**TELOS: Revista de Estudios Interdisciplinarios en Ciencias Sociales**
ISSN 1317-0570 / ISSN (e) 2343-5763
Vol. 27(1). Enero-Abril 2025

# ESTUDIO DEL RENDIMIENTO ACADÉMICO MEDIANTE LA COMPARACIÓN DE MODELOS DE REGRESIÓN Y ÁRBOLES DE CLASIFICACIÓN

Johanna Enith Aguilar-Reyes*
https://orcid.org/0000-0002-1230-2503
Edwin Fernando Mejía-Peñafiel**
https://orcid.org/0000-0001-6888-4621
Tania Paulina Morocho-Barrionuevo***
https://orcid.org/0000-0002-1019-6049
Geoconda-Marisela Velasco Castelo****
https://orcid.org/0000-0001-8120-1549

**RECIBIDO**: 15/10/2024 / **ACEPTADO**: 17/12/2024 / **PUBLICADO**: 15/01/2025

**Cómo citar:** Aguilar-Reyes, J., Mejía-Peñafiel, E., Morocho-Barrionuevo, T., Velasco Castelo, G. (2025). Estudio del rendimiento académico mediante la comparación de modelos de regresión y árboles de clasificación. *Telos: Revista de Estudios Interdisciplinarios en Ciencias Sociales, 27*(1), 94-115. www.doi.org/10.36390/telos271.08

## RESUMEN

Este artículo tiene como objetivo identificar los factores que afectan el rendimiento académico a partir de la comparación de los modelos de regresión y árboles de decisión para determinar los factores que involucra. La metodología adoptada es de naturaleza cuantitativa utilizando modelo de regresión y árboles de clasificación y la población estudiada incluye a los estudiantes de nivelación en la carrera de estadística, a quienes se les realizó un análisis exploratorio y descriptivo, empleando dos métodos estadísticos. Se emplearon dos técnicas de modelado: la regresión logística multinomial y los árboles de clasificación. Las variables evaluadas incluyeron factores sociodemográficos, rendimiento académico previo y características del entorno educativo. Los resultados mostraron que el modelo de regresión logística alcanzó una precisión del 100% con un área bajo la curva de 1, lo que indica una capacidad perfecta de clasificación. En comparación, el modelo de árbol de clasificación tuvo una precisión del 70.83% con un área bajo la curva de 0.7042, lo que refleja una capacidad moderada de clasificación. A partir de estos resultados, se identificaron factores clave que afectan el rendimiento académico, como los

* *Autor de correspondencia.* Ingeniera en Estadística Informática. Máster Universitario en Estadística Aplicada. Docente en la Escuela Superior Politécnica de Chimborazo – Ecuador. johannae.aguilar@espoch.edu.ec
** Ingeniero en Sistemas. Magíster en Informática Aplicada por la Escuela Superior Politécnica de Chimborazo – Ecuador. Especialista en Inteligencia Artificial. Docente en la Escuela Superior Politécnica de Chimborazo – Ecuador. efmejia@espoch.edu.ec
*** Biofísica. Máster Universitario en Estadística Aplicada. Docente en la Escuela Superior Politécnica de Chimborazo – Ecuador. tpaulina.morochob@espoch.edu.ec
**** Ingeniera Mecánica. Magíster en Física Aplicada, mención Física Computacional. Magíster en Gestión Industrial y Sistemas Productivos. Docente en la Escuela Superior Politécnica de Chimborazo – Ecuador. geoconda.velasco@espoch.edu.ec

hábitos de estudio, el interés en la carrera y aspectos psicológicos. En conclusión, la regresión logística multinomial fue más eficaz y precisa para analizar las relaciones cuantitativas entre las variables que afectan el rendimiento académico, superando al método de árboles de clasificación.

**Palabras clave**: Rendimiento Académico, Regresión Logística, Árboles De Clasificación, Matriz De Confusión

*Study of academic performance through comparison of regression models and classification trees*

**ABSTRACT**

This article aims to identify the factors that affect academic performance by comparing regression models and decision trees to determine the factors involved. The methodology adopted is quantitative in nature, focused on the collection of numerical data and its statistical analysis, in order to evaluate the relationships between different variables and determine those factors that influence academic performance. The population studied includes remedial students in the statistics career, who underwent an exploratory and descriptive analysis, using two statistical methods. Two modeling techniques were used: multinomial logistic regression and classification trees. The variables evaluated included sociodemographic factors, previous academic performance, and characteristics of the educational environment. The results showed that the logistic regression model achieved 100% accuracy with an AUC of 1, indicating perfect classification ability. In comparison, the classification tree model had an accuracy of 70.83% with an AUC of 0.7042, reflecting moderate classification ability. From these results, key factors that affect academic performance were identified, such as study habits, interest in the career and psychological aspects. In conclusion, multinomial logistic regression was more effective and accurate in analyzing the quantitative relationships between the variables that affect academic performance, outperforming the classification tree method.

**Keywords**: Academic Performance, Logistic Regression, Classification Trees, Confusion Matrix.

**Introducción**

El uso de análisis predictivos y modelización estadística puede considerarse un procedimiento muy útil cuando se trata de interpretar y anticipar comportamientos de datos en diversos escenarios (Álvarez et al., 2024). En ese sentido, es una valiosa herramienta en el análisis de aquellos factores que intervienen, positiva o negativamente, en el contexto que rodea el rendimiento académico de los estudiantes universitarios, bien sea en Ecuador u otros países latinoamericanos. Al respecto, la comparación de modelos de regresión y árboles de clasificación puede proporcionar una visión profunda sobre los factores que influyen en el desempeño académico en las universidades.

De acuerdo con Lizares (2017), en la actualidad existe una preocupación intelectual, en cuanto a la necesidad de implementar la estadística aplicada en el estudio del rendimiento académico, ya sea en los niveles medios o universitarios. Se trata de una iniciativa que procura dotar a las instituciones educativas de las herramientas adecuadas en la predicción de ese

rendimiento y, la toma de las previsiones más adecuadas a las características de sus alumnos. Pérez y González (2018) consideran que este tipo de esfuerzos académicas son pertinentes, no solo como una forma de mejorar los procesos educativos, sino de cara a la pertinencia de la formación universitaria para el trabajo que realizarán los egresados.

En ese orden de ideas, es fundamental comprender que las herramientas que provee la estadística permiten establecer la relación entre diferentes variables valorando el significado de cada una de ellas, en la presentación de una determinada realidad (Norambuena et al., 2022). Por ejemplo, en el contexto del rendimiento académico, un modelo de regresión puede ayudar a identificar o determinar aquellas variables que tienen un impacto significativo en las calificaciones de los estudiantes, como el tiempo dedicado al estudio, la asistencia a clases, el nivel socioeconómico, el apoyo familiar, entre otros (Garbanzo, 2014).

Ahora bien, en el contexto de este trabajo, hay que tener presente que hasta ahora se han implementado diversas técnicas de clasificación basadas en árboles de decisión, entre las cuales los árboles de clasificación han ganado relevancia. Sin embargo, los estudios que analizan su desempeño en comparación con otros métodos, como la regresión logística, son escasos. La regresión logística es comúnmente utilizada como una opción frente al análisis discriminante, especialmente cuando los requisitos de normalidad y homocedasticidad no se cumplen, lo que dificulta el uso de este último (Mures et al., 2005).

Aun así, los estudios comparativos entre la regresión y los árboles de clasificación suelen centrarse en los resultados de la matriz de confusión, lo que a veces genera conclusiones contradictorias. Ante estos conflictos, los analistas suelen optar por interpretar cada métrica por separado o recurrir a estrategias que no siempre son objetivas ni replicables para llegar a una conclusión general. La regresión es útil cuando se desea predecir valores numéricos con un alto grado de precisión, pero puede no ser tan efectiva para identificar patrones complejos en los datos mientras que los árboles de clasificación son ideales para identificar relaciones no lineales y jerárquicas entre variables (Rodríguez, 2023).

Por lo tanto, el objetivo de este artículo es comparar el rendimiento de los árboles de clasificación y la regresión logística, buscando identificar las situaciones en las que uno es superior al otro, específicamente en términos de la tasa de error de clasificación. Los datos utilizados provienen de estudiantes de nivelación y se analizan para entender la relación entre las variables predictoras y el rendimiento académico. Este estudio también pretende servir de referencia para otros investigadores que busquen comparar modelos de clasificación, brindándoles herramientas para seleccionar el más adecuado según el contexto de su investigación.

**Fundamentación Teórica**

El conjunto de factores asociados al aprendizaje y más específicamente al rendimiento académico en estudiantes universitarios ha sido abordado desde diferentes perspectivas, por lo que se enfatiza en una u otra variable, según el interés de quien realiza el análisis. Sobre esta complejidad que requiere el tratamiento intelectual del progreso académico en los estudiantes de educación superior en Latinoamérica, Real et al. (2024), afirman que “El rendimiento depende de diversos aspectos cotidianos propios del estudiante como la capacidad de trabajo, el esfuerzo, las competencias, la intensidad del estudio, la personalidad y la aptitud, la memoria y motivación” (p. 45).

Teóricamente, las dimensiones del rendimiento académico deben mostrarse cuantitativamente según las escalas de medición que estén establecidas en cada universidad, por lo cual, normalmente, se concreta en un índice académico que pretende informar hacer de las habilidades y conocimientos que poseen los alumnos. Sin embargo, está claro que las expresiones cuantitativas del rendimiento estudiantil son apenas un indicador de las competencias que han desarrollado los aprendices y, en ocasiones, pueden dejar por fuera aspectos que están relacionados con el desarrollo de la personalidad y del pensamiento crítico, en especial cuando se trata de alumnos universitarios (Borja et al., 2021).

Entre los factores que los autores destacan cuando analizan el rendimiento académico de los estudiantes universitarios, se encuentra la motivación. Según los planteamientos de Durán et al. (2021), los estudiantes que están intrínsecamente motivados tienden a tener un mejor desempeño académico, ya que se muestran comprometidos con sus estudios y buscan constantemente superarse a sí mismos, por el contrario, la falta de motivación puede conducir a la apatía y al bajo rendimiento académico, siendo así, es fundamental que los estudiantes encuentren la motivación intrínseca necesaria para alcanzar sus metas académicas y asumir, la pertinencia de los conocimientos.

Igualmente, cuando se trata de los factores determinantes en el rendimiento académico, conviene atender la relevancia de los recursos económicos con que cuentan los alumnos. Es común que aquellos alumnos que provienen de entornos desfavorecidos pueden enfrentar mayores desafíos en su educación universitaria, debido a la falta de recursos económicos y apoyo familiar (Román, 2013). Aunque no tenga una influencia lineal, esta circunstancia es importante, porque la brecha socioeconómica puede manifestarse en diferencias en el acceso a materiales de estudio, tecnología y tutorías personalizadas, lo que impacta directamente en el rendimiento académico de los estudiantes.

Sobre la importancia del factor económico en el rendimiento estudiantil, autores como Procel y Sánchez (2019), señalan que es necesario que la figura del Estado tenga una presencia constante en la creación y dotación de las instituciones de educación superior, de forma que los jóvenes de pocos recursos puedan entrar al sistema, mantenerse y tener éxito. Se entiende que, en la Universidad, algunas carreras presentan una mayor demanda de recursos que otras, pero es importante que se mantenga el apoyo institucional para todos aquellos estudiantes que puedan requerirlo y demuestren las competencias necesarias para rendir en los estudios que cursan.

Ahora, en atención al objetivo de esta investigación, es importante también la comprensión conceptual de lo que implican los modelos de regresión y, los árboles de clasificación, en el contexto estadístico. Respecto a los primeros, Montero (2016), afirma que, en general, esta herramienta de la estadística ha tenido un impacto positivo en diversas disciplinas científicas, desde la economía hasta la medicina, permitiendo a los investigadores analizar la relación entre variables y prever resultados futuros con un alto grado de precisión. Por supuesto, existen diferentes tipos de modelos de regresión, cada uno adaptado a las características específicas de los datos y los objetivos de la investigación que se esté realizando.

La regresión lineal, por ejemplo, es ampliamente utilizada cuando se busca establecer una relación lineal entre una variable independiente y una dependiente. Por otro lado, la regresión logística es fundamental en la predicción de variables categóricas binarias, como sí o no, éxito o fracaso facilitando el ejercicio de los investigadores sociales que deben tomar ciertas

decisiones a la hora de valorar la importancia de ciertos procesos o procedimientos. Al respecto, Berlanga y Vila (2014), consideran que, en un entorno académico, los modelos de regresión son esenciales para validar hipótesis y comprobar teorías.

Contento (2019), afirma que estas herramientas heurísticas tienen una gran funcionalidad a la hora de la identificación de tendencias, la predicción de resultados y la formulación de políticas basadas en evidencia empírica. Así mismo, el autor plantea que, al analizar la significancia estadística de los coeficientes y la bondad de ajuste del modelo, los investigadores pueden determinar la fiabilidad de sus resultados y contribuir al avance del conocimiento en sus respectivas disciplinas. Por ejemplo, un estudio que investigue la relación entre el nivel educativo y el ingreso familiar podría utilizar un modelo de regresión para demostrar si existe una tendencia significativa entre ambas variables y en qué medida se relacionan.

Por su parte, de acuerdo con Caro et al., (2018) los árboles de clasificación como técnica de la estadística aplicada a la investigación combinan elementos de la estadística y la inteligencia artificial para identificar patrones, relaciones y segmentaciones significativas en conjuntos de datos multivariables. Estos instrumentos estadísticos destacan por su capacidad para manejar tanto variables categóricas como continuas, lo que los hace ideales para analizar una amplia gama de datos. Al construir un árbol de clasificación, el investigador puede visualizar de manera clara y concisa cómo se dividen las observaciones en diferentes grupos basados en diversas características.

A su vez, Fernández (2022), profundiza en la cualidad de los árboles de clasificación como un apoyo para la investigación empírica facilitando la identificación de patrones subyacentes y la toma de decisiones informadas. A diferencia de otros métodos más opacos como las redes neuronales, los árboles de clasificación permiten a los investigadores entender fácilmente cómo se llega a una determinada conclusión sobre una situación o circunstancia particular que se ha sometido a estudio, lo cual es determinante en entornos científicos donde la transparencia y la replicabilidad son fundamentales.

En suma, el estudio del rendimiento académico a través de modelos de regresión y árboles de clasificación es fundamental en el ámbito profesional, ya que permite analizar y predecir el desempeño de los estudiantes con un enfoque basado en datos. Estas herramientas estadísticas ofrecen perspectivas que pueden guiar las decisiones y políticas educativas, con el objetivo de promover el éxito académico de todos los alumnos. En ese sentido, conviene que los docentes de los diferentes campos, especialmente en el ámbito universitario, se apoyen este tipo de instrumentos con mayor regularidad para aprovechar sus potencialidades.

**Metodología**

Es un estudio cuantitativo que se detiene en el análisis del rendimiento académico mediante la comparación de modelos de regresión y árboles de clasificación. Según su objetivo se puede considerar un diseño de campo con un carácter descriptivo, al tiempo que se enfoca en la relación entre variables y se utilizan algunas técnicas estadísticas (Hernández et al., 2014). La población estudiada está compuesta por los estudiantes de nivelación en la carrera de estadística de la Carrera de Estadística en la Escuela Superior Politécnica de Chimborazo (ESPOCH).

En congruencia con estas circunstancias se realizó un muestreo intencional entre aquellos estudiantes que mostraron disposición positiva para participar. Ese fue un primer criterio de selección donde, además, se tomó en cuenta que fueran estudiantes con una actitud positiva

hacia la carrera, con diferentes niveles de rendimiento y, con la intención de llevar la carrera hasta alcanzar la titulación, quedando en total, una selección de 25 estudiantes. Respecto al instrumento, se utilizó cuestionarios estandarizados que incluyeron preguntas de los test psicológicos que miden el rendimiento académico utilizados normalmente por los profesionales de psicología

Una vez definida la estructura del cuestionario, la recolección de datos se llevó a cabo en un solo punto en el tiempo por lo cual se puede decir que según el período temporal es transversal y se trabajó con ciertas variables significativas, de acuerdo al objetivo planteado, entre ellas: Edad y Estado civil; Motivación y hábitos de estudio; Interés en la carrera; Estudios y calificaciones previas; Nivel educativo de los padres; Factores psicológicos. Una vez validado por juicio de expertos de la ESPOCH, el instrumento se aplicó y los datos fueron acumulados en una hoja de Excel para los cálculos y, construcción de los gráficos respectivos.

**Resultados**

Primero se realizó un análisis descriptivo de las variables mediante representación gráfica de la frecuencia de cada variable y se realizó un análisis estadístico básico de las variables cuantitativas seleccionadas. El detalle de los resultados encontrados es el siguiente:

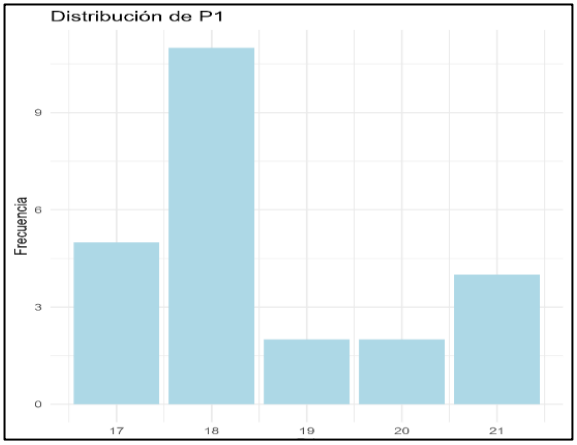



**Gráfico 1.**
Edad de los estudiantes encuestados

Según los planteamientos de Portillo et al., (2017), las edades no son concluyentes, los estudios que han analizado la influencia de la edad de los alumnos en su rendimiento académico, no son concluyentes; sin embargo, se ha observado qua la madurez es una variable importante en cuanto a su nivel de compromiso. En este caso, el gráfico 1 muestra que, en el promedio de edad de los estudiantes de nivelación de la carrera de estadística, predominan los 18 años y la edad que menos frecuencia tiene es de 19 a 20 años, es decir, son personas con una edad coherente con las exigencias de este nivel.

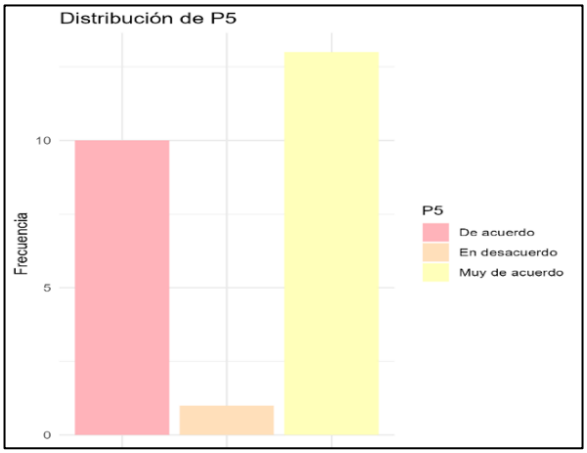



**Gráfico 2.**
Tengo una actitud positiva hacia mí mismo/a.

La mayoría de los estudiantes consideran que tienen una actitud positiva hacia ellos mismos, mientras que son pocos que no están de acuerdo con esta afirmación presentada en el instrumento. Al respecto, se puede afirmar con Gutiérrez que, "las variables que condicionan el rendimiento de los alumnos son numerosas y constituyen una intrincada red en la que resulta muy complejo ponderar la influencia específica de cada una de ellas" (p.141).

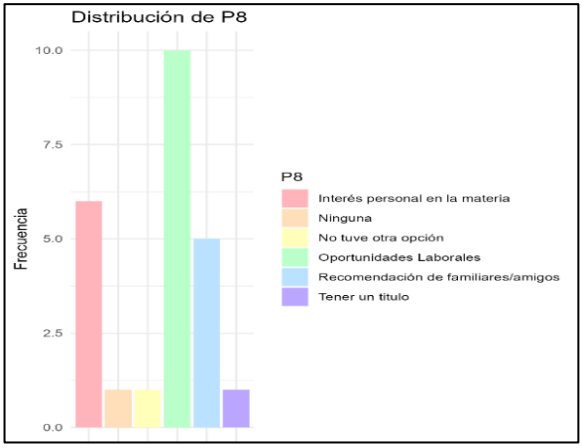



**Gráfico 3.**
¿Cuál es tu principal motivación para estudiar estadística en la ESPOCH?

Como señalan Buenaño et al. (2023), la motivación debe ser revalorizada como componente para mejorar el aprendizaje de estudiantes universitarios con actitud positiva. En las respuestas del gráfico 2, los estudiantes exponen que su principal motivación para

estudiar en la ESPOCH, es la oportunidad laboral que ofrece una vez se terminan los estudios en esta institución.

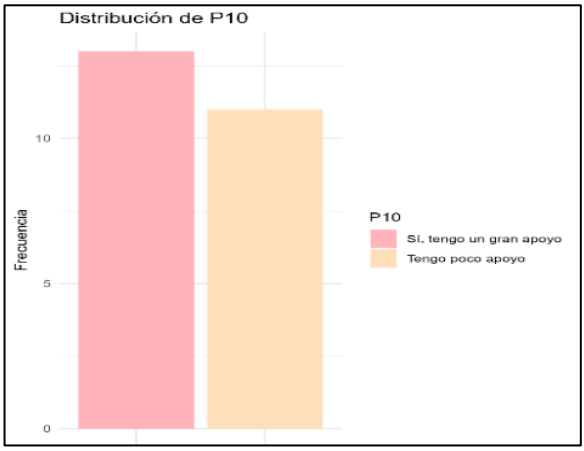



**Gráfico 4.**
¿Sientes que cuentas con un buen sistema de apoyo social (amigos, familiares) para enfrentar los desafíos académicos?

Los estudiantes cuentan con un buen sistema de apoyo social para enfrentar desafíos académicos, pero se puede observar en las respuestas que, también existen personas que manifiestan tener poco apoyo para la prosecución de sus estudios, lo que puede representar una amenaza para que, en el futuro se acreciente la deserción en el contexto de la ESPOCH por lo cual deben tomarse ciertas previsiones, ofreciendo alternativas de apoyo a estos alumnos.

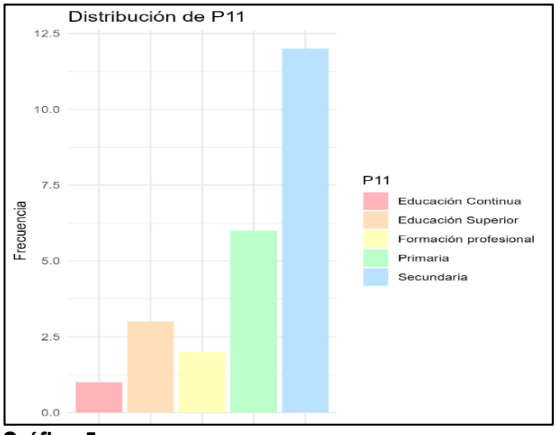



**Gráfico 5.**
¿Cuál es el nivel educativo más alto alcanzado por el padre?

El nivel educativo más alto que alcanzaron los padres de los estudiantes de nivelación de estadística es la secundaria seguida por el nivel de primaria, este dato resulta importante al momento de valorar el significado que para las familias tiene el hecho que los alumnos se mantengan en sus estudios a pesar de las circunstancias adversas que pudieran atravesar. Algunos estudios demuestran que, hay un mayor incentivo para rendir académicamente en los estudios, entre aquellos alumnos, cuyos padres son profesionales, sin que ello implique el descarte de aquellos que tienen padres con menor nivel educativo (Espejel y Jiménez, 2019).

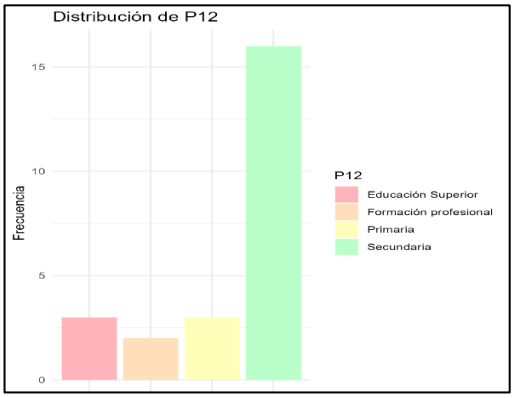



**Gráfico 6.**
¿Cuál es el nivel educativo más alto alcanzado por la madre?

El nivel educativo más alto que alcanzaron las madres de los estudiantes de nivelación de estadística es el de secundaria, seguido por el nivel de primaria. Más allá de su significado estadístico, esta circunstancia puede tener influencia al momento de apoyar la prosecución de los estudios de sus hijos, presentando debilidades al respecto.

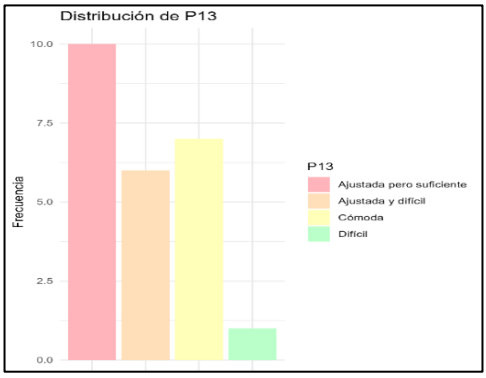



**Gráfico 7.**
¿Cómo describirías tu situación económica personal durante tus estudios?

La situación económica personal de los estudiantes durante sus estudios es ajustada pero suficiente, seguido por cómoda, estas dos opciones predominan, por lo que se puede que se encuentran en una posición económica relativamente favorable para la continuidad de sus estudios. Sobre este tema, Carrasco (2018) considera que, en la medida que los estudiantes universitarios progresan en su carrera, la necesidad de recursos suele crecer, por lo que se hace necesario mantenerse alertas a los factores que coadyuvan para que los alumnos no se sientan obligados a desertar por motivos económicos.

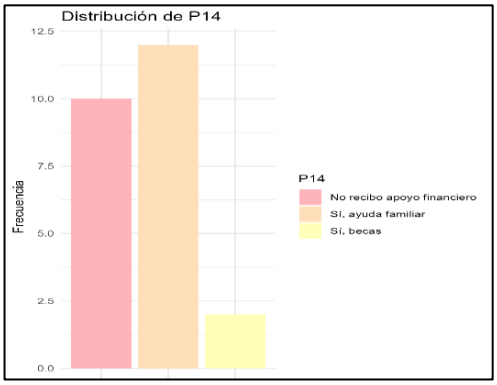



**Gráfico 8.**
¿Recibes algún tipo de apoyo financiero externo como becas, préstamos o ayuda familiar para tus estudios?

Precisamente, con relación a los aspectos económicos, en el grafico 8, se observa que los alumnos no reciben apoyo financiero distinto al que pueden proveerle su familia, esto se presenta como un elemento que preocupa en el caso de aquellos alumnos que provienen de hogares de bajo recursos que requerirían, por tanto, un apoyo institucional que, por los momentos no se da.

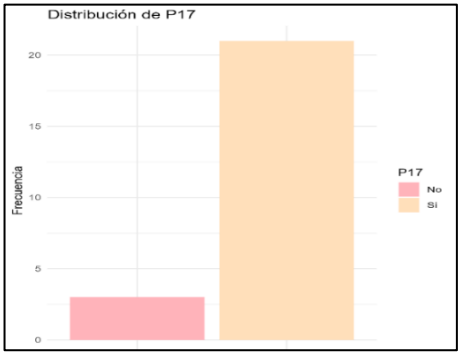



**Gráfico 9.**
¿Planificas detalladamente tus trabajos antes de empezar?

Planificar es una acción que contribuye a asegurar la consecución de los objetivos propuestos en cualquier labor, mucho más en el ámbito de los estudios universitarios (Reyes et al., 2022). Siendo así, las respuestas que ilustra el gráfico 9, reafirma a la planificación como uno de los elementos fundamentales para ordenar y darle sentido a las diferentes actividades académicas que se le exigen a los estudiantes de la ESPOCH, pues, la mayoría de los consultados optan por la respuesta "sí", esto quiere decir que los estudiantes se consideran a sí mismos, personas organizadas.

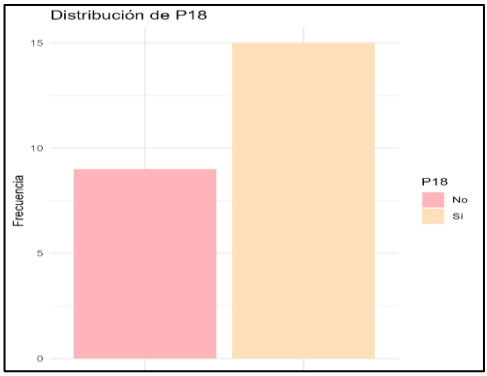



**Gráfico 10.**
¿Distribuyes tu horario del día adecuadamente para poder hacer todo lo planeado?

La planificación es exitosa en la medida en que responde a una gestión de los diferentes factores que intervienen en el alcance de los objetivos que se ha fijado la misma, en ese orden de ideas, los estudiantes que responden, "sí", a la indagatoria anterior, demuestran ser planificadores para sus actividades académicas.

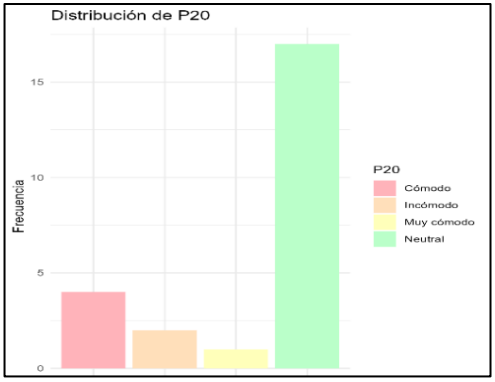



**Gráfico 11.**
¿Cuán cómodo te sientes trabajando con números y datos?

Al tratarse de alumnos en una carrera que obliga a la familiarización con datos numéricos, la pregunta del gráfico 11, profundiza en la motivación estudiantil para este tipo de labor. Allí se observa que los estudiantes responden que se sienten neutral dejando para la interpretación esa actitud; sin embargo, en el transcurso del estudio, se pudo corroborar con la observación que, hay una actitud positiva de los alumnos frente al trabajo cuantitativo, algo que es alentador para su formación en la carrera de estadística.

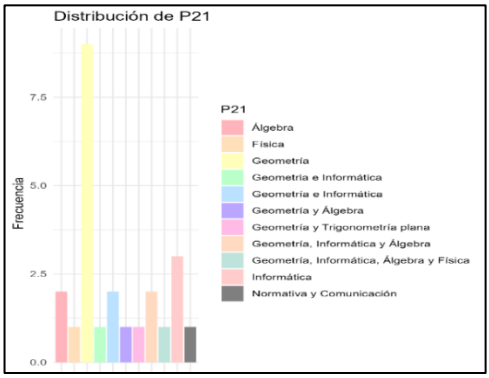



**Gráfico 12.**
¿Qué materia le resulta más difícil hasta el momento?

A pesar de lo que se afirma como interpretación del grafico 11, en los resultados del 12, La materia que resulta más difícil hasta el momento para los estudiantes de estadística ha sido la geometría. Este resultado puede considerarse un indicador de las dificultades que representan las materias de carácter cuantitativo para los alumnos de la ESPOCH, a pesar de lo señalado respecto a los resultados del gráfico precedente y, de estar incorporados en una carrera con esa orientación.

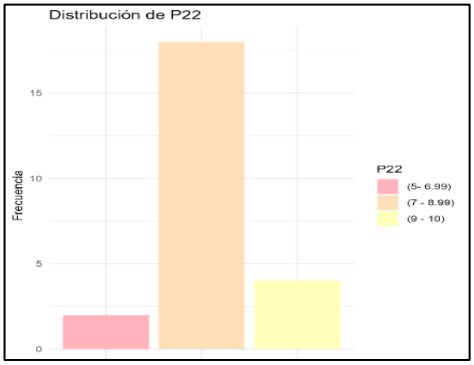



**Gráfico 13.**
¿En qué rango se encuentra su calificación promedio al graduarse?

El rango que la mayoría de los estudiantes se encuentra es de (7-8.99) en su calificación de promedio de grado, además podemos decir que la calificación entre (9-10) es común, pero menos comunes que el anterior. Si bien, este dato numérico no refleja la complejidad que representa la medición del rendimiento académico universitario, se entiende que, aquellos alumnos con un rango mayor, pueden sentirse más dispuestos a continuar y concluir sus estudios.

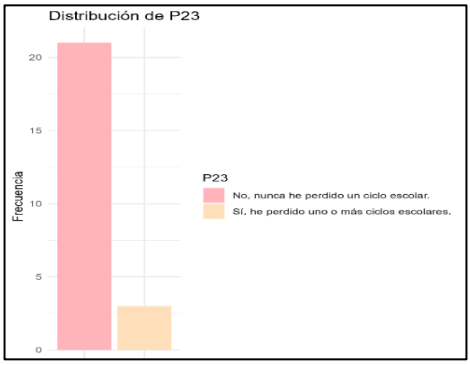



**Gráfico 14.**
¿Ha perdido algún ciclo escolar?

En este caso se les ha cuestionado acerca de si han perdido o no un ciclo escolar, a lo que los estudiantes han respondido que no, nunca han perdido un periodo académico, pero existe un pequeño grupo que sí, con esto se puede deducir la necesidad de fortalecer la vocación de aquellos alumnos que han presentado la dificultad, pues, de acuerdo a Borja et al. (2021), una de las características de quienes desertan de la educación superior, es la búsqueda de apoyos transitorios para evitar los pequeños fracasos que se presentan a lo largo de la carrera y, de contar con ese apoyo, muchos de ellos se quedarían en la universidad.

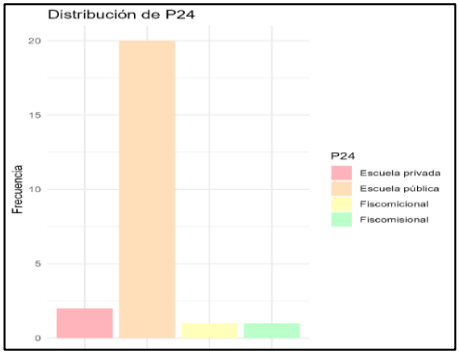



**Gráfico 15.**
¿De qué tipo de escuela proviene?

La frecuencia más alta es la escuela pública, esto quiere decir que la mayoría de los estudiantes provienen de escuelas públicas y, es probable que la mayoría de ellos vean en la posibilidad de una carrera universitaria, una opción de movilidad social. Al respecto, Bustamante y Cabrera enfatizan en los diversos factores que afectan al alumno que ingresa a la universidad, destacando, el lugar de donde proceden.

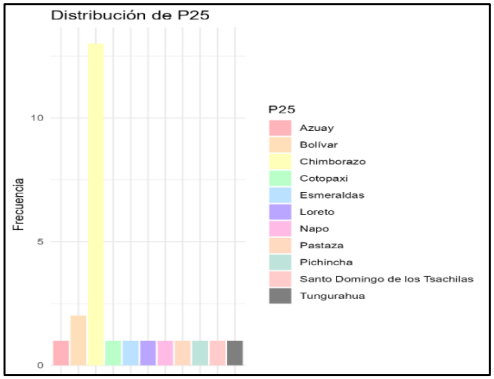



**Gráfico 16.**
¿En qué provincia cursó el colegio?

La mayoría de los estudiantes de la carrera de estadística de nivelación, provienen de un colegio ubicado en la provincia de Chimborazo, un aspecto significativo porque no deberían tener mayores inconvenientes de adaptación como si ocurriría si procedieran de regiones alejadas.

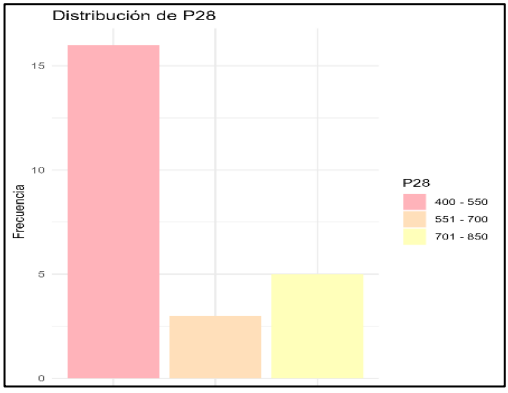



**Gráfico 17.**
¿Qué puntaje obtuvo en el examen realizado para el ingreso a la universidad?

La mayoría de los estudiantes que ingresaron a la carrera de estadística obtuvieron de 400 a 550 puntos en el examen, esto preocupa ya que en el futuro podría haber deserción en la

carrera si no existe un refuerzo constante, tanto motivacional, como formativo, para contribuir a elevar el sentido de pertenencia de los alumnos, algo que -según González et al. (2023)-, puede favorecer la permanencia en el sistema de educación superior y, el rendimiento académico

**Modelo de clasificación: Regresión logístico multinomial**

La clasificación mediante regresión logística tiene como objetivo estimar, clasificar e identificar variables que influyen en el rendimiento académico para la construcción de modelos mediante esta técnica se utilizó todo el conjunto de estudio.

```
> modelo_logistico <- multinom(P28 ~ ., data = data)
# weights:  249 (164 variable)
initial  value 26.366695
iter  10 value 0.076659
iter  20 value 0.003904
final  value 0.000091
converged
```

**Gráfico 18.**
Ajuste del modelo

El ajuste del modelo de regresión logística multinomial ha convergido exitosamente. La función de pérdida disminuyó significativamente, desde un valor inicial de 26.367 hasta un valor final de 0.000091, esto indica que el modelo ha encontrado parámetros que se ajustan bien a los datos.

**Prueba de la significancia del modelo de Regresión Logística**

```
> deviance(modelo_logistico)
[1] 0.0001812641
> AIC(modelo_logistico)
[1] 96.00018
```

**Gráfico 19.**
Resultados de la prueba de significancia

Como podemos observar tenemos un AIC del 96 esto confirma que el modelo si se ajusta correctamente a nuestros datos, se puede utilizar para comprender y mejorar el rendimiento

**Tabla de clasificación de Regresión Logística**

```
> print("Matriz de confusión para el modelo logístico:")
[1] "Matriz de confusión para el modelo logístico:"
> print(confusion_logistico)
Confusion Matrix and Statistics

          Reference
Prediction  400 - 550 551 - 700 701 - 850
  400 - 550        16         0         0
  551 - 700         0         3         0
  701 - 850         0         0         5

Overall Statistics

               Accuracy : 1
                 95% CI : (0.8575, 1)
    No Information Rate : 0.6667
    P-Value [Acc > NIR] : 5.94e-05

                  Kappa : 1

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 400 - 550 Class: 551 - 700 Class: 701 - 850
Sensitivity                    1.0000           1.000           1.0000
Specificity                    1.0000           1.000           1.0000
Pos Pred Value                 1.0000           1.000           1.0000
Neg Pred Value                 1.0000           1.000           1.0000
Prevalence                     0.6667           0.125           0.2083
Detection Rate                 0.6667           0.125           0.2083
Detection Prevalence           0.6667           0.125           0.2083
Balanced Accuracy              1.0000           1.000           1.0000
```

**Gráfico 20.**
Tabla de clasificación. Modelo Logístico

El desempeño del modelo logístico, según la matriz de confusión, alcanzó una precisión total del 100% y un índice Kappa de 1, lo que indica que todas las observaciones han sido clasificadas correctamente en sus categorías correspondientes. Cada una de las clases (400 - 550, 551 - 700, 701 - 850) muestra una sensibilidad, especificidad y valor predictivo positivo del 100%, lo que implica que no ha habido errores en las predicciones. La distribución de las clases revela que la mayor parte de las observaciones se encuentra en la categoría 400 - 550, que representa el 66.67%, mientras que las otras dos categorías tienen prevalencias significativamente menores (12.5% y 20.83%). Asimismo, la tasa de detección y la prevalencia equilibrada son perfectas para todas las clases, lo que reafirma la capacidad del modelo para clasificar sin cometer errores.

**Tabla de clasificación del Árbol de Clasificación**

La matriz de confusión para el modelo de árbol de clasificación indica una precisión general del 70.83%, que supera levemente la tasa de información nula del 66.67%, aunque esta diferencia no es significativa (p-valor de 0.4238). El índice Kappa, que es de 0.4227, sugiere una concordancia moderada entre las predicciones del modelo y las clases reales. Analizando más a fondo, el modelo clasifica correctamente el 81.25% de las observaciones en la clase 400 - 550, pero no logra clasificar adecuadamente las observaciones en la clase 551 - 700, donde tanto la sensibilidad como el valor predictivo positivo son 0.

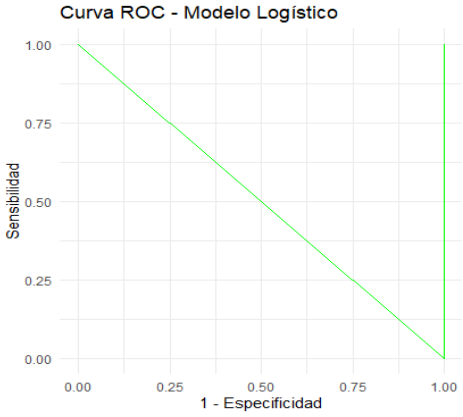
```
> print("Matriz de confusión para el modelo de árbol de clasificación:")
[1] "Matriz de confusión para el modelo de árbol de clasificación:"
> print(confusion_arbol)
Confusion Matrix and Statistics

          Reference
Prediction  400 - 550 551 - 700 701 - 850
  400 - 550        13         1         1
  551 - 700         0         0         0
  701 - 850         3         2         4

Overall Statistics

               Accuracy : 0.7083
                 95% CI : (0.4891, 0.8738)
    No Information Rate : 0.6667
    P-Value [Acc > NIR] : 0.4238

                  Kappa : 0.4227

 Mcnemar's Test P-Value : 0.2615

Statistics by Class:

                     Class: 400 - 550 Class: 551 - 700 Class: 701 - 850
Sensitivity                    0.8125           0.000           0.8000
Specificity                    0.7500           1.000           0.7368
Pos Pred Value                 0.8667             NaN           0.4444
Neg Pred Value                 0.6667           0.875           0.9333
Prevalence                     0.6667           0.125           0.2083
Detection Rate                 0.5417           0.000           0.1667
Detection Prevalence           0.6250           0.000           0.3750
Balanced Accuracy              0.7812           0.500           0.7684
```

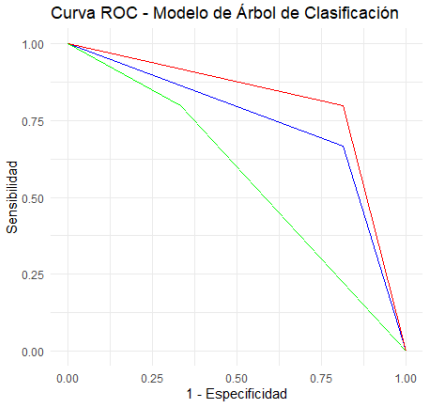
**Gráfico 21.**
Árbol de clasificación

En cuanto a la clase 701 - 850, su sensibilidad es del 80%, pero su valor predictivo positivo es relativamente bajo, alcanzando solo el 44.44%. La especificidad es alta para la clase 551 - 700 (100%), aunque menos precisa para las otras dos categorías. La tasa de detección es mayor para la clase 400 - 550 (54.17%) y significativamente más baja para la clase 701 - 850 (16.67%). Por último, la precisión balanceada del modelo muestra un rendimiento aceptable para las clases 400 - 550 y 701 - 850, pero es deficiente para la clase 551 - 700.

**Característica de funcionamiento del receptor (Curva ROC) del Árbol de clasificación**

El área bajo la curva ROC (AUC) para el modelo de árbol de clasificación es de 0.7042, lo que indica que el modelo tiene una capacidad moderada para diferenciar entre las tres categorías (400 - 550, 551 - 700, 701 - 850). Aunque un AUC superior a 0.7 sugiere un desempeño aceptable y cierta habilidad para clasificar correctamente las instancias, este valor se aleja de 1, lo que implica que el modelo de árbol no es tan eficiente en la clasificación en comparación con el modelo logístico, que alcanzó un AUC de 1. Este resultado destaca las limitaciones del modelo de árbol y sugiere que podría mejorar con ajustes adicionales para optimizar su capacidad de clasificación.




```
> print("Modelo Logístico:")
[1] "Modelo Logístico:"
> print(confusion_logistico$overall['Accuracy'])
Accuracy
       1
> print("Área bajo la curva ROC:")
[1] "Área bajo la curva ROC:"
> print(roc_logistico$auc)
Multi-class area under the curve: 1
```

```
> print("Modelo de Árbol de Clasificación:")
[1] "Modelo de Árbol de Clasificación:"
> print(confusion_arbol$overall['Accuracy'])
 Accuracy
0.7083333
> print("Área bajo la curva ROC:")
[1] "Área bajo la curva ROC:"
> print(roc_arbol$auc)
Multi-class area under the curve: 0.7042
```

**Gráfico 22.**
Comparación de los métodos

Al comparar ambos modelos, se observa que el modelo logístico presenta un rendimiento notablemente superior al del modelo de árbol de clasificación. La matriz de confusión del modelo logístico revela una precisión del 100%, y su área bajo la curva ROC (AUC) alcanza el valor de 1, lo que significa que puede clasificar todas las categorías de manera perfecta. Por otro lado, el modelo de árbol de clasificación tiene una precisión aproximada del 70.83% y un AUC de 0.7042, lo que indica una capacidad moderada para diferenciar entre las distintas clases. Si bien el modelo de árbol ofrece un rendimiento aceptable en la clasificación, el modelo logístico demuestra una clara superioridad en precisión y en la capacidad de distinguir entre categorías. Esto sugiere que el modelo logístico es más apropiado para esta tarea específica de clasificación.

**Discusión**

La presente investigación ha permitido analizar de manera integral los factores que influyen en el rendimiento académico de los estudiantes de nivelación. A través de métodos estadísticos como la regresión logística multinomial y el árbol de clasificación, se han obtenido resultados significativos que contribuyen a la comprensión de la dinámica del rendimiento académico en este grupo de estudiantes.

Según Martínez et al (2020), cuando se analizan las variables sociodemográficas y académicas que impactan el rendimiento académico, resalta la complejidad del fenómeno educativo. La regresión logística multinomial se ha mostrado como una herramienta potente y precisa para evaluar estas relaciones. La precisión del 100% y el AUC de 1 en la matriz de confusión del modelo logístico indican no solo una capacidad de clasificación perfecta, sino también la relevancia de las variables seleccionadas para explicar el rendimiento académico. Esto sugiere que factores como la satisfacción con la calidad de la educación y características demográficas son determinantes clave en el desempeño estudiantil (Almeida y Freire, 2024).

Por otro lado, el uso del árbol de clasificación ha proporcionado un enfoque complementario, permitiendo una visualización clara y comprensible del proceso de toma de decisiones. Sin embargo, su rendimiento inferior, con una precisión del 70.83% y un AUC de 0.7042, pone de manifiesto sus limitaciones en comparación con el modelo logístico. A esto se refieren Gómez et al (2019), cuando plantean la necesidad de utilizar múltiples enfoques estadísticos para enriquecer la comprensión del rendimiento académico y proporcionar un marco más robusto para la toma de decisiones en el ámbito educativo.

Adicionalmente, la discrepancia en el rendimiento entre ambos modelos subraya la importancia de seleccionar adecuadamente las metodologías analíticas en función de los objetivos de investigación. Al respecto, autores como Pérez y González (2018) y; Devincenzi et al (2017), señalan que, la regresión logística se ha mostrado más eficaz en la identificación de relaciones cuantitativas, mientras que el árbol de clasificación, aunque menos preciso, puede ser útil en contextos donde la interpretación es crucial.

Finalmente, esta investigación no solo contribuye al cuerpo de conocimiento existente sobre el rendimiento académico, sino que también proporciona implicaciones prácticas para diseñadores de políticas educativas y académicos. La identificación de variables influyentes puede guiar la implementación de estrategias educativas específicas que aborden las necesidades de los estudiantes y, potencialmente, mejoren su rendimiento académico. Las limitaciones observadas en el modelo de árbol de clasificación sugieren que futuras

investigaciones deberían considerar ajustes y mejoras en su aplicación, así como la incorporación de técnicas de modelado complementarias para optimizar los resultados

## Conclusiones

A partir de la revisión de literatura y el análisis estadístico de los datos recolectados, se han identificado varias variables sociodemográficas y académicas que tienen un impacto significativo en el rendimiento académico de los estudiantes de nivelación. Estas variables incluyen, entre otras, el nivel de satisfacción con la calidad de la educación recibida y características demográficas específicas.

La aplicación de la regresión logística multinomial ha demostrado ser una metodología robusta para explorar la relación cuantitativa entre las variables independientes y el rendimiento académico. La matriz de confusión resultante evidencia una precisión del 100% y un área bajo la curva (AUC) de 1, lo que valida la capacidad del modelo para clasificar correctamente todas las instancias en sus respectivas categorías, reflejando su idoneidad para este tipo de análisis.

El modelo de árbol de clasificación, aunque proporciona una representación visual comprensible del proceso de toma de decisiones, presentó un rendimiento inferior al de la regresión logística, con una precisión del 70.83% y un AUC de 0.7042. Estos resultados indican que, a pesar de su utilidad en la visualización, el árbol de clasificación tiene limitaciones en su capacidad discriminativa, sugiriendo que su aplicación debería complementarse con otras técnicas más precisas para un análisis exhaustivo del rendimiento académico.

En cuanto a las limitaciones de este estudio, se pueden mencionar la poca literatura que se ha producido en los últimos años para poder analizar con mayor detalle los resultados que ha arrojado la investigación. Ciertamente en la discusión de los mismos se puede observar algo de debilidad por carecer de mayores referencias que permitan contrastar o comparar lo que arrojó el cuestionario aplicado a los alumnos y avizorar las perspectivas que se pueden presentar más adelante en el contexto estudiado.

## Declaración de Conflictos de Interés

No declaran conflictos de interés.

## Contribución de autores

Los autores declaran haber contribuido en la misma proporción.

## Financiamiento

Este artículo no recibió financiamiento.

## Referencias

Almeida, L., & Freire, D. (2024). Análisis de la satisfacción estudiantil de acuerdo con el nivel académico del profesorado de un Instituto de Educación Superior. *Digital Publisher CEIT, 9*(4), 596-611. https://doi.org/https://doi.org/10.33386/593dp.2024.4.2529

Álvarez, M., Ponce, M., Alban, J., & Zambrano , L. (2024). Revisión de modelos estadísticos para pronosticar el desempeño académico en estudiantes universitarios. *Journal Scientific MQRInvestigar, 8*(2), 3806-3823. https://doi.org/10.56048/MQR20225.8.2.2024.3806-3823

Berlanga, V., & Vila, R. (2014). Cómo obtener un Modelo de Regresión Logística Binaria con SPSS. *Reire, 7*(2), 105-118. http://www.ub.edu/ice/reire.htm

Borja, G., Martínez, J., Barreno, S., & Haro, O. (2021). Factores asociados al rendimiento académico: Un estudio de caso. *Educare, 23*(3), 54-68. https://portal.amelica.org/ameli/journal/375/3752842004/3752842004.pdf

Buenaño, N., Palacios, C., Soplapuco, J., & Reluz, F. (2023). Importancia de la motivación para el aprendizaje universitario: Una revisión integradora. *Revista De Ciencias Sociales, 29*(4), 371-385. https://produccioncientificaluz.org/index.php/rcs/article/view/41261/47527

Caro, N., Guardiola, M., & Ortiz, P. (2018). Árboles de clasificación como herramienta para predecir dificultades financieras en empresas Latinoamericanas a través de sus razones contables. *Contaduría y Administración, 63*(1), 1-14. https://doi.org/https://doi.org/10.22201/fca.24488410e.2018.1148

Carrasco, Y. (2018). Los estudiantes universitarios y el aspecto socioeconómico. *Conciencia Digital, 1*(2), 6-15. https://doi.org/10.33262/concienciadigital.v1i2.946

Contento, M. (2019). *Estadística con Aplicaciones en R*. Utadeo. https://www.utadeo.edu.co/sites/tadeo/files/node/publication/field_attached_file/libro_estadistica_con_aplicaciones_en_r_def_ago_11.pdf

Devincenzi, G., Rohde, G., Giraudo, M., Bonaffini, M., & Bernaola, G. (2017). El análisis del rendimiento académico a través de modelos matemáticos y estadísticos. *Trim,* 13, 59-72. https://dialnet.unirioja.es/descarga/articulo/7609224.pdf

Durán , C., Casadiegos , M., & Carrascal, A. (2021). Motivación en estudiantes universitarios como factor generador de la calidad educativa. *Boletín Redipe, 10*(13), 443-455. https://revista.redipe.org/index.php/1/article/download/1758/1673/2863

Espejel, M., & Jiménez, M. (2019). Nivel educativo y ocupación de los padres: Su influencia en el rendimiento académico de estudiantes universitarios. *Ride, 10*(19), 1-20. https://doi.org/10.23913/ride.v10i19.540

Fernández, S. (2022). *Métodos de Regresión y Clasificación Basados en Árboles. Tesis de Grado en Ingeniería y Organización Industrial*. Universidad de Valladolid. https://uvadoc.uva.es/bitstream/handle/10324/53822/TFG-I-2181.pdf?sequence=1

Garbanzo, G. (2014). Factores asociados al rendimiento académico tomando en cuenta el nivel socioeconómico: Estudio de regresión múltiple en estudiantes universitarios. *Educare, 18*(1), 119-154. https://www.scielo.sa.cr/pdf/ree/v18n1/a06v18n1.pdf

Gómez, M., Maza, A., & Rocha , D. (2019). Análisis estadísticos de los Estilos de aprendizaje y su relación con la evaluación del rendimiento académico. *Revista Ingeniería desarrollo e innovación, 2(*1), 50-79. http://dx.doi.org/10.32012/26195259/2019.v2i1.73

González, J., Agredo, J., Campo, M., Hernández, M., & Oviedo, N. (2023). El sentido de pertenencia y el compromiso estudiantil en el contexto de educación superior a distancia con mediación virtual. *Revista Educación, 47*(1), 1-19. https://www.redalyc.org/journal/440/44072432022/44072432022.pdf

Hernández, S., Fernández, C., & Baptista, P. (2014). *Metodología de la Investigación*. Mc Graw Hill. https://www.uca.ac.cr/wp-content/uploads/2017/10/Investigacion.pdf

Lizares, M. (2017). *Comparación de modelos de clasificación: regresión logística y árboles de clasificación para evaluar el rendimiento académico*. Tesis de Grado en Estadística. Universidad Nacional Mayor de San Marcos. https://cybertesis.unmsm.edu.pe/backend/api/core/bitstreams/47cd4eb4-855f-4410-80ce-c3ecb2df1d86/content

Martínez, J., Ferrás, Y., Bermúdez, L., Ortiz, Y., & Pérez, E. (2020). Rendimiento académico en estudiantes Vs factores que influyen en sus resultados: una relación a considerar. *Edumecentro, 12*(4), 105-121. http://scielo.sld.cu/pdf/edu/v12n4/2077-2874-edu-12-04-105.pdf

Montero, R. (2016). *Modelos de regresión lineal múltiple. Documentos de Trabajo en Economía Aplicada.* Universidad de Granada. http://www.ugr.es/~montero/matematicas/regresion_lineal.pdf

Mures, J., García , A., & Vallejo, E. (2005). Aplicación del Análisis Discriminante y Regresión Logística en el estudio de la morosidad en las entidades financieras. Comparación de resultados. *Pecvnia, 1*(1), 175-199. https://dialnet.unirioja.es/descarga/articulo/1281700.pdf

Norambuena, J., Badilla, M., & López, Y. (2022). Modelos predictivos basados en uso de analíticas de aprendizaje en educación superior: una revisión sistemática. *Linguagem e Tecnologia,* 15, 1-22. https://doi.org/10.35699/1983-3652.2022.36310

Pérez, J., & González, F. (2018). Árboles de clasificación vs regresión logística en el desarrollo de competencias genéricas en ingeniería. *Computación y Sistemas, 22*(4), 1519–1541. https://doi.org/10.13053/CyS-22-4-2804

Portillo, P., Marín, L., & Scarpellini, S. (2017). La edad y el rendimiento académico universitario. *Revista de la Universidad de Zaragoza, 1(*1), 28-34. https://doi.org/10.26754/CINAIC.2017.000001_002

Procel, M., & Sánchez, D. (2019). . Factores socioeconómicos que afectan el rendimiento académico de los estudiantes de nivelación de la Escuela Superior Politécnica de Chimborazo, período 2018-1S campus matriz. Investigación e innovación en la Enseñanza Superior. *Nuevos contextos, nuevas ideas,* 1(1), 1021-1031. https://rua.ua.es/dspace/bitstream/10045/98998/1/Investigacion-e-innovacion-en-la-ES_096.pdf

Real, R., Guevara, A., Morales, I., Chibas, E., Cañete, E., Carballo, M., . . . Noldin, A. (2024). Factores asociados al rendimiento académico en estudiantes universitarios de Latinoamérica en 2023. *Investigación en Educación Médica, 13*(51), 42-52. https://doi.org/10.22201/fm.20075057e.2024.51.23580

Reyes, N., Meneses, A., & Díaz, A. (2022). Planificación y gestión del tiempo académico de estudiantes universitarios. *Formación Universitaria, 15*(1), 57-72. http://dx.doi.org/10.4067/S0718-50062022000100057

Rodríguez, D. (Junio de 2023). *Análisis de Datos y Machine Learning. Tesis de Grado en Estadística.* Universidad de Salamanca. https://gredos.usal.es/bitstream/handle/10366/157350/TFG%20Damia%CC%81n%20Rodri%CC%81guez%20Go%CC%81mez.pdf?sequence=1&isAllowed=y

Román, M. (2013). Factores asociados al abandono y la deserción escolar en América Latina: Una mirada en conjunto. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, 11*(2), 33-59. https://www.redalyc.org/pdf/551/55127024002.pdf