P. A., Proaño del Castillo, M. G., & Peñaloza Camacho, K. I. (2024). La brecha digital en la educación ecuatoriana: Desafíos post pandemia. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, 5(5). <https://doi.org/10.56712/latam.v5i5.2907>
- IBM Corp. (2017). *IBM SPSS-Árboles de decisiones 25*.
- Larruzea-Urkixo, N., & Ramírez, O. C. (2020). Individual differences in self-regulated learning among education degree students: Gender, teaching specialty, grades and academic performance. *Revista de Investigacion Educativa*, 38(2), 453–473. <https://doi.org/10.6018/RIE.334301>
- Montaño-Blacio, M., González-Escarabay, J., Pesántez-Bravo, F., Briceño-Sarmiento, J., & Jimenez-Sarango, O. (2023). Machine learning applied to the analysis of the COVID-19 confinement and its relationship with the performance of higher education students; [Aprendizaje automático aplicado al análisis del confinamiento por COVID-19 y su relación con el rendimiento de los estudiantes de educación superior]. *RISTI - Revista Iberica de*



- Sistemas e Tecnologias de Informacao*, 2023(E52), 102 – 116.
<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85184687483&partnerID=40&md5=b1192982578ca754cc31b9abc0178fdb>
- Moya-Muñoz, L. D., Lopez-Velez, C. R., Pérez-Marquin, J. M., & Cedeño-León, M. M. (2025). La retroalimentación formativa como estrategia para mejorar el desempeño en el aula. *Revista Científica Zambos*, 4(2), 155–170. <https://doi.org/10.69484/rcz/v4/n2/118>
- Naciones Unidas, & CEPAL. (2020). *La educación en tiempos de la pandemia de COVID-19*. https://www.siteal.iiep.unesco.org/respuestas_educativas_covid_19.
- Navarro, R. E. (2003). *EL RENDIMIENTO ACADÉMICO: CONCEPTO, INVESTIGACIÓN Y DESARROLLO* (Vol. 1, Number 2). <http://www.ice.deusto.es/rinace/reice/vol1n2/Edel.pdf>
- Pazmiño, L., Rodríguez, C., Barrón, J., & Cuevas, H. (2023). Factores predominantes que influyen en el indicador de rendimiento académico en los universitarios in situ. *Acta Universitaria*. <http://doi.org/10.15174/au.2023.3878>
- Rodríguez-Garcés, C. R., Fuentes, M. G., & Fuentes, G. P. (2021). Autoestima en Niños, Niñas y Adolescentes chilenos: análisis con árboles de clasificación. *Revista Reflexiones*, 100(1). <https://doi.org/10.15517/rr.v100i1.43342>
- Theobald, M. (2025). Study longer or study effectively? Better study strategies can compensate for less study time and predict goal achievement and lower negative affect. *British Journal of Educational Psychology*, 95(2), 405–420. <https://doi.org/10.1111/bjep.12725>
- Useche, L., & Mesa, D. (2006). Una introducción a la imputación de valores perdidos UNA INTRODUCCIÓN A LA IMPUTACIÓN DE VALORES PERDIDOS AN INTRODUCTION TO THE IMPUTATION OF LOST VALUES. In *Terra: XXII* (Number 31).

Agradecimiento:

A mi esposo Erick Arequipa e hijo Emmanuel Arequipa y a mis amados padres Janneth y Edgar.

Nota:

El artículo no es producto de una publicación anterior.