

Modelo de predicción de deserción: Un estudio de caso de estudiantes de la Universidad Politécnica Estatal del Carchi

Dropout prediction model: A case study of students at the Universidad Politécnica Estatal del Carchi.

Francisco Javier Montalvo Márquez^[0000-0003-2597-1113], Nadia Nathaly Sánchez-Pozo^[0000-0003-0509-2082]

¹ Universidad Politécnica Estatal del Carchi, Posgrado. Tulcán, 040101, Ecuador.

¹{javier.montalvo, nadia.sanchez}@upec.edu.ec

CITA EN APA:

Montalvo Márquez, F., & Sánchez Pozo, N. (2023). Modelo de predicción de deserción: Un estudio de caso de estudiantes de la Universidad Politécnica Estatal del Carchi. *Prometeo Conocimiento Científico*, 3(1), e34. <https://doi.org/10.55204/pcc.v3i2.e34>

Recibido: 2023-06-09

Revisado: 2023-06-13 al 2023-06-23

Corregido: 2023-06-30

Aceptado: 2023-07-03

Publicado: 2023-07-05

Prometeo

Conocimiento Científico

ISSN: 2953-4275



Los contenidos de este artículo están bajo una licencia de Creative Commons Attribution 4.0 International (CC BY 4.0)

Los autores conservan los derechos morales y patrimoniales de sus obras.

Resumen.

Introducción: La tasa de abandono estudiantil en las instituciones de educación superior es un problema global que requiere soluciones personalizadas y adaptadas a cada institución. El presente artículo se realiza en la Universidad Politécnica Estatal del Carchi (UPEC) en Ecuador.

Objetivo: Identificar los factores determinantes del abandono universitario mediante un modelo logístico que permita predecir si un estudiante puede abandonar la universidad de manera prematura.

Métodos: Como técnica de modelado estadístico multivariante se utiliza un modelo de regresión logística LR. La muestra de estudiantes corresponde a 3821 registros de matrículas en la UPEC.

Resultados: Los resultados evidencian que el modelo logístico propuesto presenta una precisión del 90%. Además, se identificó una relación inversa entre la deserción estudiantil, el número de miembros en el hogar y el nivel educativo del padre y la madre; por lo que se evidencia la necesidad de instaurar medidas de apoyo y prevención específicas para los estudiantes de hogares con menores niveles educativos y mayores cargas familiares.

Conclusiones: El abandono estudiantil en la UPEC ocurre principalmente durante el primer nivel de carrera, determinando además que los factores relevantes que influyen en la deserción estudiantil en esta universidad son: la edad, el número de miembros en el hogar, el nivel de estudio y el promedio académico. Esta información permitirá a las autoridades universitarias plantear medidas preventivas para mejorar la permanencia y el éxito académico de sus estudiantes.

Palabras Clave: deserción; educación; estudiantes universitarios; regresión logística.

Abstract:

Introduction: The student dropout rate in higher education institutions is a global problem that requires customized solutions adapted to each institution. This article is carried out at the Universidad Politécnica Estatal del Carchi (UPEC) in Ecuador.

Objective: To identify the determinants of college dropout using a logistic model to predict whether a student may leave college prematurely.

Methods: A LR logistic regression model was used as a multivariate statistical modeling technique. The sample of students corresponds to 3821 enrollment records at UPEC.

Results: The results show that the proposed logistic model is 90% accurate. In addition, an inverse relationship was identified between student dropout, the number of members in the household and the educational level of the father and mother; thus, the need to implement specific support and prevention measures for students from households with lower educational levels and higher family burdens is evident.

Conclusions: The UPEC found that the most important factors influencing student attrition are: age, number of household members, level of study and academic average. This information will allow the university authorities to propose preventive measures to improve the retention and academic success of their students.

Keywords: dropout; education; university students; logistic regression.

1. INTRODUCCIÓN

La deserción o abandono universitario es un problema que incide de manera significativa a las universidades a nivel mundial; este fenómeno se refiere a la situación en la que un estudiante abandona sus estudios antes de completar su programa académico, provocando que los estudios de la desincorporación de estudiantes de su carrera universitaria sin lograr titularse cada vez sean más frecuentes (Ramírez et al., 2017). Además, la deserción o abandono estudiantil es considerado como un parámetro relevante en el seguimiento académico que se realiza en las universidades, para (de Oliveira et al., 2021).

El progreso de un país depende de la educación de sus ciudadanos, por lo que se puede decir que el abandono universitario es un problema grave que influye en todo el mundo (Niyogisubizo et al., 2022). Algunos autores abordan esta temática desde el estudio de la retención estudiantil; y otros autores proponen su análisis desde la perspectiva de la deserción o abandono estudiantil (Caselli Gismondi y Urrelo Huiman, 2021). En definitiva, los dos puntos de vista son válidos ya que la retención es el complemento de la deserción estudiantil.

A nivel mundial, según Tomás y Gutiérrez (2019) la tasa de deserción de estudiantes universitarios es elevada, generando inconvenientes no solo en ellos, sino también a las instituciones a las cuales pertenecen y la sociedad. Además, esta problemática representa pérdidas de capital humano, y financiero; provocando cierto impacto negativo en los universitarios que interrumpen sus estudios y en sus familias (Bedregal-Alpaca et al., 2020).

Según estudios presentados por el Global Education Monitoring hasta el 2014 se registran 207 millones de universitarios y la UNESCO en su último informe emitido para el 2016 indica que la deserción alcanza el 40%, y la tasa de abandono universitario es de alrededor del 30% para los Estados Unidos, lo que sugiere que casi un tercio de los estudiantes no completan su programa de estudios.

En un estudio presentado por Ferreyra et al. (2017) señalan que en América Latina y el Caribe se cuenta con 20 millones de estudiantes a nivel universitario. Sin embargo, muestran una gran preocupación debido a que apenas el 50% de la población que algún momento inició estudios de nivel superior ha logrado su titulación. Para Pierella et al. (2020) el abandono se evidencia en el primer año de universidad, asimismo se identifica en Colombia que la tasa de deserción corresponde al 37% y el 36% que se evidencian en el primer año de estudio (Ferreyra et al., 2017).

Para el Ecuador, la Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT) indica en su boletín anual que en el 2018 se registra 632.541 matrículas en todas las Instituciones de Educación Superior, así mismo en el 2019 se registra en promedio una tasa de deserción del 26%. En la UPEC, en los últimos tres periodos académicos el promedio de abandono es del 33%.

El Rendimiento académico en las universidades a nivel nacional e internacional ha sido objeto de estudio como una estrategia del cumplimiento de la calidad académica (Caselli y Huiman, 2021). La UPEC centra su atención en la tasa de retención, considerándola un indicador del modelo de evaluación de

carreras; al ubicarse en el norte del Ecuador, en la frontera con Colombia, posiblemente su ubicación geográfica pueda afectar el abando estudiantil.

Suárez et al. (2021) estudian esta problemática en dos ciudades de frontera Cúcuta - Norte de Tumaco-Nariño y Santander; estas localidades fronterizas presentan problemas de inseguridad, desatención del Estado, movimiento migratorio, desempleo, conflicto armado y contrabando de estupefacientes; se determinó que los problemas económicos en las familias son la causa principal del abandono estudiantil en dichas zonas.

Para Álvarez et al. (2020) comprender las causas de la desvinculación escolar es fundamental en procura de minimizar este efecto y estas no se pueden generalizar, cada estudiante presenta diferentes realidades, López-Aguilar et al. (2022) lo definen como un problema multicausal. Poveda (2019) identifica seis factores de deserción universitaria: salud, económico, académico, social, familiar, motivacional. Para Muzumdar et al. (2020) las razones de esta problemática son: psicoeducativas, evolutivas, familiares, económicas, institucionales y sociales.

Zárate-Valderrama et al. (2021) concluyen que el factor más importante es la relación de créditos aprobados por un estudiante con respecto a los créditos que debía haber tomado; además, el examen de admisión de los estudiantes no es significativo en la deserción estudiantil. Bedregal-Alpaca et al. (2020) proponen que el desempeño escolar es un posible indicador a este problema. Sin embargo, precisan que, además de las calificaciones, el comportamiento académico, el desempeño relativo en relación con la cohorte y las asignaturas aprobadas son factores determinantes del éxito académico en los estudiantes.

La estimación del comportamiento del ausentismo estudiantil puede realizarse utilizando diferentes técnicas estadísticas y computacionales, dependiendo de los recursos disponibles. Pérez et al. (2018) utilizan un modelo de regresión logística en donde estiman hasta en el 86.4 % la probabilidad de abandono para cada estudiante que registra su ingreso a la universidad.

Poveda (2019) realiza un análisis cuantitativo apoyándose en ecuaciones estructurales para determinar los factores influyentes en deserción universitaria al igual que Muzumdar et al. (2020) quienes analizan los datos obtenidos con dos modelos de ecuaciones estructurales mediante variables latentes. Zárate-Valderrama et al. (2021) utilizan una técnica de extracción de datos para buscar patrones en la deserción estudiantil mediante un algoritmo que permite clasificar en una primera fase de aprendizaje, luego en una segunda fase clasifica los datos.

Bedregal-Alpaca et al. (2020) plantean descubrir patrones en datos con gran volumen apoyándose en minería de datos con métodos de árbol de decisión o redes neuronales, (Kittinan et al., 2018) plantean modelos de predicción con una precisión de hasta el 92%. Así mismo, Quiñones Huatangari et al. (2020) emplean una metodología Cross Industry Standard Process for Data Mining (CRISP-DM) planteando tres modelos que permiten estimar la deserción estudiantil.

Muzumdar et al. (2020) utilizan el análisis de componentes principales para estudiar los factores intrínsecos que apoyan a la permanencia universitaria, y luego determinan la influencia de cada factor en

la probabilidad de retención a través de un análisis de regresión múltiple. Hasan (2019) compara el análisis de regresión logística ARL y el análisis discriminante lineal ADL determinando que el ADL puede ser más apropiado para técnicas de clasificación, sin embargo, la selección del modelo más adecuado depende del lugar y características en donde se aplique.

En cuanto al aseguramiento de la calidad en las Instituciones de Educación Superior, el Modelo Genérico de Evaluación del entorno de aprendizaje de carreras en Ecuador se calcula en los dos primeros años de estudio y recomienda que la tasa de deserción estudiantil no sea mayor al 20%. En consecuencia, la visión de crecimiento institucional no debe centrarse únicamente aumentar el número de matriculados, sino que requiere de estrategias de acción que garanticen la permanencia de los estudiantes en la institución hasta que finalicen sus estudios con éxito (Pérez et al., 2018).

El objetivo de este estudio es analizar el comportamiento de la deserción estudiantil en la Universidad Politécnica Estatal del Carchi (UPEC), utilizando información institucional y aplicando un modelo de regresión logística con el propósito de identificar los factores determinantes de esta problemática a fin de que las autoridades competentes puedan formular propuestas efectivas de retención estudiantil.

2. METODOLOGÍA O MATERIALES Y MÉTODOS

2.1 Población y muestra

La población de estudio hace referencia a 3821 estudiantes que registraron su matrícula en el primer nivel de carrera en la UPEC desde el segundo periodo del 2016 hasta el primer periodo del 2020 (8 periodos académicos). Toda la información se recogió de la base de datos institucional UPEC y sus características principales se detallan en la tabla 1.

Tabla 1

Características de la población

<i>Característica</i>	<i>Carrera</i>	<i>Estudiantes</i>	<i>Representación</i>
<i>Sexo</i>	Masculino	1577	41.27 %
	Femenino	2244	58.73 %
<i>Carrera</i>	Administración de Empresas	451	11.8 %
	Administración Pública	501	13.1 %
	Agropecuaria	403	10.6 %
	Alimentos	431	11.3 %
	Comercio Exterior	394	10.3 %
	Computación	394	10.3 %
	Enfermería	485	12.7 %
	Logística y Transporte	437	11.4 %
	Turismo	325	8.5 %
<i>Periodo académico</i>	Oct2016-Feb2017	516	13.5 %
	Abr2017-Ago2017	509	13.3 %
	Oct2017-Feb2018	505	13.2 %
	Abr2018-Ago2018	376	9.8 %
	Oct2018-Feb2019	447	11.7 %
	Abr2019-Ago2019	496	12.9 %
	Oct2019-Feb2020	466	11.2 %
	Abr2020-Ago2020	506	13.2 %

Además, la edad media de los estudiantes oscila entre 17 y 21 años, y en promedio el 31.3% de estudiantes abandonaron su carrera universitaria.

2.2 Procedimiento de recogida y análisis de datos

Con la base de datos obtenida se aplica una metodología experimental que consiste en cuatro etapas: preparación, preprocesamiento, clasificación y análisis de resultados (Findiana et al., 2020). La preparación

de la data consiste en comparar los estudiantes matriculados en el sistema en un periodo académico determinado con los estudiantes que aún continúan realizando el proceso de matrícula dos años o cuatro periodos después, de manera ideal un estudiante que se matricula en primer nivel, dos años después debería matricularse en quinto nivel.

Sánchez-Pozo et al. (2021) realizan un preprocesamiento de los datos que consiste en eliminar duplicados e información irrelevante, además de una integración, limpieza, transformación y discretización. La base cuenta con 13 variables de interés como se muestra en la tabla 2. Se agrega una variable “Estado” para poder clasificarlo según el modelo de clasificación supervisado.

El modelo de clasificación lineal debe permitir identificar a los estudiantes con ciertas características particulares quienes posiblemente pueden abandonar sus estudios durante los dos primeros años y asignarlos en una de las dos categorías de la variable “Estado”. El análisis lineal discriminante LDA o la regresión logística LR se pueden utilizar como técnica de modelado estadístico multivariante, cuyas posibles variables de salida son de tipo categórico (Pohar et al., 2004).

No todas las variables disponibles en la base de datos aportarán para predecir o clasificar a la variable de interés. Liong y Foo (2013) afirmaron que la técnica LR es más robusta y además se ajusta bien a cualquier distribución de datos, incluso cuando no se cumplan con los supuestos de normalidad, homogeneidad y homocedasticidad.

La evaluación del rendimiento del modelo se realiza mediante la exactitud, sensibilidad, especificidad, curva ROC; y se miden a partir de una matriz de confusión. Estas métricas se calculan utilizando una validación cruzada, que es un método de comprobación del rendimiento de los modelos, divide los datos en dos partes: la primera para entrenar el modelo y la segunda para comprobar el modelo (Nuankaew et al., 2020).

La precisión del modelo se expresa matemáticamente como una relación del número correcto de casos predichos y el total de predicciones y se calcula mediante la siguiente expresión:

$$\text{Precisión} = \frac{Vp + Vn}{Vp + Vn + Fp + Fn}$$

La sensibilidad indica la probabilidad de clasificar las predicciones verdaderas positivas considerando también los falsos negativos, se determina mediante la expresión:

$$\text{Sensibilidad} = \frac{Vp}{Vp + Fn}$$

La especificidad calcula la probabilidad de clasificar las predicciones verdaderas negativas considerando también los falsos positivos, se obtiene mediante la aplicación de la siguiente expresión matemática:

$$\text{Especificidad} = \frac{Vn}{Vn + Fp}$$

La curva ROC se construye a partir de la sensibilidad vs el complemento de la especificidad (1-especificidad), el área bajo la curva se utiliza para comparar el modelo de clasificación.

Tabla 2
Preprocesamiento de datos

<i>N</i>	<i>Variable</i>	<i>Descripción</i>	<i>Valor</i>
1	Carrera	Carreras de la UPEC	1: Administración de Empresas 2: Administración Pública 3: Agropecuaria 4: Alimentos 5: Comercio Exterior 6: Computación 7: Enfermería 8: Logística y Transporte 9: Turismo
2	Sexo	Sexo	1: Masculino 2: Femenino
3	Edad	Edad al ingreso a la universidad	-
4	Discapacidad	Algún tipo de discapacidad	0: No tiene 1: Si tiene
5	Etnia	Tipo de etnia	1: Afroecuatoriana 2: Indígena 3: Mestizo 4: Blanco 5: Montubio
6	Residencia	Zona de planificación en Ecuador según la provincia de residencia	1: Esmeraldas, Imbabura, Carchi, Sucumbíos 2: Pichincha, Napo, Orellana 3: Cotopaxi, Tungurahua, Chimborazo, Pastaza 4: Manabí, Santo Domingo 5: Santa Elena, Guayas, Bolívar, Los Ríos, Galápagos 6: Cañar, Azuay, Morona Santiago 7: El Oro, Loja, Zamora Chinchipe 10: Extranjeros
7	N_padre	Nivel de formación académica del padre	0: ninguno, jardín de infantes 1: primaria, secundaria, básica 2: media, universidad incompleta 3: tercer nivel 4: cuarto nivel
8	N_madre	Nivel de formación académica del padre	0: ninguno, jardín de infantes 1: primaria, secundaria, básica 2: media, universidad incompleta 3: tercer nivel 4: cuarto nivel
9	M_hogar	Número de miembros en el hogar	-
10	T_colegio	Tipo de colegio que realizó sus estudios	1: Privado 2: Fiscomisional 3: Público
11	Nivel	Nivel en que registra su última matrícula	1: primero 2: segundo 3: tercero 4: cuarto 5: quinto
12	Promedio	Promedio de todo su récord académico	-
13	Estado	Estado de continuidad en la universidad dos años después de ingresar a la universidad	0: deserción 1: permanencia

3. RESULTADOS Y DISCUSIÓN

Se analizan en total a 3821 estudiantes quienes han registrado su matrícula en el primer nivel de una de las nueve carreras de la UPEC a partir del 2016, y la tasa de abandono estudiantil corresponde al 31.3%. Considerando que Tomás y Gutiérrez (2019) asocian a los factores de deserción en seis grupos para estudiar la influencia de cada variable predictor, en este estudio las variables disponibles se agrupan en tres grupos o factores: factores individuales, factores familiares y factores académicos, como se muestra en la tabla 3.

Tabla 3
Análisis estadístico de factores definidos

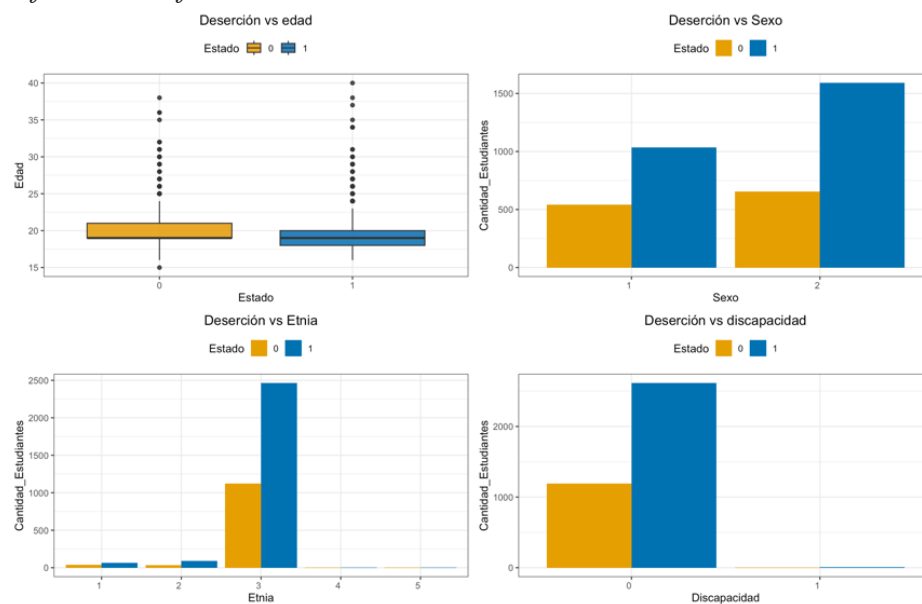
Tipo de factor	Variable	Prueba estadística	Valor p	Diferencia significativa
Factores individuales	Edad	Prueba de Wilcoxon	1.021e-13	Si
	Sexo	test de chi-cuadrado	0.0005943	Si
	Etnia	test exacto de Fisher	0.3926	No
	Discapacidad	test de chi-cuadrado	1	No
Factores familiares	Residencia	Prueba exacta de Fisher	0.04083	Si
	N_padre	Prueba chi-cuadrado	2.2e-16	Si
	N_madre	Prueba chi-cuadrado	2.2e-16	Si
	M_hogar	Prueba exacta de Fisher	0.00049	Si
Factores académicos	Carrera	Prueba chi-cuadrado	2.2e-16	Si
	T_colegio	Prueba chi-cuadrado	0.6764	No
	Nivel	Prueba exacta de Fisher	2.2e-16	Si
	Promedio	Prueba de Wilcoxon	2.2e-16	Si

Nota: *Elaboración propia con RStudio*

El factor individual hace referencia a los datos demográficos de cada estudiante, la figura 1 muestra una ligera diferencia en la edad de los estudiantes que continúan en sus estudios en comparación con aquellos que abandonan. Sin embargo, no se logra apreciar una diferencia aparente en función del género. La mayoría de los estudiantes que participaron en la investigación pertenecen a la etnia mestiza, siendo las demás etnias minoritarias y no susceptibles de comparación.

Los resultados sugieren que la etnia y la discapacidad no parecen ser factores determinantes en la decisión de abandonar los estudios. En la tabla 5 se presentan los análisis estadísticos correspondientes a cada variable para establecer si existen diferencias significativas de forma analítica.

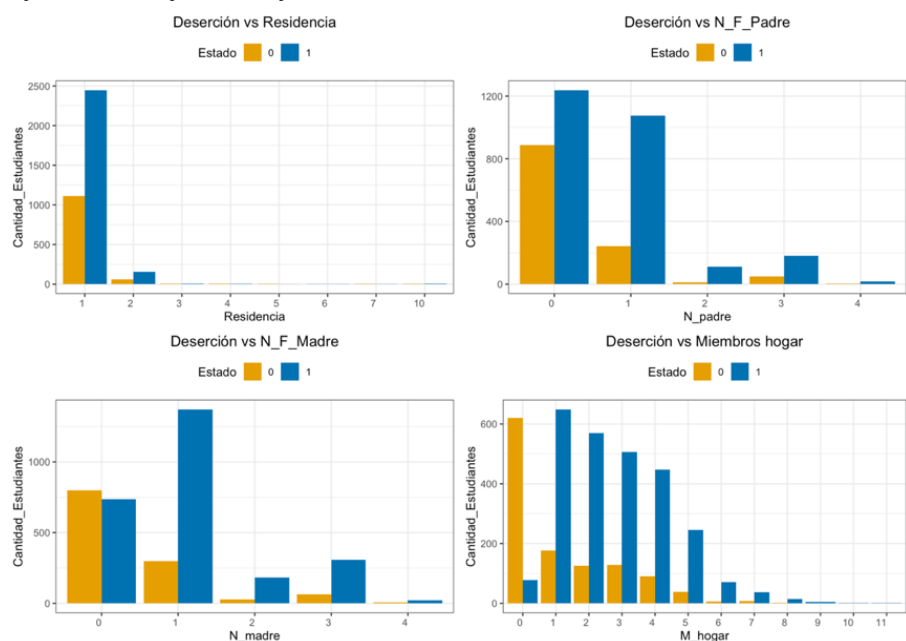
Figura 1
Influencia de factores individuales en el abandono estudiantil.



Se observa que el 93% de los estudiantes se concentran en la Zona de Planificación 1, donde se encuentra ubicada la universidad, por lo que las demás zonas no presentan representatividad para poder llevar a cabo alguna comparación significativa. Además, se ha evidenciado que el nivel educativo de los padres es un factor determinante en la deserción estudiantil, siendo más elevado el número de estudiantes que abandonan la universidad cuando el nivel de formación académica de los padres es bajo. Asimismo, se ha podido evidenciar que la tasa de abandono disminuye conforme aumenta la cantidad de integrantes en el hogar, siendo esta problemática más evidente cuando el estudiante vive solo.

En la figura 2 se muestra el comportamiento de las variables que pertenecen a factores familiares del estudiante que hacen referencia a nivel socioeconómica de su entorno familiar.

Figura 2
Influencia de factores familiares en el abandono estudiantil.



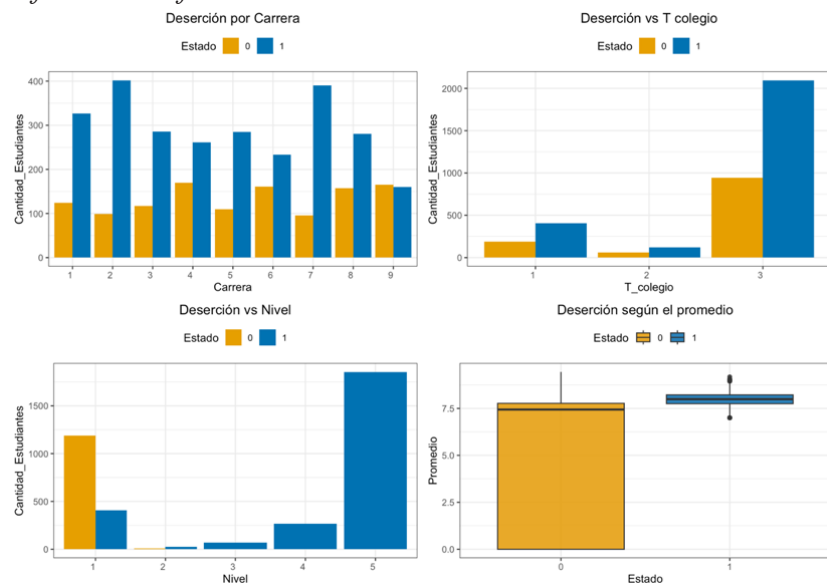
Cada una de las variables del factor familiar presenta una diferencia estadísticamente significativa en lo que respecta al abandono estudiantil, razón por la cual resulta imperativo llevar a cabo un análisis de correlación con el fin de prevenir la posible presencia de colinealidad en el modelo.

En la Figura 3 se ha identificado que la problemática de la deserción estudiantil no es homogénea en todas las carreras vigentes en la universidad, siendo la carrera de Turismo aquella en la que se ha registrado una tasa más elevada de abandono en comparación con aquellos estudiantes que han permanecido en sus estudios, mientras que, en las carreras de Enfermería y Administración Pública, la tasa de deserción ha sido menor.

En cuanto al tipo de colegio de procedencia, se ha constatado que la mayoría de los estudiantes provienen de colegios públicos, por lo que no se puede inferir que este factor sea determinante en la descripción de la deserción estudiantil. Siguiendo la línea de investigación de Nuankaew et al. (2020) se ha podido observar que el primer año de estudios es el período en el que se presenta la mayor tasa de abandono, y en este caso en particular, el fenómeno se produce en el primer nivel de carrera.

Figura 3

Influencia de factores académicos en el abandono estudiantil.



Finalmente, se ha constatado que los estudiantes que abandonaron la universidad presentaron un promedio de rendimiento académico inferior en comparación con aquellos que permanecieron en sus estudios. La tabla 3 muestra su análisis cuantitativo.

Del análisis estadístico presentado, se puede apreciar que con excepción de la variable que indica el tipo de colegio de procedencia (particular, privado o mixto), se ha encontrado una relación significativa del 95% entre la deserción universitaria y la carrera de estudio, el nivel de estudios y el promedio académico, lo cual indica que estos factores pueden ser importantes predictores de la deserción estudiantil.

Estos resultados sugieren la necesidad de implementar medidas específicas para prevenir y reducir la deserción en los estudiantes que se encuentran en situaciones de mayor riesgo, como aquellos que cursan ciertas carreras, se encuentran en niveles académicos específicos o presentan un bajo desempeño académico.

Para evitar tener información redundante en el modelo de este estudio se emplearon dos medidas de correlación no paramétricas para analizar la asociación monótonica entre las variables. La correlación de Spearman y la correlación de Kendall, para evaluar la relación entre dos variables. La correlación de Spearman fue aplicada para datos ordinales e intervalos, mientras que la correlación de Kendall fue adecuada para datos nominales, ordinales e intervalos. Se observó que ambas medidas de correlación proporcionaron resultados similares y consistentes como se muestra en la figura 4.

De la figura 4 se sugiere que existe una correlación entre el número de miembros del hogar y el nivel de formación de los padres. Asimismo, se ha evidenciado que el nivel de preparación académica de los padres presenta una relación con la variable estado. En general, se ha encontrado que no existe una relación lineal significativa entre todas las variables, por lo que no es posible descartar alguna variable para su inclusión en el modelo, dado que ninguna variable explica por sí sola el comportamiento de las demás.

Se propone tres modelos de regresión logística; el primero incluye todas las variables disponibles y se utiliza como modelo teórico. El segundo se ajusta mediante el método de selección de variables

backward. Finalmente, el tercero se propone a partir de los dos anteriores, con la premisa que la carrera y el nivel de formación de los padres pueden aportar con información relevante. En la tabla 4 se describen las variables utilizadas para cada modelo.

Figura 4
Prueba de correlación de Spearman

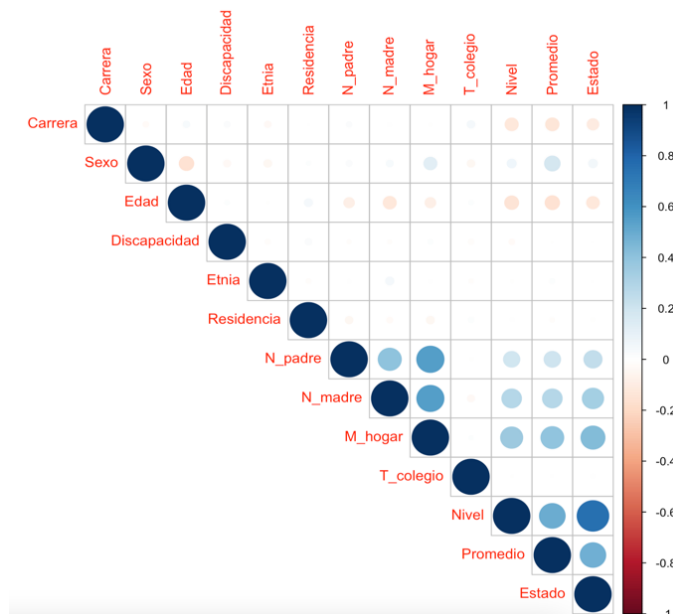


Tabla 4
Modelos de regresión logística propuestos

Modelo	Variables
Modelo_1	glm(Estado ~ Carrera+ Sexo + Edad + Discapacidad + Etnia+ Residencia + N_padre + N_madre + M_hogar + T_colegio + Nivel + Promedio)
Modelo_2	glm(Estado ~ Edad + M_hogar + Nivel + Promedio)
Modelo_3	glm(Estado ~ Carrera + Edad + N_padre + N_madre + M_hogar + Nivel)

Estos modelos buscan identificar qué variables son las más importantes para predecir la deserción estudiantil y contribuir a la implementación de estrategias efectivas para reducir este fenómeno en la universidad, la tabla 5 presenta los valores de cuatro medidas de evaluación de tres modelos de regresión logística.

Tabla 5
Medidas de evaluación del modelo

Tipo de factor	AIC	Precisión	Sensibilidad	Especificidad
Modelo_1	1280.1	0.9032	0.9375	0.8878
Modelo_2	1249.5	0.8963	0.9417	0.8760
Modelo_3	1257.8	0.9022	0.9364	0.8869

Se observa que el modelo 2 presenta el valor de AIC más bajo, lo que indica que es el que mejor ajuste presenta a los datos disponibles. Los valores de precisión obtenidos se encuentran dentro de un rango aceptable en los tres modelos analizados (0.8963 - 0.9032), lo que indica que todos ellos tienen un desempeño similar en términos de la cantidad de predicciones correctas que realizan. Es importante destacar que, aunque los valores obtenidos para la precisión son similares, pueden existir otras métricas que permitan una evaluación más exhaustiva de los modelos y que sean relevantes para el problema en cuestión.

En cuanto a la Sensibilidad y Especificidad, estas métricas reflejan la capacidad que tiene el modelo para identificar correctamente a estudiantes que abandonarán la universidad (Sensibilidad) y a los que no

lo harán (Especificidad). Se observa que los tres modelos presentan una Sensibilidad adecuada, con valores entre 0.9364 y 0.9417, lo que indica que son capaces de identificar correctamente a la mayoría de los estudiantes que abandonarán sus estudios.

La Especificidad presenta valores más bajos, con un rango entre 0.8760 y 0.8878, lo que indica que los modelos presentan un mayor porcentaje de falsos positivos, es decir, estudiantes que son clasificados como propensos a abandonar, pero que en realidad no lo hacen. Estos resultados indican que el modelo 2 es el más adecuado para predecir la deserción estudiantil en la UPEC y podría ser utilizado por las autoridades académicas para implementar medidas preventivas y de apoyo a los estudiantes en riesgo de abandonar sus estudios.

Se utiliza la matriz de confusión para evaluar la precisión de los modelos predictivos empleados en la identificación de estudiantes que podrían abandonar la universidad durante los dos primeros años de carrera. Los resultados obtenidos, presentados en la tabla 6, indican que el modelo 2 tiene una precisión aceptable en la predicción de ambos grupos de estudiantes, con una tasa de aciertos que varía entre 0.8963 y 0.9032.

Tabla 6

Matriz de confusión del Modelo 2

<i>Tipo de factor</i>	<i>Datos de entrenamiento</i>		<i>Datos de prueba</i>	
	Positivo	Negativo	Positivo	Negativo
<i>Puede abandonar la universidad durante los dos primeros años de estudio</i>	889	262	243	70
<i>Puede permanecer la universidad durante los dos primeros años de estudio</i>	55	1851	10	441

Es importante destacar que se utilizaron el conjunto de datos de entrenamiento y los datos de prueba en la construcción y evaluación del modelo, lo que asegura una validación adecuada de su capacidad predictiva en situaciones reales. En conjunto, estos hallazgos contribuyen al desarrollo de estrategias efectivas para prevenir el abandono universitario y mejorar la retención estudiantil.

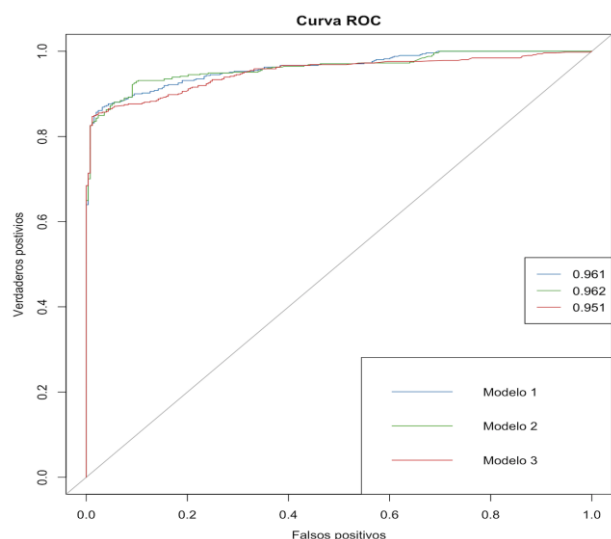
Los resultados de esta investigación indican que la etnia de los estudiantes, la discapacidad y el tipo de colegio del que provienen no influyen en la deserción universitaria, pero se observa una relación lineal inversa entre la deserción estudiantil y el nivel de formación del padre, el de la madre y el número de miembros en el hogar, a mayor nivel de formación de los padres menor es la cantidad de estudiantes que abandonan la universidad.

Finalmente, se evalúa la capacidad predictiva de los tres modelos para predecir la deserción estudiantil en una población de estudiantes universitarios mediante el uso de una curva ROC. La figura 5 presenta los resultados obtenidos y los valores del área bajo la curva (AUC) para cada modelo. Los resultados indican que el Modelo_2 presentó un mejor desempeño en la predicción de la deserción estudiantil en comparación con los otros dos modelos con un AUC de 0.962.

El análisis de variables relacionadas con el fenómeno de la deserción estudiantil es un objetivo compartido por los profesionales de la educación superior. Revisión de la documentación científica indica que los elementos determinantes de la deserción varían según la realidad de cada universidad. En este

estudio, se encontró que los motivos principales que provocan la deserción de los estudiantes de la UPEC (edad, familiar, rendimiento académico) son distintos de los factores encontrados en otras universidades.

Figura 5
Curva de la Característica Operativa del Receptor (ROC)



Poveda (2019) coincide con Muzumdar et al. (2020) en los aspectos determinantes de la deserción los factores: académico, económico, familiar, motivacional, salud y social. Bedregal-Alpaca et al., (2020) estudian el abandono universitario únicamente con aspectos de rendimiento académico y recomiendan estudiar los aspectos socioeconómicos, Por lo tanto, es conveniente considerar la realidad específica de cada institución al abordar el problema de la deserción estudiantil y desarrollar soluciones personalizadas para cada caso.

La predicción o clasificación de poblaciones o grupos a partir de variables predictoras se puede realizar mediante modelos LDA o LRA (Liong y Foo, 2013). Generalmente, estos modelos pueden comportarse de manera similar, y su aplicación depende lugar y supuestos aplicados (Hasan, 2019). En este estudio se compararon diferentes modelos de regresión lineal variando el número de variables predictivas. Niyogisubizo et al., (2022) proponen que el indicador de evaluación del rendimiento de un modelo de predicción es el AUC.

El modelo 2 es el más adecuado para predecir la deserción estudiantil en esta población de estudiantes universitarios, lo que sugiere que podría ser utilizado como una herramienta efectiva para prevenir la deserción y mejorar la retención estudiantil con una precisión del 90%. Estos hallazgos son particularmente relevantes en el contexto actual de la educación superior, donde el abandono universitario representa un problema significativo en muchos países.

El sistema de calidad de Educación Superior en el Ecuador sugiere medir la deserción estudiantil en los dos primeros años de carrera; Pierella et al. (2020) y Ferreyra et al., (2017) demuestran que el abandono universitario se concentra en el primer año de estudio. En la UPEC el abandono ocurre principalmente durante el primer semestre de carrera. Este hallazgo permite identificar un espacio oportuno para intervenir y aplicar medidas preventivas y de apoyo dirigidas a los estudiantes en riesgo de desertar en sus estudios, lo que puede contribuir a mejorar la tasa de retención y graduación en la universidad.

4. CONCLUSIONES

Este estudio evidenció que aunque el modelo de evaluación institucional mide la retención estudiantil en los dos primeros años de estudio, en la UPEC la deserción de estudiantes más significativa se concentra en el primer nivel de carreras.

El modelo logístico propuesto demuestra una alta precisión del 90%, que los factores principales que influyen en la deserción estudiantil en la UPEC son: la edad, el número de miembros en el hogar, el nivel de estudio y el promedio académico. Estos factores podrían ser útiles para el desarrollo de políticas y estrategias efectivas para reducir la tasa de abandono estudiantil. Sin embargo, se recomienda que se realicen futuras investigaciones cualitativas para profundizar en las experiencias de los estudiantes que experimentan la intención de desertar y complementar este trabajo, ya que se basó únicamente en datos cuantitativos disponibles en la base de datos de la institución.

En conclusión, este estudio destaca la importancia de contar con modelos predictivos precisos para identificar a tiempo a los alumnos universitarios en riesgo de deserción. Estas investigaciones proporcionan una visión más completa de las causas del abandono y permiten diseñar políticas y estrategias más efectivas para prevenirlo. Es decir, disponer de modelos predictivos precisos es fundamental para abordar el problema del abandono estudiantil de manera efectiva.

FINANCIACIÓN

Los autores no recibieron financiación para el desarrollo de la presente investigación.

CONFLICTO DE INTERESES

Los Autores declaran que no existe conflicto de intereses

CONTRIBUCIÓN DE AUTORÍA

En concordancia con la taxonomía establecida internacionalmente para la asignación de créditos a autores de artículos científicos (<https://credit.niso.org/>). Los autores declaran sus contribuciones en la siguiente matriz:

	Montalvo F.	Sánchez N.
<i>Participar activamente en:</i>		
Conceptualización	X	X
Análisis formal	X	X
Adquisición de fondos	X	X
Investigación	X	X
Metodología	X	X
Administración del proyecto	X	X
Recursos	X	X
Redacción –borrador original	X	X
Redacción –revisión y edición	X	X
La discusión de los resultados	X	X
Revisión y aprobación de la versión final del trabajo.	X	X

RECONOCIMIENTO A REVISORES:

La revista reconoce el tiempo y esfuerzo del editor Raúl Lozada Yáñez, y de revisores anónimos que dedicaron su tiempo y esfuerzo en la evaluación y mejoramiento del presente artículo.

REFERENCIAS

- Álvarez, N. L., Callejas, Z., & Griol, D. (2020). Predicting Computer Engineering Students' Dropout In Cuban Higher Education With Pre-Enrollment and Early Performance Data. *Journal of Technology and Science Education*, 10(2), 241–258. <https://doi.org/10.3926/jotse.922>
- Bedregal-Alpaca, N., Tupacyupanqui-Jaén, D., & Cornejo-Aparicio, V. (2020). Análisis del rendimiento académico de los estudiantes de Ingeniería de Sistemas, posibilidades de deserción y propuestas para su retención Analysis of the academic performance of Systems Engineering students, desertion possibilities and proposals for retention. In *Revista chilena de ingeniería* (Vol. 28, Issue 4). <https://doi.org/10.4067/S0718-33052020000400668>
- Caselli Gismondi, H. E., & Urrelo Huiman, L. V. (2021). Características para un modelo de predicción de la deserción académica universitaria. Caso Universidad Nacional de Santa. *Llamkasun*, 2(4), 02–22. <https://doi.org/10.47797/llamkasun.v2i4.61>
- de Oliveira, C. F., Sobral, S. R., Ferreira, M. J., & Moreira, F. (2021). How does learning analytics contribute to prevent students' dropout in higher education: A systematic literature review. *Big Data and Cognitive Computing*, 5(4). <https://doi.org/10.3390/bdcc5040064>
- Ferreira, M., Avitabile, C., Botero Álvarez, J., Haimovich Paz, F., & Urzúa, S. (2017). *At a Crossroads Higher Education in Latin America and the Caribbean Human Development*. <https://doi.org/10.1596/978-1-4648-1014-5>
- Findiana, R., Yuniarno, E. M., & Endroyono. (2020). Classification of Graduates Student on Entrance Selection Public Higher Education through Report Card Grade Path Using Support Vector Machine Method. *2020 3rd International Conference on Information and Communications Technology, ICOIACT 2020*, 7–11. <https://doi.org/10.1109/ICOIACT50329.2020.9332072>
- Hasan, M. N. (2019). A comparison of logistic regression and linear discriminant analysis in predicting of female students attrition from school in Bangladesh. *N 2019 4th International Conference on Electrical Information and Communication Technology (EICT)*, 1–3. <https://doi.org/10.1109/EICT48899.2019.9068776>
- Kittinan, L., Kanda, T., & Tanasin, Y. (2018). Dropout Prediction System to Reduce Discontinue Study Rate of Information Technology Students. *5th International Conference on Business and Industrial Research (ICBIR)*, 110–114. <https://doi.org/10.1109/ICBIR.2018.8391176>
- Liong, C. Y., & Foo, S. F. (2013). Comparison of linear discriminant analysis and logistic regression for data classification. *AIP Conference Proceedings*, 1522, 1159–1165. <https://doi.org/10.1063/1.4801262>
- López-Aguilar, D., Álvarez-Pérez, P. R., & Ravelo-González, Y. (2022). Adaptability skills and the intention to drop out in university students. *Revista de Investigacion Educativa*, 40(1), 237–255. <https://doi.org/10.6018/rie.463811>
- Muzumdar, P., Basyal, G. P., & Vyas, P. (2020). Antecedents of Student Retention: A Predictive Modelling Approach. *International Journal of Contemporary Research and Review*, 11(11). <https://doi.org/10.15520/ijcrr.v11i11.860>

- Niyogisubizo, J., Liao, L., Nziyumva, E., Murwanashyaka, E., & Nshimyumukiza, P. C. (2022). Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization. *Computers and Education: Artificial Intelligence*, 3. <https://doi.org/10.1016/j.caeai.2022.100066>
- Nuankaew, P., Nuankaew, W., Phanniphong, K., Fooprateepsiri, R., & Bussaman, S. (2020). Analysis dropout situation of business computer students at University of Phayao. In *Advances in Intelligent Systems and Computing: Vol. 1134 AISC*. https://doi.org/10.1007/978-3-030-40274-7_42
- Pérez, A. M., Escobar, C. R., Toledo, M. R., Gutierrez, L. B., & Reyes, G. M. (2018). Prediction model of first-year student desertion at Universidad Bernardo O'Higgins (UBO). *Educacao e Pesquisa*, 44. <https://doi.org/10.1590/S1678-4634201844172094>
- Pierella, M. P., Peralta, N. S., & Pozzo, M. I. (2020). The first year of college. Working conditions, admission procedures and academic dropout from the teachers' perspective. *Revista Iberoamericana de Educacion Superior*, 11(31), 68–84. <https://doi.org/10.22201/iissue.20072872e.2020.31.706>
- Pohar, M., Blas, M., & Turk, S. (2004). Comparison of logistic regression and linear discriminant analysis. *Advances in Methodology and Statistics*, 1(1), 143–161. <https://doi.org/10.51936/ayrt6204>
- Poveda Velasco, I. M. (2019). Los factores que influyen sobre la deserción universitaria. Estudio en la UMRPSFXCh-Bolivia, análisis con ecuaciones estructurales. *Revista Investigación y Negocios*, 12(20), 63–80. http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S2521-27372019000200007&nrm=iso
- Quiñones Huatangari, L., Jara, D. M., Alvarado, N., Milla, M. E., & Gamarra, O. A. (2020). Modelo para la estimación de la deserción estudiantil Awajún y Wampis empleando minería de datos. *Argentine Journal of Science and Technology*, 34(1), 45–60. <https://doi.org/10.36995/j.recyt.2020.34.006>
- Ramírez, T., Díaz, R. B., & Salcedo, A. (2017). ¿Abandono o deserción estudiantil? Una necesaria discusión conceptual. *Investigación y Postgrado*, 32(1), 63–74. <https://dialnet.unirioja.es/servlet/articulo?codigo=6430685>
- Sanchez-Pozo, N. N., Mejia-Ordóñez, J. S., Chamorro, D. C., Mayorca-Torres, D., & Peluffo-Ordóñez, D. H. (2021). Predicting High School Students' Academic Performance: A Comparative Study of Supervised Machine Learning Techniques. *Future of Educational Innovation Workshop Series - Machine Learning-Driven Digital Technologies for Educational Innovation Workshop 2021*. <https://doi.org/10.1109/IEEECONF53024.2021.9733756>
- Suárez, A. A. G., Núñez, R. P., & Suárez, C. A. H. (2021). Deserción estudiantil en contextos vulnerables: comparativo entre dos ciudades fronterizas colombianas. *Revista Boletín Redipe*, 10(4), 178–189. <https://doi.org/10.36260/rbr.v10i4.1261>
- Tomás, J. M., & Gutiérrez, M. (2019). Contributions of the self-determination theory in predicting university students' academic satisfaction. In *Revista de Investigacion Educativa* (Vol. 37, Issue 2, pp. 471–485). Asociacion Interuniversitaria de Investigacion en Pedagogia. <https://doi.org/10.6018/rie.37.2.328191>
- Zárate-Valderrama, J., Bedregal-Alpaca, N., & Cornejo-Aparicio, V. (2021). Classification models to recognize patterns of desertion in university students. In *Revista chilena de ingeniería* (Vol. 29). <https://doi.org/10.4067/S0718-33052021000100168>.