

Factores incidentes de probabilidad de deserción estudiantil en un instituto tecnológico superior: un análisis estadístico

Incident factors in the probability of student dropout in a higher technological institute: a statistical analysis

Fatores incidentes na probabilidade de deserção de alunos em um instituto superior de tecnologia: uma análise estatística

Salazar-Siguenza, Daniela Michelle ¹
Universidad Politécnica del Carchi
daniela.salazar@upec.edu.ec
<https://orcid.org/0000-0002-4979-2067>



Valencia-Nuñez, Edison Roberto ²
Universidad Técnica de Ambato
edisonvalencia@uta.edu.ec
<https://orcid.org/0000-0003-2280-9129>



 DOI / URL: <https://doi.org/10.55813/gaea/ccri/v5/n1/371>

Como citar:

Salazar-Siguenza, D. M., & Valencia-Nuñez, E. R. (2024). Factores incidentes de probabilidad de deserción estudiantil en un instituto tecnológico superior: un análisis estadístico. *Código Científico Revista De Investigación*, 5(1), 59–86.

Recibido: 18/05/2024

Aceptado: 06/06/2024

Publicado: 30/06/2024

¹ Ingeniera en Finanzas con un Maestría en Diseño y Evaluación de Proyectos, actualmente docente de educación Superior.

² Máster universitario en estadística aplicada por la Universidad de Granada España, docente universitario a nivel de pregrado y posgrado desde el año 2008

Resumen

La deserción estudiantil se produce como un hecho global multicausal, en la totalidad instituciones de educación ya sean de carácter público, o particular, debido a factores socio económicos, sociales, individuales y académicos, con un abandono voluntario o forzado, de una manera temporal o definitiva del sistema de educación superior. Se desarrolló la investigación con el fin de determinar los factores incidentes de deserción vinculados a los estudiantes del Instituto Superior Tecnológico Nelson Torres en Cayambe. Se propuso un estudio con enfoque cuantitativo, con una investigación descriptiva, correlacional, y también explicativa, tomando una muestra estratificada de 260 estudiantes, y una encuesta estructurada de los distintos factores resultantes de la revisión teórica, se empleó herramientas correlacionales, y también de regresión logística, a través de un modelo evaluado mediante matriz de confusión y Curva ROC. Este estudio proporciona de forma más detallada las características asociadas a la deserción estudiantil, en el Instituto Nelson Torres en Cayambe, como son el rendimiento académico, las enfermedades catastróficas, el grado educativo del padre y gusto por la carrera, permitiendo el reconocimiento de estudiantes, con una probabilidad alta de abandono de estudios, con la consecuente implementación de planes para intervenir de manera adecuada y temprana. Los resultados también podrían ser utilizados por las instituciones educativas y los involucrados en la toma de decisiones para implementar políticas que contribuyan en la retención estudiantil, y mermar la tasa de deserción.

Palabras clave: Factores, Incidencia, Deserción, Educación Superior.

Abstract

Student desertion occurs as a global multicausal fact, in all educational institutions, whether public or private, due to socioeconomic, social, individual and academic factors, with a voluntary or forced abandonment, temporarily or permanently from the higher education system. The research was developed with the purpose of determining the incident factors of desertion linked to the students of the Instituto Superior Tecnológico Nelson Torres in Cayambe. A study with a quantitative approach was proposed, with a descriptive, correlational, and also explanatory research, taking a stratified sample of 260 students, and a structured survey of the different factors resulting from the theoretical review, correlational tools were used, and also logistic regression, through a model evaluated by confusion matrix and ROC Curve. This study provides more detailed characteristics associated with student dropout at the Nelson Torres Institute in Cayambe, such as academic performance, catastrophic illnesses, father's educational level and taste for the career, allowing the recognition of students with a high probability of dropping out, with the consequent implementation of plans to intervene appropriately and early. The results could also be used by educational institutions and decision makers to implement policies that contribute to student retention and reduce the dropout rate.

Keywords: Factors, Incidence, Dropout, Higher Education.

Resumo

A deserção estudantil ocorre como um fato global multicausal, em todas as instituições de ensino, sejam elas públicas ou privadas, devido a fatores socioeconômicos, sociais, individuais e acadêmicos, com abandono voluntário ou forçado, temporário ou definitivo, do sistema de ensino superior. A pesquisa foi realizada com o objetivo de determinar os fatores incidentes de deserção vinculados aos alunos do Instituto Superior Tecnológico Nelson Torres, em Cayambe. Foi proposto um estudo com abordagem quantitativa, com uma pesquisa descritiva, correlacional e também explicativa, tomando uma amostra estratificada de 260 alunos, e uma

pesquisa estruturada dos diversos fatores resultantes da revisão teórica, foram utilizadas ferramentas correlacionais e também regressão logística, por meio de um modelo avaliado por matriz de confusão e curva ROC. Este estudo fornece informações mais detalhadas sobre as características associadas à evasão escolar dos alunos do Instituto Nelson Torres em Cayambe, tais como desempenho acadêmico, doenças catastróficas, nível de escolaridade do pai e gosto pela carreira, permitindo o reconhecimento de alunos com alta probabilidade de evasão escolar, com a consequente implementação de planos para intervir de forma adequada e precoce. Os resultados também podem ser usados por instituições educacionais e tomadores de decisão para implementar políticas que contribuam para a retenção de alunos e reduzam a taxa de evasão.

Palavras-chave: Fatores, Incidência, Desistência, Ensino Superior.

Introducción

A lo largo de los años, la deserción estudiantil se ha producido en todas las instituciones educativas, producto de varios factores tanto endógenos como exógenos, lo que, afectado negativamente a los centros educacionales, a los alumnos, y a la colectividad educativa en todos sus estamentos. Esta problemática, se ha tratado de resolver hace un tiempo aproximado de 80 años (Cevallos, 2014). La deserción en la vida estudiantil en todos los niveles incluido el superior, es un fenómeno de gran complejidad que ha sido analizado desde diversas perspectivas a lo largo de décadas, sin embargo, a nivel de educación superior en sus instituciones, aún persiste convirtiéndose en un desafío no resuelto (Da Re & Clerici, 2017).

En el plano mundial, números determinan que de todos los alumnos que empiezan sus estudios a nivel superior solo el 50%, termina y se gradúa, de acuerdo con un estudio elaborado a nivel de América Latina (Ferreira et al., 2017). La deserción estudiantil se presenta como un hecho educativo en todos los países, incluso en los desarrollados, dentro del entorno latinoamericano, las estadísticas mencionan que solo 1 de 10 estudiantes, entre los 25 y 29 años han completado su formación profesional superior, según la Oficina Regional de Educación para América Latina y el Caribe (OREALC), (Lázaro et al., 2020).

Dentro del mismo entorno internacional, los autores Gutiérrez et al. (2021) muestran en el nivel superior universitario, estadísticas citadas en el informe Education at the glance (2016),

que en este tipo de educación la tasa de deserción en los países y regiones parte de la OCDE (Organización para la Cooperación y el Desarrollo Económicos) es del 31%, específicamente para los sectores Europeos la tasa de abandono varía entre el 20% y el 55%, y en el espacio latinoamericano, las tasas de abandono van desde el 8% hasta el 48%.

En Ecuador, de acuerdo con números presentados como estadísticas del CES (Consejo de Educación Superior del Ecuador), el porcentaje de deserción en el país es del 48%, entendiéndose que 48 de cada 100 estudiantes que logran ingresar en instituciones de formación superior, pueden cumplir el proceso de graduación acorde tiempos determinados para el propósito, sin el abandono de sus estudios (Lázaro et al., 2020). También, en Ecuador con base en cifras y estadísticas emitidas por el INEC en publicaciones, se registra un índice elevado de deserción estudiantil y un alto nivel de analfabetismo que ha empeorado provocando la inasistencia y el cierre de las aulas desde el 2020 hasta el 2022.

Es así que cifras muestran el 4.1% de población estudiantil se retiraron de sus casas de estudio en el 2022, además de existir una tasa de matriculación con tendencia a la disminución de un 3%, por otro lado es importante conocer que la tasa de matriculación en el nivel superior de la educación ecuatoriana es del 36%, sin embargo existe una deserción en la misma formación del 26% especialmente en los primeros años, además que la tasa presenta un incremento con respecto al 2015 de 6 puntos (INEC, 2022).

De acuerdo con estudios del INEC, 11 son las causas más relevantes de deserción en el Ecuador, entre las cuales se puede mencionar en orden de importancia, los recursos económicos escasos, poco interés por su formación y estudios, por cupos limitados, por cuidado de hijos, por embarazo, por enfermedad, discapacidad, por asistir a nivelación y por fracaso escolar con escasa orientación vocacional (INEC, 2022).

En este sentido de acuerdo con Tinto (1975), la deserción se produce cuando un estudiante no logra finalizar sus metas educativas, considerando desertor a aquel estudiante que

no muestra desempeño académico alguno en 3 meses, sin embargo, es importante considerar que existen varios comportamientos de deserción y que después de ese tiempo no es posible saber si el estudiante retornará o si optará por otro programa académico, surgiendo el concepto de primera deserción. En términos más actuales la deserción representa el abandono voluntario del estudiante de su formación, tanto por condicionantes internas o externas, no registrando su matrícula para estudiar el siguiente periodo académico (B. Rodríguez et al., 2018).

En cuanto a los factores usados como predictores de la deserción estudiantil, autores como Tinto (1975), y años posteriores Roland et al. (2018), han concluido que el retiro de un estudiante se explica mediante cinco dimensiones la psicológica, sociológica, económica, institucional y de interacción. Por otro lado, Díaz Peralta, (2008) y Lopera Oquendo, (2007) coinciden con varios autores en los cuatro macro factores presentados como predictores de la no permanencia de los estudiantes, encontrando los factores individuales, académicos, económicos e institucionales.

Para el tratamiento del abandono de estudios se beneficia significativamente de la aplicación de diversas técnicas y herramientas estadísticas que permiten analizar y comprender de manera más profunda la actuación de los estudiantes, entre las técnicas más utilizadas se encuentran el análisis descriptivo, que proporciona medidas resumidas como promedios, medianas y desviaciones estándar para caracterizar la distribución de datos. Asimismo, los análisis de correlación, el uso de regresiones logísticas junto a técnicas multivariadas puede ayudar a identificar factores predictivos probables que influyen en la decisión de continuar la formación o no, actualmente es de uso los estudios de supervivencia, la minería de datos e incluso la inteligencia artificial (Gutiérrez et al., 2015).

Por lo dicho, la problemática de la deserción estudiantil tiene una gran relevancia social, debido a que la educación en cualquier nivel es un factor que genera movilidad social y desarrollo personal mitigando las barreras y desigualdades que enfrentan determinados grupos

de estudiantes, como aquellos de bajos recursos económicos y condiciones socioeconómicas desfavorables, el estudio de la deserción puede contribuir en el trabajo de lograr, en cuanto a oportunidades educativas, la igualdad. La deserción escolar fomenta la exclusión y el marginamiento social, perjudicando a toda la colectividad (Venegas et al., 2017).

Por otro lado en un Institución de Educación Superior es importante el análisis del abandono de los estudios, puesto que en el Instituto cada periodo la deserción ha presentado un incremento observado en el Sistema SIGA (Sistema Integrado de Gestión Administrativa), lo que genera el cierre de aulas, siendo importante identificar qué factores en los estudiantes hacen que deserten de sus estudios en las diferentes carreras y niveles, comprendiendo a la vez la problemática que enfrentan los estudiantes incluyendo aquellas relacionados con la gestión educativa y su entorno, mejorando así la comprensión del objeto de estudio, para el diseño de estrategias, formulación de políticas y reformas, programas de retención y persistencia estudiantil, para disminuir y mitigar la evolución de la deserción, permitiendo la supervivencia de la institución con el manejo eficiente de recursos, considerando que la institución es de carácter público.

Dentro del aspecto económico es importante considerar, que con la deserción existe una inversión económica desperdiciada, tanto para los alumnos como para la institución educativa, y en general para la sociedad, por lo que al lograr incrementar, la retención estudiantil, la optimización de inversión en capital humano incluyendo la educación, se reduce el índice de abandono y aumenta la retención estudiantil, siendo una prioridad ineludible para los líderes gubernamentales, que la equidad mejore, y se mitigue a la vez exclusión en aspectos económicos y sociales, de las poblaciones desfavorecidas (Donggeun & Seoyong, 2018).

Resulta importante entonces el estudio del objetivo de determinar los factores específicos que están incidiendo para que la tasa de deserción estudiantil se eleve en el Instituto Tecnológico Nelson Torres por parte de los alumnos para contribuir a la formulación de

políticas en la gestión educativa que permitan la retención estudiantil, considerando que la problemática aún genera discusión, pues aún existe poco conocimiento sobre este tema en Latinoamérica y especialmente en Ecuador, por lo que en la esfera académica y de gestión, con la investigación se obtendrán aportes en las metodologías estadísticas que expliquen el comportamiento de los datos relacionados a la deserción estudiantil, con especificidad del entorno de la institución, con el uso de técnicas de regresión, análisis discriminante y dimensional.

Metodología

El enfoque que se empleó para la investigación fue de perspectiva cuantitativa, al tratarse específicamente de mediciones numéricas y análisis estadístico inferencial para establecer los diferentes factores determinantes en la deserción estudiantil en la IES establecida. Se trató de una investigación a realizarse de manera secuencial y probatoria, partiendo de una idea, que se delimitó con objetivos y preguntas.

Se realizó también en cuanto a variables propuestas tanto dependientes como independientes, su medición, en un determinado contexto, analizando a la vez las mediciones obtenidas, para llegar a conclusiones con referencia de las hipótesis, comprobando la realidad cuantitativa.

De acuerdo con el enfoque, el tipo de investigación que se propuso para el estudio en un primer punto fue descriptiva, para determinar la situación general de la deserción estudiantil, describiendo los datos y características de los diferentes factores y riesgos planteados, recolectados a través de encuestas e información disponible en la institución, en el SIGA (Sistema Integrado de Gestión Administrativa).

En otro aspecto la investigación fue explicativa, puesto que se ha buscado comprender las razones y los mecanismos subyacentes que contribuyen a la deserción en la institución,

profundizando en cómo y por qué ocurren los factores, a través del análisis de las relaciones causales entre las diversas variables propuestas a partir del marco teórico y definidas en la tabla 2, y como estas influyen para que un estudiante tome la decisión de retirarse de sus estudios.

La investigación también fue de tipo correlacional, determinando en qué grado se relacionan y asocian, las variables de deserción estudiantil y diferentes factores determinados en la hipótesis y marco teórico, en la muestra trabajada, con la aplicación de métodos estadísticos de regresión y longitudinales.

Respecto a la unidad de análisis propuesta para la investigación, la población estudiantil empleada fue la cohorte que cursa desde el 2021II al 2023I, de acuerdo con una muestra probabilística estratificada, a la que se aplicaron las encuestas con las diferentes perspectivas de factores.

El número de estudiantes correspondientes a la cohorte propuesta para el estudio es 785 distribuidos en las carreras de Administración, Diseño Gráfico y Desarrollo de Software oferta, con estudiantes de primero a cuarto nivel. Por lo que se tomó una muestra estratificada, considerando un 95% de confianza y un error permisible del 5%, con un total de 259 estudiantes a los que se aplicó la encuesta.

Tabla 1

Muestra estratificada

Carreras	Estudiantes	%	Muestra
Administración	515	66%	170
Diseño Gráfico	110	14%	36
Desarrollo Software	158	20%	52
Total	783		259

Nota: Autores (2024)

Para la recolección de datos, en base al enfoque metodológico propuesto y tipo de investigación, se usó una encuesta que incluyó los cuatro tipos de factores incidentes

(individuales, académicos, institucionales, socio económicos), que se aplicó mediante técnica de muestreo de estratos, a un total de 783 estudiantes.

Para determinar si y que características de factores económicos, institucionales, estudiantiles o académicos inciden en la evolución del retiro de estudios, en el Instituto Superior Tecnológico Nelson Torres, las variables usadas se detallan en la tabla 2, de acuerdo a la revisión bibliográfica y documental teórica realizada, determinado el enfoque, el modelo y factores individuales, académicos, institucionales, o socio económicos que expliquen las variables de abandono de los estudios en la institución.

Tabla 2

Variables y categorías usadas por factores, enfoques y modelos de la deserción estudiantil

Enfoque	Modelo	Variable y Tipo	Factor	Categorías
		Deserción (Dependiente)	Binario	0 (No deserta) 1 (Si deserta)
Interaccionista, y Sociológico	Spady, Tinto, Bean (1970, 1975, 1980)	(Independientes)		
		Relaciones Sociales	Individuales	Malas - (5) Excelentes
		Rendimiento académico	Académicos	Deficiente – (5) Excelente
		Antecedentes familiares (Nivel de estudio de la madre y padre, estado civil de padres)	Socio económicos	(1) Sin estudios, (2) Primaria, (3) Secundaria, (4) Superior, (5) Maestría, (6) Doctorado
		Problemas familiares	Socio económicos	(1) Nunca - (5) Siempre
		Motivación familiar	Socio económicos	(1) Nunca - (5) Siempre
		Gusto de carrera	Individuales	(1) No me gusta - (5) Me gusta mucho
		Financiamiento de estudios	Socio económicos	(1) Cuenta propia, (2) Padres, (3) Cónyuge, (4) Familiar, (5) Otros
		Género	Individuales	(1) Masculino,

		Edad	Individuales	(2) Femenino
		Estado Civil	Individuales	Numérica (1) Soltero, (2) Casado, (3) Divorciado, (4) Viudo, (5) Unido
		Enfermedades	Individuales	(1) Cayambe urbanas (2) Cayambe rurales (3) De otra ciudad o provincia
		Lugar de residencia	Individuales	(1) Cayambe urbanas (2) Cayambe rurales (3) De otra ciudad o provincia
		Condición laboral (situación, horario)	Socio económicos	(1) No trabaja - (5) Trabaja
		Gratuidad	Socio económicos	(0) No, (1) Sí
Psicológico	Ethington, (1990)	Nota de Ingreso	Académicos	(1) Deficiente - (5) Excelente
		Orientación Vocacional	Académicos	(0) No, (1) Sí
		Habilidades y técnicas de estudio	Académicos	(1) Nunca - (5) Siempre
		Tipo de plantel del que proviene	Individuales	(1) Particular, (2) Público, (3) Fiscal, (4) Fiscomisional, (5) Otros
Económico	Cabrera, (1992)	Subsidios y Becas	Socio económicos	(0) No, (1) Sí
		Dificultades económicas	Socio económicos	(1) Nunca - (5) Siempre
		Accesibilidad (Transporte)	Socio económicos	(1) Nunca - (5) Siempre
		Nivel de ingresos	Socio económicos	(1) Bajos - (5) Altos
Organizacional	Braxton, (1997)	Satisfacción	Institucionales	(1) Totalmente satisfecho - (5) Totalmente insatisfecho
		Calidad de docentes	Institucionales	(1) Deficiente - (5) Excelente
		Calidad de recursos	Institucionales	(1) Deficiente - (5) Excelente
		Calidad de contenidos académicos	Institucionales	(1) Deficiente - (5) Excelente
		Motivación institucional	Institucionales	(1) Nunca - (5) Siempre

Nota: Autores (2024)

La encuesta aplicada de forma aleatoria y directa a los estudiantes de la institución constó de un total de 31 ítems basados en la tabla 2, distribuidos de acuerdo con las 4 dimensiones identificadas 14 en los socios económicos, 6 académicos, 5 institucionales y 6 en el aspecto individual. La variable dependiente la deserción medida a través de la matrícula o no para el periodo, se mide a través de una variable binaria donde cero (0) representa la no deserción, y (1) la deserción, para las variables independientes existe una numérica la edad, y las demás se representan como variables categóricas con escala de, Likert, variables cualitativas y dicotómicas usando el uno o cero.

Para comprobar la confiabilidad del cuestionario estructurado a ser aplicado, el coeficiente de Alfa de Cronbach empleado en las preguntas métricas con escala del Likert, arrojó un valor de 0.72, afirmando que el instrumento es consistente, de acuerdo con los siguientes cálculos, y considerando el alfa estandarizada.

Tabla 3

Alfa de Cronbach

raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
0.69	0.72	0.74	0.26	2.5	0.026	3.6	0.52	0.25

Nota: Autores (2024)

Respecto al tratamiento y estudio de los datos, que se recolectaron, la información se organizó, y se transformó en base de datos, siendo procesada en el programa estadístico Rstudio, después de la limpieza de la base, determinación de valores perdidos, extremos e imputación, a través de herramientas de regresión y longitudinales estadísticas, se determinaron los factores incidentes.

Para el análisis en un primer momento se realizó un estudio descriptivo y exploratorio, posteriormente se realizó la correlación de las variables propuestas para este estudio, incluyendo pruebas de hipótesis respectivas para comprobación de supuestos, usando un

análisis entre cada variable independiente y la variable binaria deserción, a través de la prueba Chi cuadrada de Pearson, para obtener las variables que manifiesten asociación o dependencia a la deserción en el instituto. Como objetivo, para las variables, la prueba tuvo la contrastación del supuesto de independencia cuando la significación del estadístico sea menor o igual al 5%, siendo que las variables son independientes como hipótesis nula, por el contrario, como alternativa que estas guardan dependencia (Rodríguez y Mora, 2021).

Con las variables que resultaron tener una asociación, a través de la regresión logística se estableció un modelo que, con las variables incluidas, explique a la deserción como variable dependiente con las variables independientes de los factores (individuales, académicos, institucionales, socio económicos).

Se usó el modelo de regresión logística que predice la probabilidad, respecto a un evento binario, en función de una o más variables independientes. A diferencia de la regresión lineal, que se utiliza para predecir valores continuos, la regresión logística utiliza una función logística para transformar la salida en un rango de 0 a 1, siendo la función lineal el logaritmo de:

$$Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + \dots + B_kX_k \quad (1)$$

En donde B_0 hasta B_k , representan los coeficientes de regresión estimados a partir de X_0 hasta X_k como variables independientes, siendo B_0 el coeficiente de intercepción o constante, función que muestra una relación entre cada una de las variables predictoras y la dependiente en la escala logarítmica (Gujarati, 2010). Siendo la variable que depende (Y) la probabilidad de deserción.

Por lo que el modelo de regresión logística permitió estimar los coeficientes asociados con cada factor incidente, indicando el sentido y la medida de su impacto sobre la probabilidad del resultado de deserción. Entonces, la regresión logística, usada para estimar la probabilidad de las distintas cantidades de las variables experimentadas en el modelo, se expresa a través de la fórmula 2, siendo Y la función lineal y e la base de los logaritmos neperianos.

$$P\left(Y = \frac{1}{X}\right) = P \frac{1}{1+e^{-y}} = \frac{1}{1+e^{-(B_0+B_1X_1+B_2X_2+\dots+B_kX_k)}} \quad (2)$$

Se planteó el modelo inicial solo de una constante, luego con todas las variables significativas determinadas en el análisis bivariado, y finalmente de manera automática en el programa estadístico Rstudio con la función stepwise se determinó con la regresión logística, un modelo, procediendo a la selección de variables, aplicando el paso hacia adelante y hacia atrás, con un modelo logit inicial y un total usado como límite superior, con una dirección both, lo que significa que se permiten adiciones y eliminaciones de variables en el modelo, evaluando iterativamente diferentes combinaciones de variables para encontrar el modelo que minimiza la medida de información de Akaike (AIC).

Con el fin de examinar, el ajuste del modelo en su idoneidad y generalización, se usó la devianza residual y la devianza nula, así como el AIC, la devianza nula representa la discrepancia entre el modelo ajustado y un modelo nulo que asume que entre las variables predictoras, no hay ninguna asociación con la variable de respuesta, la devianza nula baja indica un mejor ajuste del modelo, por otro lado, la devianza residual mide la discrepancia entre el modelo ajustado y los datos observados, y una devianza residual pequeña indica que respecto a los datos, el modelo tiene un buen ajuste.

El AIC, por su parte, es un criterio de selección de modelos, que castiga la complejidad del modelo, seleccionando a los que logran adecuadamente el ajuste con un menor número de parámetros, este criterio se define como: $AIC = -2 (\log \text{verosimilitud}) + 2K$, donde, el log-likelihood (función verosimilitud) es una medida de la probabilidad de encontrar los valores, establecidos los datos poblacionales de un modelo estadístico, encontrando en qué medida modelo logístico concuerda con las observaciones, se obtiene tomando el logaritmo natural de la función de verosimilitud, un valor más alto indica un mejor ajuste, y K representa el número de parámetros, en general se selecciona el modelo con el AIC más bajo como el mejor (Bruce et al., 2022).

Con el modelo obtenido con mejor ajuste, se determinaron los efectos marginales que proporcionan información sobre cómo cada variable contribuye al cambio en la probabilidad de deserción en la institución, se ha considerado la dirección y magnitud de estos efectos, entendiendo como las variables predictoras se relacionan con la variable de respuesta en el contexto del modelo logit, ajustado por las otras variables.

Se evaluó además al modelo través de su capacidad de predicción, usando en un primer momento la matriz de confusión, para lo cual se dividió la data en dos partes de entrenamiento y prueba en un 80% y 20% respectivamente. En la matriz se determinó el número de predicciones acertadas y no acertadas hechas por el modelo en confrontación con los valores reales, dando como resultado cuatro elementos: verdaderos positivos abreviados por TP, falsos positivos o FP, verdaderos negativos identificados por TN y los falsos negativos o FN. A partir de esta matriz se valoró las métricas, de precisión, exactitud, así como de sensibilidad y también de especificidad.

La precisión corresponde a la parte de predicciones acertadas (TP + TN) sobre la totalidad de las predicciones, correspondiendo el cálculo a: $(TP + TN) / (TP + FP + TN + FN)$. Para la tasa de verdaderos positivos o sensibilidad, es la parte de casos positivos que el modelo identifica acertadamente, el cálculo es: $TP / (TP + FN)$. La especificidad es la parte de casos negativos que el modelo identifica exactamente, la fórmula corresponde a: $TN / (TN + FP)$. Y la exactitud es la parte de casos positivos correctamente estimados entre todos los casos identificados como positivos, siendo su cálculo: $TP / (TP + FP)$.

La curva ROC (Receiver Operating Characteristic), como un instrumento eficaz también permitió la evaluación del rendimiento del modelo y la cuantificación de su calidad, a partir de la sensibilidad en función a la diferencia de la especificidad (1-especificidad) para diferentes límites de clasificación, representando esta curva la relación entre la tasa de verdaderos positivos (TPR) y la tasa de falsos positivos (FPR), se calcula a partir del

entrenamiento del modelo obteniendo las probabilidades predichas para cada observación en los datos, luego se obtiene la tasa de verdaderos positivos (TPR) y la tasa de falsos positivos (FPR) para diferentes umbrales, uno más bajo clasificará más observaciones como positivas, lo que aumentará tanto el TPR como el FPR, un umbral más alto clasificará menos observaciones como positivas, lo que disminuirá tanto el TPR como el FPR, para graficar se coloca en el eje y el TPR, y en el eje x el FRP, cada punto en la curva ROC representa una combinación de TPR y FPR para un umbral de clasificación puntual, la diagonal de referencia (línea punteada) representa el rendimiento aleatorio, para una área del 0.50, cuanto más se aleje la curva ROC de la diagonal de referencia hacia el punto (0,1), mejor será el rendimiento del modelo (Irrizari, 2020). En general se hizo uso de análisis de correlación, y también de la regresión logística.

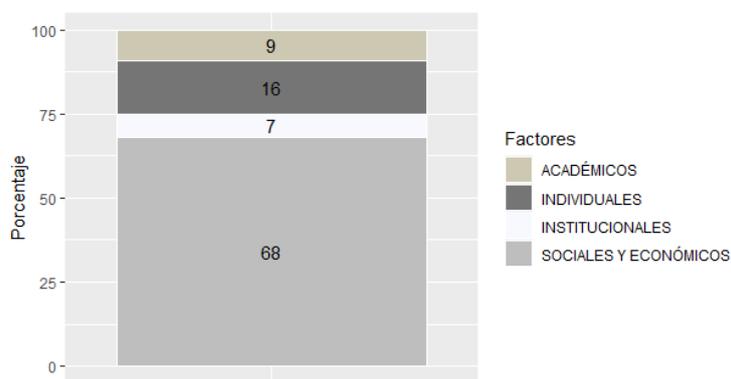
Resultados

Análisis Descriptivo

De acuerdo con un análisis de frecuencias, los estudiantes del Instituto Nelson Torres en general guardan una proporción parecida entre hombres y mujeres, tienen una concentración de edad entre los 19 y 21 años, más del 80% son solteros, una sexta parte proviene de Cayambe y sus parroquias aledañas, y los demás de otras provincias y ciudades, es importante distinguir que el 40% del estudiantado vive en áreas rurales. Con referencia a los factores encontrados, tanto socio económicos, individuales, académicos e institucionales en la población estudiantil de la institución, se presentan estadísticos descriptivos y posteriormente inferenciales, que permitan conocer la situación actual. La deserción encontrada para el periodo es del 10%, que considerando el total de estudiantes representa al menos 100 que abandonan sus estudios, en mayor proporción se presentan los factores socio económicos, seguido de los institucionales, y con una proporción menor los factores académicos, estando en último lugar los institucionales.

Figura 1

Percepción de factores de deserción Instituto Nelson Torres

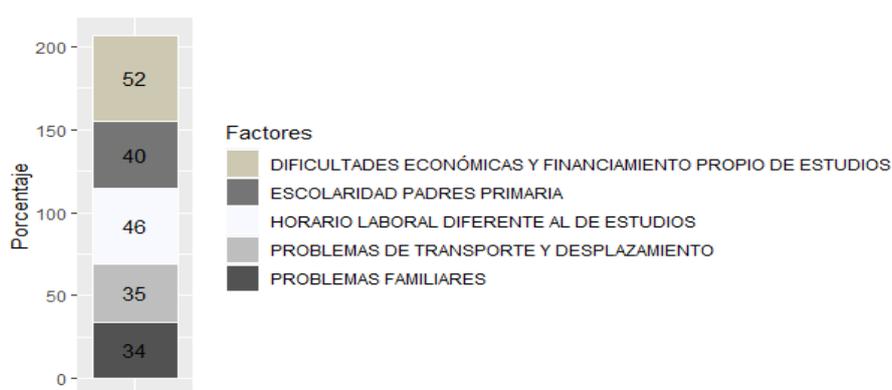


Nota: Autores (2024)

De acuerdo a la información encontrada en los factores socio económicos la mayoría presenta dificultades económicas y más de la cuarta parte se financia de manera autónoma los estudios, también en la misma proporción no existe compatibilidad del horario del trabajo actual con el horario de la institución educativa, por otro lado, está el factor de baja escolaridad de los padres, así como en menor proporción los problemas de transporte y desplazamiento.

Figura 2

Factores sociales y económicos de deserción Instituto Nelson Torres



Nota: Autores (2024)

En cuanto a los factores institucionales, los estudiantes consideran en una tercera y cuarta parte que la calidad buena y regular de los recursos de la IES, la calidad de los docentes, y de contenidos curriculares es de riesgo para la deserción, en una menor proporción la insatisfacción del servicio que percibe el estudiante, a través de todos los productos de oferta académica.

Figura 3

Factores institucionales de deserción Instituto Nelson Torres

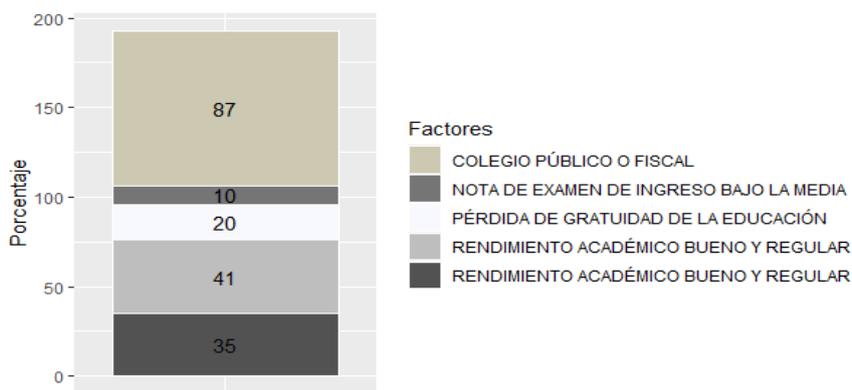


Nota: Autores (2024)

Referente a los factores académicos, más de la mayoría de los estudiantes provienen de un colegio público o fiscal, además un cuarto de los estudiantes tiene un rendimiento bueno y regular y rara la vez aplican métodos de estudios, en menor proporción constituye un factor de probabilidad la pérdida de gratuidad de le educación y la nota de ingreso no muy desarrollada de los postulantes.

Figura 4

Factores académicos de deserción Instituto Nelson Torres

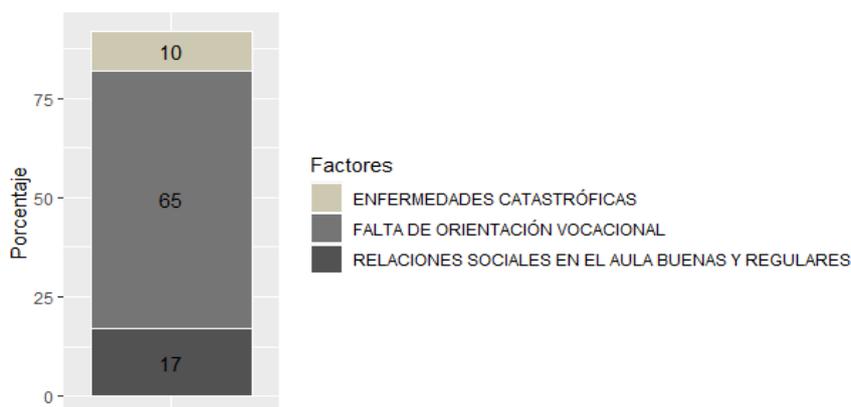


Nota: Autores (2024)

Dentro de los factores individuales representa un riesgo la falta de orientación vocacional en una sexta parte, y en una proporción menor las relaciones sociales en el aula buenas y regulares, así como el hecho de que los estudiantes tengan una enfermedad catastrófica, que genere una probabilidad de deserción de la institución.

Figura 5

Factores individuales de deserción Instituto Nelson Torres



Nota: Autores (2024)

Análisis Correlacional

Al determinar la correlación de las variables independientes de los distintos factores, con la variable deserción de respuesta, como una análisis bivariado, se establece que los factores que guardan asociación con la el abandono de estudios en el instituto, al cumplir con la prueba de hipótesis de dependencia, median el estadístico chi cuadrado, son el que le guste la carrera al estudiante, el hecho de que tenga alguna enfermedad catastrófica que interfiera en sus estudios, el nivel de escolaridad del jefe de familia, el apoyo y motivación que se percibe por parte de la familia, el rendimiento académico reflejado por el desempeño del estudiante, y la calidad de los docentes reflejada en la ejecución de la docencia, percibida por el estudiantado.

Tabla 4

Pearson's Chi-squared test variables significativas

Variable	Prueba Chi Cuadrado
Gusto Carrera	data: df\$varmatricula and df\$vargustocarrera X-squared = 10.704, df = 4, p-value = 0.03011
Motivación Familiar	data: df\$varmatricula and df\$varmotivacion X-squared = 9.368, df = 3, p-value = 0.02478
Rendimiento Académico	data: df\$varmatricula and df\$varendimeintocademico X-squared = 10.539, df = 4, p-value = 0.03226
Calidad Docentes	data: df\$varmatricula and df\$varcalidaddocentes X-squared = 7.7275, df = 3, p-value = 0.04199
Escolaridad Padre	data: df\$varmatricula and df\$varescolaridadpadre X-squared = 10.273, df = 5, p-value = 0.04787
Enfermedad Catastrófica	data: df\$varmatricula and df\$varenfermedadcastrostrof X-squared = 3.1606, df = 1, p-value = 0.07543

Nota: Autores (2024)

En el análisis exploratorio se observa, que aquellos estudiantes que tienen un menor gusto por la carrera, la matrícula es más baja, así mismo aquellos que reciben menor motivación familiar, se matriculan en menor cantidad, al igual que aquellos que no tiene un rendimiento académico excelente, la calidad de los docentes también se refleja en que aquellos que perciben una calidad baja se matriculan menos, por otro parte una menor escolaridad del padre, como

aquellos que tienen primaria, presentan un menor matrícula y finalmente el estudiante que posee una enfermedad catastrófica, la no matriculación es mayor.

Análisis Logit

En base a las variables que resultaron significativas se presenta un análisis de regresión logit, que permita establecer la relación entre la dependiente binaria de abandono o no, y las demás variables independientes sean cualitativas, cuantitativas, o categóricas, en este caso, gusto carrera, motivación familiar, rendimiento académico, calidad docente, escolaridad padre y enfermedad catastrófica.

Después de las iteraciones, mostrando la devianza residual, el AIC, para agregar o eliminar variables en cada paso se presenta el modelo más equilibrado, parsimonioso y mejor ajustado:

variablematricula \sim varenfermedadcastrotof + varendimeintocademico + varescolaridadpadre + vargustocarrera), el mismo que ha sido el que posee mejor ajuste, ha minimizado el AIC durante el procedimiento, este muestra variables significativas en la deserción estudiantil del Instituto Nelson Torres en Cayambe, en la cohorte de estudio, como el rendimiento o promedio académico, el hecho de sufrir una enfermedad catastrófica, el gusto por la carrera, y la escolaridad del padre, eliminando variables del modelo como la motivación y la calidad de los docentes que corresponden al factor institucional. La tabla 5 presenta un resumen del modelo seleccionado, mostrando los coeficientes estimados, errores estándar, valores t, valores p y otros estadísticos relacionados con el ajuste del modelo.

Tabla 5*Significancia de variables del modelo LOGIT*

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.02441	0.17177	-0.142	0.8871
varenfermedadcastrotof	0.17828	0.07803	2.285	0.0230 *
Varendimeintocademico	-0.06715	0.03399	-1.976	0.0491 *
varescolaridadpadre	0.10421	0.17108	0.609	0.5429
vargustocarrera	0.06096	0.04296	1.419	0.1570

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.08552036)

Null deviance: 27.920 on 311 degrees of freedom

Residual deviance: 25.827 on 302 degrees of freedom

(1 observation deleted due to missingness)

AIC: 130.05

Number of Fisher Scoring iterations: 2

Nota: Autores (2024)

El modelo se presenta de la siguiente manera, de acuerdo a los valores mostrados en la tabla 5.

$$Y = -0.02 + 0.18 \text{ Enfermedad Catastr\u00f3fica} - 0.07 \text{ Rendimiento Acad\u00e9mico} + 0.10 \text{ Escolaridad Padre} + 0.06 \text{ Gusto Carrera.}$$

Los coeficientes resultantes estimados para cada variable en el modelo, representan el cambio en el logaritmo de la odds (o probabilidad logar\u00edtmica) de la variable de respuesta, es decir en la desercion explicado por con un cambio de una unidad en la variable del predictor correspondiente, considerando constantes las otras variables, es as\u00ed que para una enfermedad catastr\u00f3fica, un estudiante que la tenga se asocia con un incremento de 0.18 veces m\u00e1s en la probabilidad de desercion del semestre. En el rendimiento acad\u00e9mico, un aumento de una unidad en esta variable, es decir un estudiante con mejor rendimiento acad\u00e9mico, se asocia con una disminuci\u00f3n de 0,07 veces en la probabilidad que deserte.

Dentro de los factores sociales y económicos se determina en cuanto al nivel de educación del padre, que tenga una educación primaria, existe un incremento de 0.10 veces más de posibilidad de que el estudiante se retire de la institución, finalmente el hecho que no le guste la carrera se asocia con una probabilidad de aumento en la decisión de deserción de 0,06 veces más.

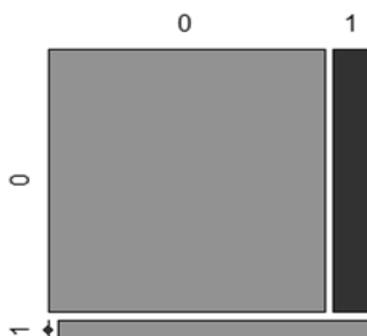
La variable deserción sigue una distribución Bernoulli, en la cual uno (1) denota el abandono de los estudios (no matriculación), a diferencia de cero (0) que significa que el estudiante realiza su matrícula (retención), en la tabla 5 se exponen las variables independientes evaluadas, por lo que usando un nivel de confianza del 95%, se determina que las variables rendimiento académico y la presencia de una enfermedad catastrófica, son significantes en el modelo, además la variables como el nivel educativo del padre y el interés por la carrera son relevantes en el contexto.

Ajuste, bondad y capacidad predictiva del modelo

Para medir el ajuste del modelo, en base a la devianza residual de 25.827, la devianza nula de 27.920 y del AIC como medida de penalización de la complejidad del modelo obtenido, con un valor de 130.05, siendo el más bajo, se determina el modelo es relativamente bueno en términos de equilibrio entre ajuste y complejidad. En cuanto a la capacidad de predicción del modelo en términos de predicciones correctas e incorrectas, según la matriz de confusión en la Figura 6 con la data de entrenamiento y test, hay 51 casos en los que tanto el modelo como los valores reales indican una clase negativa de no deserción siendo los verdaderos positivos, hay 4 casos en los que tanto el modelo como los valores reales indican una clase positiva (1) de deserción siendo los verdaderos negativos, hay 7 casos donde el modelo predijo una clase positiva (1), pero los valores reales indican una clase negativa (0), es decir falsos positivos, por último no hay casos donde el modelo predijo una clase negativa (0), pero los valores reales indican una clase positiva, por lo que no hay falsos negativos.

Figura 6

Matriz de confusión de clasificación del modelo de deserción

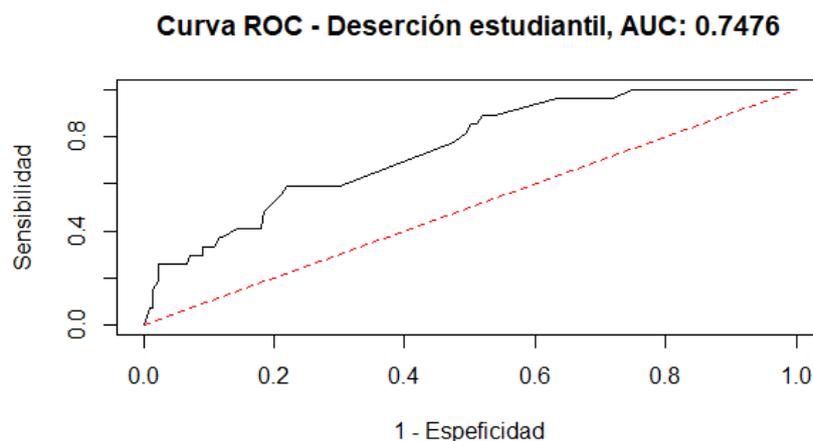


Nota: Autores (2024)

En términos generales el modelo parece tener un rendimiento moderado en la clasificación de casos, con una capacidad del 83%, de hacerlo de manera correcta, complementando con el uso de la curva ROC, el resultado final es un AUC de aproximadamente 0.75, siendo este un valor del modelo que muestra capacidad discriminativa, donde un modelo perfecto esta dado por el valor de uno, y un rendimiento aleatorio esta dado por 0.50, el valor encontrado para la investigación sugiere un rendimiento moderado en la clasificación de casos de deserción por el modelo, lo que se confirma en la Figura 7, en la cual se observa la tasa de falsos positivos en el eje x la tasa de verdaderos positivos en el eje y, mostrando cómo varía la TPR en función de la FPR de acuerdo al ajuste del umbral de clasificación, finalmente es curva se aleja de la diagonal de referencia y se acerca al vértice izquierdo, por lo que el modelo está clasificando correctamente una proporción considerable de casos positivos como positivos y una baja proporción de casos negativos como positivos.

Figura 7

Curva ROC del modelo de clasificación



Nota: Autores (2024)

Discusión

En cuanto los hallazgos de variables con una correlación favorable, se contrasta con Hernández Prado et al. (2016) que en un estudio de enfoque mixto, presentan que el abandono de los estudios, tiene origen en circunstancias económicas, familiares, falta de motivación, así como escasez de instituciones educacionales en las zonas circundantes, también se atribuye que la deserción estudiantil depende de otras condiciones como factores culturales, pudiendo tener una incidencia mayor en las zonas rurales y en poblaciones con identificaciones étnicas distintas. Sin embargo, en la investigación actual no se encontró relevante la variable de lugar y área de procedencia, así como factores culturales y de ruralidad.

Es así que, también en una investigación que combina métodos cuantitativos y cualitativos, los investigadores Lázaro Alvarez et al. (2020) identificaron los factores que provocan incidencia y predicen el retiro de estudiantes de nivel superior en la educación cubana, tomando a estudiantes de nuevo ingreso con una muestra de 485 a través del uso correlación, de regresión logística y análisis discriminante, técnicas usadas también en la

presente investigación, obteniendo factores similares como que la calidad educativa de la institución, el puntaje del test de ingreso y el promedio académico son variables que inciden en el abandono estudiantil. Sin embargo, cabe destacar que debido a la falta de disponibilidad de datos, no se pudieron considerar variables socio económicas y motivacionales en el análisis.

En otro estudio, que también se contrasta de acuerdo a las variables identificadas, Solís et al. (2022), después de investigar en una Universidad de Manabí, tomando una muestra de información existente del Sistema de Gestión Académica SIGA, se usa la regresión logística binaria para obtener variables predictoras significativas con pruebas de Wald, teniendo como resultados que el retiro de la educación se explica por variables como el gusto por la carrera elegida, ingresos, escolaridad del jefe de hogar, y el apoyo motivacional que el estudiante percibe en su entorno familiar.

Por otro lado, en el modelo final de mejor ajuste, se evidencia que el aspecto institucional no muestra una influencia estadísticamente significativa en el hecho de desertar para un estudiante, corroborando así los hallazgos propuestos por otros autores Viera et al. (2020) que realizaron estudios de determinantes de deserción en Perú, a estudiantes universitarios, con un enfoque cuantitativo utilizando técnicas estadísticas como el análisis factorial y la regresión logística, y considerando el modelo de interacción propuesto por Tinto (1982), llegando a una conclusión similar, de que los factores académicos, sociales, económicos e individuales guardan una asociación directa y significativa con el abandono de los estudios, también, encontraron que el grado de formación académica de los padres produce influencia en esta situación.

Respecto a los dos variables encontradas significativas, los autores Hernández et al. (s.f.), en un investigación documental afirman que un factor individual que puede afectar en el abandono estudiantil, son los problemas de salud, ya sean físicos o mentales, estos pueden impactar negativamente en el compromiso con los estudios, las enfermedades crónicas o

agudas, e incluso las de carácter mental pueden generar ausencias frecuentes, afectar la concentración y disminuir la energía necesaria para enfrentar las demandas académicas, y tener un impacto profundo en el bienestar emocional. En otros estudios de carácter cuantitativo, no se ha encontrado significativa estadísticamente la variable sin embargo en el Instituto Nelson Torres en el periodo estudiado las enfermedades catastróficas son un factor de riesgo a considerar para el abandono.

Por último, siendo el rendimiento académico otra variable significativa, varias son las investigaciones que lo encuentra como explicativo en el fenómeno de la deserción, como (Martínez et al., 2021), en su trabajo realizado en la educación cubana con cohortes de estudiantes de la carrera de medicina, aplican métodos de análisis bivariados para determinar las variables influyentes y luego un análisis multivariado a través de la regresión logística, mostrando una relación significativa de la deserción con el tiempo de estudio, notas de historial académico, la repetición de materias y un bajo nivel de rendimiento académico.

Conclusión

Como mejor método para determinar y encontrar los factores incidentes en la deserción estudiantil que se viene presentando en el Instituto Nelson Torres en la Ciudad de Cayambe, se ha usado herramientas de correlación, y un modelo logit, que permitió manejar las variables categóricas planteadas en la investigación de los factores individuales, económicos - sociales, académicos e institucionales, proporcionando estimaciones de probabilidad, para comprensión más profunda de los factores.

Se ha identificado las variables más significativas en la deserción en la institución, siendo estas el rendimiento académico, y el hecho de sufrir una enfermedad aguda, y en el contexto el nivel de escolaridad del padre y el gusto que el estudiante muestra por su carrera, por lo que los mencionados constituyen los principales factores para diseños de intervenciones

y políticas de retención y prevención de deserción en la institución, siendo estas variables pertenecientes a los factores académicos, económicos-sociales e individuales, por lo que no resultan explicativos y significantes para el instituto los factores institucionales.

Las probabilidades encontradas de las variables incidentes muestran que tener una enfermedad grave o aguda incrementa la probabilidad de abandono, un rendimiento académico excelente o muy bueno por el contrario disminuye la probabilidad para deserción, un estudiante al que no le gusta o le gusta muy poco la carrera tiene mayor incidencia en dejar sus estudios y un nivel de escolaridad bajo del padre o jefe de hogar también incrementa la probabilidad de abandono.

El modelo logit formulado para identificar factores incidentes de deserción, presenta buen ajuste y una precisión moderada, con una capacidad de predicción, del 84% para proporcionar información sobre la asociación entre las variables predictoras y la variable de respuesta en el contexto.

Referencias bibliográficas

- B. Rodríguez, A., Espinoza, J., Ganga, A., & J. Ramirez, L. (2018). Deserción Universitaria: Nuevo Análisis Metodológico. Scielo, 107-118.
- Bernardo Gutiérrez, A., Cerezo Menéndez, R., Núñez Pérez, J., & Tuero Herrero, E. (2015). Predicción del abandono universitario: Variables explicativas y medidas de prevención. Revista Fuentes, 63-84.
- Bruce, P., Bruce, A., & Gedeck, P. (2022). Estadística práctica para ciencia de datos con r y python. Maroombo.
- Cevallos, F. (2014). Deserción, calidad y reforma universitaria. Quito: Contrato Social.
- Da Re, L., & Clerici, R. (2017). Abandono, rendimiento académico y tutoría. *Educatio Siglo XXI*.
- Díaz Peralta, C. (2008). Modelo conceptual para la deserción estudiantil universitaria chilena. *Estudios pedagógicos (Valdivia)*, 65-86.
- Donggeun, K., & Seoyong, K. (2018). Sustainable education: Analyzing the determinants of

- university student dropout by nonlinear panel data models. *Sustainability. Research Gate*, 1-18.
- Ferreira, M., Avitabile, C., & Alvarez, J. (2017). *Momento decisivo: La Educación Superior en América Latina y el Caribe*. Washington: Grupo Banco Mundial.
- Gujarati, D. (2010). *Econometría*. México: Mc Graw Hill.
- Gutierrez, D., Vélez, J., & López, J. (2021). *Indicadores de deserción universitaria y factores asociados*. Educat.
- Hernández Prado, M., Álvarez Muñoz, J., & Aranda Martínez, A. (2016). *El problema de la deserción escolar en la producción científica educativa*. *Revista Internacional de Ciencias Sociales y Humanidades*.
- Hernández, E., Álvarez, E., & Ruiz, A. (s.f.). *Salud y desempeño en estudiantes de distintos niveles académicos*.
- INEC. (2022). *La educación en cifras*.
- Irrizari, R. (2020). *Introduction to Data Science Data Analysis and Prediction Algorithms with R*. London - New York: CRC Press.
- Lázaro Alvarez, N., Callejas, Z., & Griol, D. (2020). *Factores que inciden en la deserción estudiantil en carreras de perfil*. *Revista Fuentes*, 105-126.
- Lopera Oquendo, C. (2007). *Determinants of Student Dropout in the Department of Economics at the University Del Rosario (Determinantes De La Deserción Universitaria En La Facultad De Economía Universidad Del Rosario)* (Spanish). *Borradores de Investigación, Facultad de Economía, Universidad del Rosario*.
- Martínez, J., Pérez, E., Fernández, Y., & Bermúdez, L. (2021). *Análisis predictivo de la deserción estudiantil en la carrera de Medicina*. *Edumecentro*.
- Rodríguez, M., & Mora, R. (2021). *Informática y Estadística: Casos y ejemplos con el SPSS*. Universidad de Alicante.
- Roland, M., Frenay, M., & Boudrenghien, G. (2018). *Understanding Academic Persistence through the Theory of Planned Behavior: Normative Factors under Investigation*. *Journal of College Student Retention: Research, Theory & Practice*, 215-235.
- Solís, J., Fernández, S., & Téllez, O. (2022). *Modelo de regresión logística para la estimación de la deserción escolar del posgrado en la Universidad Técnica de Manabí, Ecuador*. *Revista bases de la ciencia*, 1-14.
- Tinto, V. (1975). *Dropout from Higher Education: A theoretical synthesis of recent research*.
- Venegas, G., Chiluisa, M., Casillas, I., & Castro, S. (2017). *La deserción en la Educación*. Dialnet.