



UNIVERSIDAD NACIONAL DE CHIMBORAZO

**VICERRECTORADO DE INVESTIGACIÓN,
VINCULACIÓN Y POSGRADO**

DIRECCIÓN DE POSGRADO

TESIS PREVIA A LA OBTENCIÓN DEL GRADO DE:

**MAGÍSTER EN MATEMÁTICA APLICADA MENCIÓN MATEMÁTICA
COMPUTACIONAL**

TEMA:

**“DISEÑO DE UN MODELO MATEMÁTICO MEDIANTE TÉCNICAS DE
ANÁLISIS MULTIVARIADO PARA ESTIMAR LA DESERCIÓN
ESTUDIANTIL EN EL INSTITUTO TECNOLÓGICO SUPERIOR
COTACACHI”**

AUTOR:

Lic. Juan Pablo Estrada Arana

TUTOR:

Ing. Juan Carlos Paz Espinoza, Mgs

Riobamba – Ecuador

2025

CERTIFICACIÓN DEL TUTOR

Certifico que el presente trabajo de titulación denominado: **“Diseño de un Modelo Matemático Mediante Técnicas de Análisis Multivariado para Estimar la Deserción Estudiantil en el Instituto Tecnológico Superior Cotacachi”**, ha sido elaborado por el Licenciado Juan Pablo Estrada Arana el mismo que ha sido orientado y revisado con el asesoramiento permanente de mi persona en calidad de Tutor. Así mismo, refrendo que dicho trabajo de titulación ha sido revisado por la herramienta antiplagio institucional; por lo que certifico que se encuentra apto para su presentación y defensa respectiva.

Es todo cuanto puedo informar en honor a la verdad.

Riobamba, 07 de enero de 2025

Ing. Juan Carlos Paz Espinoza, Mgs

TUTOR

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

Yo, **Juan Pablo Estrada Arana**, con número único de identificación **100387674-3**, declaro y acepto ser responsable de las ideas, doctrinas, resultados y lineamientos alternativos realizados en el presente trabajo de titulación denominado: “DISEÑO DE UN MODELO MATEMÁTICO MEDIANTE TÉCNICAS DE ANÁLISIS MULTIVARIADO PARA ESTIMAR LA DESERCIÓN ESTUDIANTIL EN EL INSTITUTO TECNOLÓGICO SUPERIOR COTACACHI”, previo a la obtención del grado de Magíster en Matemática Aplicada con mención en Matemática Computacional.

- Declaro que mi trabajo investigativo pertenece al patrimonio de la Universidad Nacional de Chimborazo de conformidad con lo establecido en el artículo 20 literal j) de la Ley Orgánica de Educación Superior LOES.
- Autorizo a la Universidad Nacional de Chimborazo que pueda hacer uso del referido trabajo de titulación y a difundirlo como estime conveniente por cualquier medio conocido, y para que sea integrado en formato digital al Sistema de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor, dando cumplimiento de esta manera a lo estipulado en el artículo 144 de la Ley Orgánica de Educación Superior LOES.

Riobamba, 07 de enero de 2025

Lic. Juan Pablo Estrada

N.U.I. 100387674-3



Dirección de
Posgrado
VICERRECTORADO DE INVESTIGACIÓN,
VINCULACIÓN Y POSGRADO



Riobamba, 16 de diciembre de 2024

ACTA DE SUPERACIÓN DE OBSERVACIONES

En calidad de miembro del Tribunal designado por la Comisión de Posgrado, CERTIFICO que una vez revisado el Proyecto de Investigación y/o desarrollo denominado "**Diseño de un Modelo Matemático Mediante Técnicas de Análisis Multivariado para Estimar la Deserción Estudiantil en el Instituto Tecnológico Superior Cotacachi**", dentro de la línea de investigación de **Ingeniería informática, presentado por el maestrante Estrada Arana Juan Pablo**, portador de la CI. 1003876743, del programa de **Maestría en Matemática Aplicada, mención Matemática Computacional**, cumple al 100% con los parámetros establecidos por la Dirección de Posgrado de la Universidad Nacional de Chimborazo.

Es todo lo que podemos certificar en honor a la verdad.

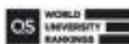
Atentamente,



GUILLERMO EDVIN
MACHADO SOTOMAYOR

Dr. Guillermo Machado Sotomayor

MIEMBRO DEL TRIBUNAL



Campus La Dolorosa
Av. Eloy Alfaro y 10 de Agosto
Teléfono (593-3) 373-0880, ext. 2002
Riobamba - Ecuador

Unach.edu.ec
en asociación



Dirección de
Posgrado

VICERRECTORADO DE INVESTIGACIÓN,
VINCULACIÓN Y POSGRADO



Riobamba, 17 de diciembre de 2024

ACTA DE SUPERACIÓN DE OBSERVACIONES

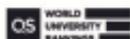
En calidad de miembro del Tribunal designado por la Comisión de Posgrado, CERTIFICO que una vez revisado el Proyecto de Investigación y/o desarrollo denominado **"Diseño de un Modelo Matemático Mediante Técnicas de Análisis Multivariado para Estimar la Deserción Estudiantil en el Instituto Tecnológico Superior Cotacachi"**, dentro de la línea de investigación de **Ingeniería informática**, presentado por el maestrante **Estrada Arana Juan Pablo**, portador de la CI. 1003876743, del programa de **Maestría en Matemática Aplicada**, mención **Matemática Computacional**, cumple al 100% con los parámetros establecidos por la Dirección de Posgrado de la Universidad Nacional de Chimborazo.

Es todo lo que podemos certificar en honor a la verdad.

Atentamente,



Ing. Henry M. Villa Yáñez, Mgs.
MIEMBRO DEL TRIBUNAL



Campus La Dolorosa
Av. Eloy Alfaro y 10 de Agosto
Teléfono (593-3) 373-0880, ext. 2002
Riobamba - Ecuador

Unach.edu.ec
en movimiento



Dirección de
Posgrado
VICERRECTORADO DE INVESTIGACIÓN,
VINCULACIÓN Y POSGRADO



Riobamba, 12 de diciembre de 2024

ACTA DE SUPERACIÓN DE OBSERVACIONES

En calidad de miembro del Tribunal designado por la Comisión de Posgrado, CERTIFICO que una vez revisado el Proyecto de Investigación y/o desarrollo denominado "**Diseño de un Modelo Matemático Mediante Técnicas de Análisis Multivariado para Estimar la Deserción Estudiantil en el Instituto Tecnológico Superior Cotacachi**", dentro de la línea de investigación de **Ingeniería informática**, presentado por el maestrante Estrada Arana Juan Pablo, portador de la CC. 1003876743, del programa de **Maestría en Matemática Aplicada**, mención **Matemática Computacional**, cumple al 100% con los parámetros establecidos por la Dirección de Posgrado de la Universidad Nacional de Chimborazo.

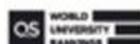
Es todo lo que puedo certificar en honor a la verdad.

Atentamente,



Ing. Juan Carlos Paz Espinoza, Mgs

TUTOR



Campus La Dolorosa
Av. Eloy Alfaro y 10 de Agosto
Teléfono (593-3) 373-0880, ext. 2002
Riobamba - Ecuador

Unach.edu.ec
en movimiento



Dirección de
Posgrado
VICERRECTORADO DE INVESTIGACIÓN,
VINCULACIÓN Y POSGRADO



Riobamba, 07 de enero de 2025

CERTIFICADO

De mi consideración:

Yo Juan Carlos Paz Espinoza, certifico que Juan Pablo Estrada Arana con cédula de identidad No. 1003876743 estudiante del programa de maestría en Maestría en matemática aplicada con mención en matemática computacional, cohorte Segunda (2023-2024), presentó su trabajo de titulación bajo la modalidad de Proyecto de titulación con componente de investigación aplicada/desarrollo denominado: Diseño de un Modelo Matemático Mediante Técnicas de Análisis Multivariado para Estimar la Deserción Estudiantil en el Instituto Tecnológico Superior Cotacachi, el mismo que fue sometido al sistema de verificación institucional de contenido identificando el 5% en el texto.

Es todo en cuento puedo certificar en honor a la verdad.

Atentamente,



VERIFICAR VERIFICACIONES
JUAN CARLOS PAZ
ESPINOZA

Juan Carlos Paz Espinoza

CI: 1002834800



Campus La Dolorosa
Av. Eloy Alfaro y 10 de Agosto
Teléfono (593-3) 373-0880, ext. 2002
Riobamba - Ecuador

Unach.edu.ec
en movimiento

AGRADECIMIENTO

A Dios por ser mi guía y fortaleza en todo momento.

Al MSc. Juan Paz por su ejemplo de profesionalismo, rectitud y amistad, siendo guía para el presente trabajo de investigación.

A mis docentes de quienes me llevo las enseñanzas humanas y profesionales.

A mis amigos más cercanos por su lealtad y apoyo incondicional.

DEDICATORIA

A mis padres Olga y Jorge, mis hermanos Katherine, Pamela y Andrés, eje central de mi crecimiento personal y profesional.

A mis tíos Susana, Luis y Manuel, (+) esencia de mi vida.

ÍNDICE GENERAL

CERTIFICACIÓN DEL TUTOR

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

ACTAS DE SUPERACIÓN DE OBSERVACIONES

CERTIFICADO ANTIPLAGIO

AGRADECIMIENTO

DEDICATORIA

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

RESUMEN

ABSTRACT

Introducción	3
Capítulo 1 Generalidades	5
1.1 Planteamiento del Problema	5
1.2 Justificación de la Investigación	6
1.3 Objetivos	7
1.3.1 Objetivo General	7
1.3.2 Objetivos Específicos.....	8
1.4 Descripción de la empresa y puestos de trabajo	8
Capítulo 2 Estado del arte y la práctica.....	10
2.1 Antecedentes Investigativos.....	10
2.2 Fundamentación Legal.....	14
2.2.1 Legislación Internacional.....	14
2.2.2 Legislación Nacional	14
2.3 Fundamentación Teórica.....	17
2.3.1 Deserción Estudiantil	17
2.3.2 Modelos matemáticos para la predicción de la deserción estudiantil	19
2.3.3 Modelos matemáticos predictivos comunes	20

2.3.4	Regresión Logística	24
Capítulo 3	Diseño Metodológico	27
3.1	Enfoque de la Investigación.....	27
3.2	Diseño de la Investigación.....	27
3.3	Tipo de Investigación.....	27
3.4	Nivel de Investigación	28
3.5	Técnicas e Instrumentos de Recolección de Datos	28
3.6	Técnicas para el Procesamiento e Interpretación de Datos.....	29
3.7	Población y Muestra	32
3.7.1	Población.....	32
3.7.2	Tamaño de la Muestra.....	32
CAPÍTULO 4	Análisis y discusión de los resultados	34
4.1	Análisis Descriptivo de los Resultados.....	34
4.2	Análisis Correlacional.....	38
4.3	Modelo de Regresión Logística	39
4.4	Validación Cruzada.....	40
4.5	Discusión de los Resultados.....	46
Capítulo 5	Marco Propositivo.....	47
5.1	Planificación de la Actividad Preventiva.....	47
5.1.1	Actividades Preventivas.....	47
5.1.2	Sistema de Evaluación	51
5.2	Impacto y Beneficios Esperados	53
Conclusiones		55
Recomendaciones		57
Referencias Bibliográficas		59
Anexos		63
Anexo A.	Base de datos.....	63
Anexo B.	Depuración y limpieza.....	64
Anexo C.	Normalización	65
Anexo D.	Programación	66

Anexo E. Correlación	67
Anexo F. Diagrama de calor.....	67
Anexo G. Matriz de confusión.....	68
Anexo H. Modelo de regresión.....	69
Anexo I. Ajuste del modelo	69

ÍNDICE DE TABLAS

Tabla 1 Factores que influyen en la deserción estudiantil	18
Tabla 2 Categorización de variables del modelo	34
Tabla 3 Resumen de métricas del modelo de regresión logística	39
Tabla 4 Validación Cruzada.....	40
Tabla 5 Coeficientes de regresión logística	41
Tabla 6 Resumen de métricas del modelo	43
Tabla 7 Resultados de la regresión logística.....	45
Tabla 8 Acciones académicas	51
Tabla 9 Acciones psicosociales	52
Tabla 10 Acciones socioeconómicas	53
Tabla 11 Evaluación general.....	53

ÍNDICE DE FIGURAS

Figura 1 Función lineal del modelo de regresión logística	24
Figura 2 Proceso metodológico	29
Figura 3 Mapa de calor	38
Figura 4 Matriz de Confusión y Desempeño del Modelo.....	39
Figura 5 Curva ROC del Modelo de Clasificación.....	40

RESUMEN

El presente trabajo de titulación, denominado "Diseño de un modelo matemático mediante técnicas de análisis multivariado para estimar la deserción estudiantil en el Instituto Tecnológico Superior Cotacachi," tiene como objetivo desarrollar un modelo predictivo que permita estimar la tasa de deserción estudiantil en la institución. La investigación analiza factores socioeconómicos, académicos y demográficos que inciden en la deserción, aplicando técnicas de análisis multivariado para su identificación y cuantificación.

La metodología empleada incluye la recolección de datos de 1,385 registros de estudiantes matriculados entre 2022 y 2024, seguida de un análisis descriptivo y la aplicación de un modelo de regresión logística multivariante. Este modelo fue validado a través de métricas de desempeño como precisión, recall y F1-score, obteniendo una exactitud del 79%.

Los resultados destacan que variables como la edad, la repetición de asignaturas y la pérdida de gratuidad influyen significativamente en la deserción, mientras que factores como el sexo y el estado civil no resultaron estadísticamente relevantes. En conclusión, el modelo predictivo propuesto responde al objetivo de identificar con precisión a los estudiantes en riesgo de deserción, permitiendo la implementación de estrategias de intervención temprana.

Palabras claves: deserción estudiantil, análisis multivariado, regresión logística, modelo predictivo, educación.

ABSTRACT

The aim of this work, called “Design of a mathematical model using multivariate analysis techniques to estimate student desertion at the Instituto Tecnológico Superior Cotacachi,” is to develop a predictive model that allows to estimate the student desertion rate at the institution. The research analyzes socioeconomic, academic and demographic factors that influence desertion, applying multivariate analysis techniques for its identification and quantification.

The methodology used includes the data collection from 1,385 student records enrolled between 2022 and 2024, followed by a descriptive analysis and the application of a multivariate logistic regression model. This model was validated through performance metrics such as precision, recall and F1-score, obtaining an accuracy of 79%.

The results highlight that variables such as age, repetition of subjects and loss of gratuity significantly influence desertion, while factors such as gender and marital status were not statistically relevant. In conclusion, the proposed predictive model responds to the objective of accurate identification of students at risk of desertion, allowing the implementation of early intervention strategies.

Keywords: student desertion, multivariate analysis, logistic regression, predictive model, education.



EDISON BERNAN
SALAZAR CALDERON

Reviewed by:
Mgs. Edison Salazar Calderón
ENGLISH PROFESSOR
I.D. 0603184698

Introducción

El fenómeno de la deserción estudiantil representa un desafío crítico en las instituciones de educación superior, afectando tanto al desarrollo académico como a las políticas educativas. Este trabajo aborda la problemática en el Instituto Tecnológico Superior Cotacachi, ubicado en la provincia de Imbabura en el Ecuador. A través del diseño de un modelo matemático basado en técnicas de análisis multivariado, se busca estimar la tasa de deserción en esta institución, lo que permite identificar los factores más influyentes y, en consecuencia, proponer estrategias de intervención y apoyo a los estudiantes en riesgo de abandono académico.

La relevancia de esta investigación radica en su contribución tanto al ámbito académico, al mejorar la comprensión de los factores que influyen en la deserción, como al contexto empresarial y social, al generar propuestas de intervención que optimicen los recursos institucionales y disminuyan las tasas de abandono, garantizando una inserción laboral en base a las necesidades sociales. Además, este estudio tiene implicaciones significativas para las políticas educativas, proporcionando herramientas para mejorar la retención estudiantil y el éxito académico.

En cuanto a la metodología, se utiliza un enfoque cuantitativo no experimental, con un diseño aplicado y documental. Los datos recolectados de los estudiantes son analizados mediante técnicas de regresión logística, que permite la construcción de un modelo predictivo que al ser validado asegura su precisión en identificar los principales factores sociales, económicos, académicos que inciden en la deserción y el desarrollo de recomendaciones específicas para mitigar este problema.

El trabajo se estructura en cinco capítulos: el capítulo I abarca las generalidades, el planteamiento del problema, la justificación y los objetivos de la investigación; el capítulo II presenta el estado del arte y el marco teórico, revisando estudios previos sobre la deserción y las leyes aplicables para el estudio dentro de la institución; en el capítulo III se aborda el diseño metodológico utilizado para recolectar y analizar los datos, los que fueron sometidos a análisis de los resultados obtenidos y su discusión; en el capítulo IV se evalúa la eficacia del modelo predictivo. Finalmente, en el capítulo V se expone una propuesta de intervención para la institución, basada en los hallazgos del estudio.

Con esta estructura, se espera que la investigación aporte una solución práctica basada en datos reales y de campo, para prevenir la deserción estudiantil en el Instituto Tecnológico Superior Cotacachi y en otras instituciones de educación.

Capítulo 1

Generalidades

1.1 Planteamiento del Problema

La deserción estudiantil es un fenómeno crítico que afecta a muchas instituciones educativas en el Ecuador, incluyendo al Instituto Tecnológico Superior Cotacachi, una entidad con dependencia fiscal. Esta problemática no solo representa una pérdida significativa para los estudiantes que ven truncadas sus aspiraciones académicas y profesionales, sino que también impacta negativamente la eficiencia y el prestigio de la institución educativa. Además, tiene implicaciones económicas para el país, dado que la educación superior es un elemento clave para el desarrollo económico y social.

Dada la complejidad de los factores socioeconómicos, personales, académicos que influyen en la deserción estudiantil, es esencial desarrollar un modelo matemático que permita a la institución identificar a los estudiantes en riesgo de desertar y tomar medidas preventivas. Un enfoque basado en técnicas de análisis multivariado ofrece una oportunidad para analizar múltiples variables simultáneamente y determinar cuáles son los factores más significativos en la deserción.

Por todo lo expuesto, la implementación de este modelo permite a las autoridades del Instituto Tecnológico Superior Cotacachi tomar decisiones informadas y desarrollar estrategias efectivas para reducir la deserción estudiantil. Además, contribuye a mejorar la eficiencia de la institución, aumentando la tasa de graduación y, fortaleciendo el vínculo entre la educación superior y el mercado laboral en la región.

1.2 Justificación de la Investigación

La deserción estudiantil en la educación superior es un problema que tiene consecuencias significativas a nivel individual, institucional y social. Al desertar el estudiante se encuentra implicado en la interrupción de su formación académica y profesional lo que puede truncar sus oportunidades de desarrollo personal y laboral. Además afecta a la eficiencia de las instituciones educativas de la mano con su reputación, y conlleva a una pérdida de recursos humanos y económicos, mientras que para la sociedad representa una disminución en la capacidad para desarrollar capital humano calificado, lo que puede limitar el crecimiento económico y el desarrollo sostenible.

El Instituto Tecnológico Superior Cotacachi, al ser una institución con dependencia fiscal ofrece una educación de calidad alineada con las políticas públicas en relación a las necesidades del mercado laboral actual. Por lo tanto, es fundamental abordar este problema para optimizar el uso de los recursos públicos y garantizar que la educación superior sea accesible y relevante para todos los estudiantes.

Actualmente, la falta de un enfoque sistemático basado en datos para identificar y abordar los factores que contribuyen a la deserción estudiantil limita la capacidad del Instituto para implementar estrategias efectivas de retención estudiantil. Por ello, la investigación propone el diseño de un modelo matemático mediante técnicas de análisis multivariado con datos del 100% de los estudiantes matriculados en el Instituto Tecnológico Superior Cotacachi entre 2022 y 2024. Este tamaño muestral es adecuado para el modelo, garantizando una adecuada potencia estadística y generalización de los resultados. Además, la inclusión de registros de diferentes periodos académicos asegura la captura de variaciones estacionales y

estructurales en las tasas de abondono estudiantil que permita estimar la deserción estudiantil en el ITSCO.

Este modelo ofrecerá una herramienta valiosa para:

- Identificar a los estudiantes en riesgo de deserción.
- Informar la toma de decisiones.
- Optimizar el uso de recursos.
- Mejorar la calidad de la educación.

En conclusión, esta investigación surge por la necesidad de abordar un problema crítico en la educación superior, utilizando un enfoque innovador y basado en datos reales. Al desarrollar un modelo matemático para estimar la deserción estudiantil, se espera mejorar la eficacia de las intervenciones educativas y promover el éxito académico y profesional de los estudiantes en las instituciones de educación superior.

1.3 Objetivos

1.3.1 Objetivo General

- Diseñar un modelo matemático basado en técnicas de análisis multivariado para estimar la tasa de deserción estudiantil en el Instituto Tecnológico Superior Cotacachi.

1.3.2 Objetivos Específicos

- Realizar un análisis descriptivo de datos relevantes de los estudiantes de la institución de educación superior tecnológica involucrados en la investigación en cuanto a su desempeño académico y condiciones socioeconómicas
- Aplicar el modelo de análisis multivariante con regresión logística para la obtención del modelo predictivo.
- Validar el modelo predictivo para determinar la probabilidad de deserción de los estudiantes involucrados.
- Proponer recomendaciones para la intervención y apoyo a estudiantes en riesgo identificados por el modelo.

1.4 Descripción de la empresa y puestos de trabajo

El Instituto Superior Tecnológico Cotacachi (ISTCO), inicia como el Instituto Tecnológico de la Industria del Cuero “Cotacachi” el 11 de agosto de 1995, según la resolución Nº 4066 del Ministerio de Educación y Cultura. En el año 2000, la institución se incorporó al Sistema de Educación Superior del Ecuador, según la Ley de Educación Superior de 2010. En 2001, el Consejo Nacional de Educación Superior le asignó el registro institucional Nº10-011, lo que permitió al Instituto ofrecer carreras tecnológicas en Diseño, Confecciones, Marroquinería, Línea de Viaje y Calzado, adoptando entonces el nombre de Instituto Tecnológico Cotacachi. En 2010, la gestión del Instituto pasó a la Secretaría de Educación Superior, Ciencia, Tecnología e Innovación. En 2013, se unió al Proyecto de Reconversión de la Educación Técnica y Tecnológica Superior Pública, trasladándose para el 2017 a una nueva sede en la calle Waminka Rumiñahui, parroquia San Francisco, cantón Santa Ana de Cotacachi,

provincia de Imbabura. Finalmente, mediante el Oficio Nro. CES-SG-2019-1354-O del 13 de junio de 2019, el Consejo de Educación Superior otorgó a la Institución la denominación oficial de Instituto Superior Tecnológico Cotacachi (ISTCO), una entidad adscrita al organismo rector en materia de política educativa, ciencia, tecnología e innovación que en la actualidad ofrece programas de Tecnologías Superiores en: Desarrollo Infantil Integral, Diseño de Modas, Confección Textil, Gastronomía, Administración, Modelado y Producción Industrial de Indumentaria y Gestión de Operaciones Turísticas, todas ellos aprobados por el Consejo de Educación Superior con un total de 653 estudiantes matriculados.

CAPÍTULO 2

ESTADO DEL ARTE Y LA PRÁCTICA

2.1 Antecedentes Investigativos

Gaviria (2024) desarrolló un modelo predictivo enfocado en identificar los factores que influyen en la deserción estudiantil universitaria, utilizando técnicas de análisis de datos en el marco de la educación por ciclos propedéuticos. El modelo permitió activar alertas tempranas y mejorar la toma de decisiones en las instituciones educativas. El estudio, basado en un diseño de casos y controles, analizó datos de 21,460 estudiantes de una institución de educación superior entre 2017 y 2022, de los cuales 1,106 abandonaron sus programas. Se aplicaron técnicas como redes neuronales, KNN y regresión logística, obteniendo resultados que señalaron múltiples factores de riesgo y protección vinculados a la deserción. El análisis evidenció que los estudiantes en ciclos propedéuticos presentaban un mayor riesgo de abandonar sus estudios, y la regresión logística alcanzó una precisión del 94%, demostrando la eficacia del enfoque analítico utilizado. Este trabajo contribuye a la validación y mejora de modelos matemáticos, al incorporar variables y contextos educativos adicionales. En sintonía con esta investigación, el presente estudio busca, a través de técnicas de análisis multivariado, diseñar estrategias de retención más personalizadas y eficaces para las instituciones de educación superior, considerando que las carreras ofertadas por el ITSCO están orientadas directamente al mercado laboral.

Álvarez, et. al. (2021) realizaron un estudio que tiene como objetivo identificar los factores que predicen la deserción en el programa de Ingeniería en Ciencias Informáticas mediante el uso del software SPSS. Para seleccionar las variables se emplearon métodos

histórico-lógicos y de análisis-síntesis, complementados con análisis estadísticos descriptivos e inferenciales. Estas herramientas permitieron relacionar variables independientes, tales como género, provincia de origen, fuente de ingreso, elección de carrera, y rendimiento en Matemática y Programación, con la variable dependiente de deserción estudiantil. La muestra incluyó a 485 estudiantes, y los resultados mostraron que los principales predictores de deserción eran la provincia de origen, la fuente de ingreso, la nota de acceso en Matemática y el rendimiento académico. El estudio concluyó que tanto las variables académicas como demográficas son claves para anticipar el abandono estudiantil, lo que permite a las instituciones educativas implementar acciones para abordar este problema. Este trabajo proporciona una base importante para la presente investigación, en la que se exploran variables adicionales en contextos diferentes, con el fin de fortalecer las estrategias de retención y mejorar su impacto en la educación y la sociedad.

El trabajo realizado por Vinueza (2021) titulado "Diseño de un modelo matemático para estimar la deserción estudiantil mediante técnicas de análisis multivariado en una institución de educación superior tecnológica," llevado a cabo en el Instituto Superior Tecnológico Luis A. Martínez Agronómico, se centra en la problemática de la deserción estudiantil con el propósito de desarrollar un modelo matemático que permita estimarla. Para ello, se empleó la metodología KDD (Knowledge Discovery in Databases), analizando información de 849 estudiantes matriculados entre 2018 y 2020. Las variables consideradas en el estudio incluyeron género, estado civil, edad, carrera, repitencia, ocupación e ingresos económicos. Durante el proceso, se generaron cuatro modelos de regresión logística, siendo refinado el modelo final para enfocarse únicamente en las variables 'carrera' y 'repitencia', logrando una clasificación correcta del 83% en los datos de entrenamiento y del 79% en los datos de testeo.

Paralelamente, se diseñó un modelo predictivo basado en árboles de decisión, donde la variable ‘carrera’ resultó ser la más significativa. No obstante, el F1_Score del modelo de regresión logística fue superior al de los árboles de decisión, lo que resalta la eficacia de la regresión logística para estimar la deserción estudiantil, subrayando la relevancia de ‘carrera’ y ‘repitencia’ como factores clave. Este enfoque permite a las instituciones educativas detectar a los estudiantes con mayor riesgo de abandono. Además, el estudio ofrece un marco de referencia útil para el presente trabajo, ya que su metodología puede ser aplicada en otras instituciones para mejorar las tasas de retención estudiantil. La consideración de un mayor número de variables podría mejorar la precisión en la predicción de la deserción.

Quintero (2002) desarrolló un modelo predictivo destinado a generar alertas tempranas sobre el riesgo de deserción en los programas de pregrado presenciales de la Facultad de Ingeniería de la Universidad de Antioquia. El objetivo principal fue diseñar una herramienta que permitiera identificar a los estudiantes en riesgo de abandonar sus estudios de manera anticipada. Para ello, se emplearon datos históricos tanto de la universidad como del ICFES, y se entrenaron modelos predictivos utilizando técnicas de machine learning, entre las que se destacaron las redes neuronales artificiales (RNA) y el algoritmo Xtreme Gradient Boosting (XGBoost). El estudio abarcó los tres primeros semestres de la carrera: el modelo del primer semestre se basó únicamente en los datos de ingreso, mientras que los de los semestres siguientes integraron variables relacionadas con el desempeño académico. Los resultados revelaron que la precisión de los modelos aumentaba conforme los estudiantes progresaban en sus estudios, concluyendo que estas herramientas predictivas son eficaces para detectar la deserción temprana y permiten a las instituciones de educación superior desarrollar estrategias de intervención más eficientes. En el contexto del Instituto Tecnológico Superior Cotacachi,

este enfoque contribuye a validar el uso de técnicas de aprendizaje automático en el análisis de la deserción estudiantil, sugiriendo la incorporación de datos adicionales para mejorar la precisión de los modelos y optimizar las estrategias de retención.

Álvarez, et al. (2020) realizaron un estudio sobre los factores que inciden en la deserción de estudiantes en carreras de Ingeniería Informática, con el fin de desarrollar un modelo teórico que explique este fenómeno. Inicialmente, llevaron a cabo una revisión exhaustiva de la literatura para identificar los principales factores relacionados con la deserción. Posteriormente, realizaron un análisis cuantitativo correlacional, multivariado y predictivo, empleando una muestra representativa de estudiantes de todas las provincias de Cuba, en el que se incluyeron tanto variables previas como posteriores al ingreso a la educación superior. Los hallazgos indicaron que el rendimiento académico en Matemática y Programación, la nota del examen de ingreso en Matemática, la elección de carrera y la provincia de origen son factores clave que influyen en el abandono estudiantil. Estos resultados proporcionaron a las instituciones educativas información valiosa para desarrollar estrategias más efectivas de retención. La investigación también resaltó que la metodología utilizada es adaptable a otros contextos, lo que permite explorar diferentes variables y mejorar la comprensión sobre la deserción en diversas instituciones educativas, facilitando el diseño de mejores estrategias para prevenir el abandono escolar.

2.2 Fundamentación Legal

2.2.1 *Legislación Internacional.*

2.2.1.1 Leyes y tratados internacionales.

La Declaración Universal de Derechos Humanos, adoptada por la Asamblea General de las Naciones Unidas en 1948, establece en su Artículo 26 que "toda persona tiene derecho a la educación", y especifica que la educación debe orientarse al pleno desarrollo de la personalidad humana y al fortalecimiento del respeto a los derechos humanos y las libertades fundamentales. Este marco internacional promueve la inclusión y la continuidad educativa como derechos fundamentales, aspectos relevantes para abordar y mitigar la deserción estudiantil.

2.2.1.2 Resoluciones y acuerdos

La Agenda 2030 para el Desarrollo Sostenible, adoptada por todos los Estados Miembros de las Naciones Unidas en 2015, incluye el Objetivo de Desarrollo Sostenible 4 (ODS 4), que busca "garantizar una educación inclusiva y equitativa de calidad y promover oportunidades de aprendizaje durante toda la vida para todos". Dentro de este objetivo, se establece la importancia de mejorar la retención escolar y reducir las tasas de abandono.

2.2.2 *Legislación Nacional*

2.2.2.1 Leyes

La Constitución de la República del Ecuador (2008) establece un marco normativo que prioriza la educación como un derecho fundamental y un deber ineludible del Estado, en su Artículo 26, se reconoce a la educación como indispensable para el desarrollo individual y

colectivo, lo que demanda la implementación de políticas que garanticen no solo el acceso, sino también la permanencia y culminación de los estudios.

De igual manera, el Artículo 27 destaca que la educación debe ser inclusiva, equitativa y orientada a la justicia social, exigiendo herramientas innovadoras que permitan reducir desigualdades y asegurar la continuidad académica. Por lo tanto, el diseño de modelos matemáticos predictivos basados en técnicas de análisis multivariado como la regresión logística, se alinea con estos principios al ofrecer una herramienta analítica que identifica de manera temprana a los estudiantes en riesgo de deserción, permitiendo la implementación de políticas proactivas.

La Ley Orgánica de Educación Superior (LOES) establece directrices específicas para la calidad y permanencia en la educación superior. El Artículo 72 enfatiza que las instituciones de educación superior deben desarrollar programas de apoyo a los estudiantes que se encuentren en riesgo de deserción, incluyendo tutorías, becas y otros tipos de asistencia. Del mismo modo, el Artículo 77 obliga a las instituciones a implementar mecanismos para garantizar la calidad educativa, entre los cuales se incluyen herramientas de análisis predictivo para mitigar la deserción. El uso de modelos matemáticos, como la regresión logística, permite cumplir con este mandato al proporcionar predicciones basadas en datos reales que facilitan la toma de decisiones fundamentadas.

Finalmente, el Artículo 90, al referirse a las políticas de bienestar estudiantil, y el Artículo 91, que menciona la creación de un sistema de información de educación superior, subrayan la importancia de contar con datos confiables y herramientas analíticas que permitan monitorear y prever tendencias como la deserción. En este contexto, el modelado matemático,

al integrar técnicas de análisis multivariado, contribuye significativamente al cumplimiento de estos objetivos legales, asegurando que las instituciones educativas implementen estrategias basadas en evidencia para mejorar la permanencia estudiantil y fortalecer el sistema educativo nacional.

2.2.2.2 Resoluciones y acuerdos

La Resolución CES No. 048/2013 aborda la necesidad de que las instituciones de educación superior implementen sistemas de monitoreo y evaluación continua de sus programas académicos, incluyendo la identificación de factores de riesgo para la deserción estudiantil. Esta resolución refuerza la importancia de desarrollar modelos matemáticos y técnicas de análisis para estimar y prevenir la deserción.

El Acuerdo Ministerial No. 140-2019 establece directrices para el fortalecimiento de los sistemas de apoyo y acompañamiento estudiantil en las instituciones de educación superior, con el objetivo de mejorar la retención y el éxito académico de los estudiantes. Este acuerdo es relevante para la investigación, ya que proporciona un marco normativo para el diseño y la implementación de estrategias dirigidas a reducir la deserción.

Los modelos matemáticos predictivos, basados en técnicas como la regresión logística, ofrecen una solución práctica y fundamentada para abordar el problema de la deserción estudiantil en las instituciones de educación superior en Ecuador. Al identificar patrones en los datos socioeconómicos, académicos y personales de los estudiantes, estos modelos permiten prever quiénes están en riesgo de abandonar sus estudios, facilitando la intervención temprana. Esta capacidad predictiva no solo contribuye al cumplimiento de los mandatos internacionales y nacionales legales de la Constitución de la República, la LOES y las normativas del CES,

sino que también fortalece la equidad y la inclusión educativa. Implementar estas herramientas no es solo una alternativa innovadora, sino una necesidad estratégica para prevenir la deserción, optimizar recursos y mejorar la calidad de la educación superior en el país.

2.3 Fundamentación Teórica

2.3.1 Deserción Estudiantil

La deserción se puede definir según Picardo, et. al. (2004): “el acto deliberado o forzado mediante el cual el estudiante deja su aula o centro educativo, cuyo abandono tiene a la base múltiples causas, entre ellas sociales y económicas” (p-72).

Aunque la definición de deserción estudiantil mantiene discusión constante, se busca precizarla enfocándola como un abandono que puede ser explicado por diferentes categorías de variables: socioeconómicas, individuales, institucionales y académicas debido a esto hay que interpretarla como una situación a la que se enfrenta un estudiante cuando aspira y no logra concluir su proyecto educativo.

2.3.1.1 Importancia del estudio

En el año 2023, la tasa de deserción universitaria en Ecuador se situó en un 20,46%. Además, los hombres mostraron una mayor tasa de deserción académica (21,28%) en comparación con las mujeres (16,02%). Las universidades públicas registraron una tasa de deserción del 13,8%, mientras que en las universidades privadas fue del 27,9%. (Teleamazonas, 2023). Este fenómeno afecta tanto el rendimiento académico como los recursos institucionales, ya que la salida de los estudiantes interrumpe su proceso de aprendizaje y representa una pérdida significativa para las instituciones educativas. A nivel nacional, también tiene un impacto negativo, ya que limita el desarrollo social y económico al reducir la cantidad de

personas con educación superior. Por ello, el estudio de la deserción resulta clave para implementar estrategias que mejoren la retención estudiantil.

2.3.1.2 Factores que influyen en la deserción estudiantil

Según Espejo (2012) "la trayectoria escolar, de logro o abandono, está influenciada por factores de diversa índole, que corresponden fundamentalmente a los entornos familiar y escolar" (p.116). Estos factores socioeconómicos, académicos y personales afectan directamente la permanencia o deserción de los estudiantes. Por lo tanto, dentro del fenómeno de la deserción es esencial identificar y abordar estos factores, con el fin de mejorar la retención estudiantil.

Tabla 1

Factores que influyen en la deserción estudiantil

Factores	Indicadores	Descripción
Socioeconómicos	Ingresos familiares	Condiciones económicas, laborales y sociales que afectan la capacidad del estudiante para continuar estudiando.
	Nivel educativo de los padres	
	Situación laboral del estudiante	
	Acceso a becas o ayudas económicas	
	Condiciones de vivienda	
Académicos	Rendimiento académico	Aspectos relacionados con el desempeño académico y la experiencia educativa del estudiante en la institución.
	Asistencia a clases	
	Carga académica	
	Satisfacción con la carrera	
Personales y Psicológicos	Motivación personal	Factores individuales que influyen en el bienestar emocional, las decisiones personales y la motivación del estudiante.
	Salud mental	
	Expectativas personales	
	Apoyo familiar y social	
	Autonomía personal	

Fuente: *Elaboración Propia*

2.3.2 Modelos matemáticos para la predicción de la deserción estudiantil

2.3.2.1 Modelo Matemático

Los modelos matemáticos juegan un papel crucial en la comprensión y análisis de fenómenos complejos, permitiendo traducir situaciones del mundo real en términos matemáticos. Según Blomhøj (2008) "un modelo matemático es una relación entre ciertos objetos matemáticos y sus conexiones por un lado, y por el otro, una situación o fenómeno de naturaleza no matemática" (p.21). Esto significa que los modelos matemáticos establecen una correspondencia entre elementos matemáticos, como ecuaciones y variables, y las características de un problema real, facilitando así su estudio y resolución.

El modelado predictivo consiste en diversas técnicas que se emplean para desarrollar y utilizar modelos que permiten hacer predicciones a partir de patrones identificados en los datos (Kelleher & D`Arcy, 2015).

2.3.2.2 Técnicas de aprendizaje supervisado

Los métodos predictivos, conocidos como aprendizaje supervisado, buscan identificar relaciones entre variables independientes y una variable dependiente. A través de estas relaciones, se construye un modelo que describe y explica un fenómeno subyacente en los datos. Este modelo se utiliza para predecir el valor de la variable dependiente a partir de los valores de las variables independientes. Los modelos supervisados se clasifican en dos categorías: modelos de clasificación y modelos de regresión (Azoumana, 2013).

2.3.2.3 Importancia de los modelos matemáticos en la predicción educativa

El uso de modelos predictivos tiene un gran potencial en el ámbito educativo, permitiendo anticipar situaciones como el abandono escolar o identificar patrones ideales en los cursos. El núcleo del análisis predictivo radica en desarrollar o entrenar modelos que permitan realizar predicciones basadas en datos históricos, empleando para ello técnicas de aprendizaje automático. La eficacia del aprendizaje automático en estos modelos se debe a su capacidad para generalizar los resultados obtenidos (Young & Lee, 2019).

2.3.3 Modelos matemáticos predictivos comunes

2.3.3.1 Árboles de Decisión

Un árbol de decisión es un diagrama que contiene: un nodo raíz donde se encuentran todas las observaciones; nodos internos que albergan los nodos de división, y los nodos hoja que contienen la clasificación final para un conjunto de observaciones (Khalilian, Mustapha, Sulaiman, & Mamat, 2011). Estos diagramas se utilizan como modelos predictivos que emplean una estructura de árbol para dividir iterativamente los datos en grupos más pequeños según criterios específicos, hasta alcanzar una clasificación final clara y precisa.

Beneficios:

- **Facilidad de interpretación:** representación visual clara y sencilla de las reglas de decisión.
- **No linealidad:** puede capturar relaciones complejas entre las variables.
- **Versatilidad:** maneja tanto datos categóricos como numéricos.

Limitaciones:

- Menor precisión en comparación con modelos más robustos (como regresión logística o redes neuronales) si el conjunto de datos no es suficientemente grande.
- Sensible al ruido y sobreajuste, especialmente con conjuntos de datos más pequeños o de tamaño moderado.

Los árboles de decisión aportan de manera significativa en la investigación ya que el objetivo principal es explorar reglas claras para clasificar estudiantes en categorías de riesgo, aunque su rendimiento puede ser limitado con datos moderados.

2.3.3.2 Redes Neuronales

Las redes neuronales artificiales (RNAs) son modelos computacionales diseñados como un intento de realizar formalizaciones matemáticas de la estructura del cerebro. Las RNAs imitan la arquitectura del sistema nervioso y se centran en el funcionamiento del cerebro humano, basado en el aprendizaje a través de la experiencia (Rivera Vergaray, 2021).

Se trata de modelos computacionales inspirados en el cerebro humano, que utilizan múltiples capas de nodos interconectados para identificar patrones complejos y no lineales en los datos.

Beneficios y limitaciones:

Las redes neuronales tienen un alto desempeño predictivo, puesto que son capaces de capturar las relaciones complejas lineales y no lineales entre las variables independientes y dependientes. Sin embargo, necesita de grandes volúmenes de información para obtener un buen desempeño. (Vásquez, 2016.)

Aunque puede proporcionar predicciones precisas, no es ideal debido al tamaño del conjunto de datos y la necesidad de interpretar los factores que influyen en la deserción.

2.3.3.3 Regresión Logística

La Regresión Logística, o Logistic Regression (LR) en inglés, es un caso especial de regresiones cuyo uso es para predecir el resultado de una variable dependiente categórica. Tiene bastante uso en los cálculos de probabilidades, donde se predice la ocurrencia de un evento en función de otros factores (Vásquez, 2016).

Beneficios:

- Interpretabilidad: permite identificar las variables que tienen un mayor impacto en la probabilidad de deserción.
- Eficiencia: requiere menos recursos computacionales que otros modelos.
- Aplicabilidad: funciona bien con conjuntos de datos de tamaño moderado como el que se utiliza en la investigación en el ITSCO.
- Probabilidad explícita: proporciona una probabilidad directa de ocurrencia del evento (probabilidad de que un estudiante deserte).

Ideal para analizar un conjunto de datos de tamaño moderado con un enfoque en identificar factores de riesgo y predecir la deserción.

2.3.3.4 Elección del modelo predictivo

A continuación se detalla un análisis de comparación entre los modelos matemáticos que permiten la predicción de la deserción educativa en el Instituto Superior Cotacachi destacando los aportes de la regresión logística para este trabajo:

- Tamaño del conjunto de datos: la regresión logística es ideal para conjuntos de datos moderados, ya que no requiere una gran cantidad de datos para ser efectiva.
- Interpretabilidad: es crucial entender qué factores contribuyen a la deserción (como ingresos familiares, rendimiento académico o pérdida de una materia). La regresión logística permite identificar y cuantificar el impacto de cada variable independiente.
- Simplicidad y recursos: comparada con redes neuronales, la regresión logística requiere menos recursos computacionales y es más fácil de implementar y ajustar.
- Robustez y generalización: mientras que los árboles de decisión son propensos al sobreajuste, especialmente con conjuntos de datos más pequeños, la regresión logística tiende a generalizar mejor.
- Probabilidades directas: proporciona una probabilidad directa de deserción, lo que facilita la toma de decisiones y la priorización de intervenciones en estudiantes con alto riesgo.

En la presente investigación se utiliza la regresión logística como el modelo más adecuado, gracias a su equilibrio entre interpretabilidad, robustez y capacidad predictiva al trabajar con conjuntos de datos de tamaño moderado. Si bien otros enfoques, como los árboles de decisión o las redes neuronales, pueden ser útiles en diferentes contextos, la regresión logística se presenta como una herramienta eficaz para estimar la deserción estudiantil y orientar estrategias de intervención fundamentadas en datos en el Instituto Tecnológico Superior Cotacachi.

2.3.4 Regresión Logística

La regresión logística es una herramienta estadística de análisis multivariado que se utiliza tanto para fines explicativos como predictivos. Es especialmente útil cuando la variable dependiente es dicotómica, es decir, puede tomar dos valores, como 0 para ausencia y 1 para presencia, y cuando las variables independientes o predictoras son cuantitativas o categóricas (Chitarroni, 2002).

Es un modelo estadístico de predicción utilizado para estimar la probabilidad de ocurrencia de un evento binario (como la deserción estudiantil) en función de una o más variables independientes.

2.3.4.1 Coeficientes del modelo

La regresión logística consiste en obtener una función lineal de las variables independientes que permita clasificar a los individuos en una de las dos subpoblaciones o grupos por los dos valores de la variable dependiente. Un modelo de regresión logística es un modelo que permite estudiar si dicha variable binaria depende de otra/s variable/s. Consecuentemente, la función lineal es el logaritmo de la figura 1 donde β es constante y x_k las variables independientes, dando lugar al modelo logístico múltiple (Pérez & Santin, 2007).

Figura 1

Función lineal del modelo de regresión logística

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$ <p>“β_0” y “β_k” son los coeficientes estimados a partir de los datos X_k son las variables independientes</p>
--

Fuente: Elaboración Propia

A partir de los coeficientes de regresión (β) de las variables independientes en el modelo, es posible calcular directamente la razón de probabilidades (odds ratio) para cada variable. Esta medida representa el riesgo de que ocurra un determinado resultado o efecto cuando la variable independiente toma un valor específico, en comparación con el riesgo cuando dicho valor se reduce en una unidad. En otras palabras, la odds ratio indica cuánto cambia la probabilidad de que ocurra el evento cuando la variable independiente aumenta en una unidad.

2.3.4.2 Definición y principios básicos

En el contexto del análisis predictivo y la investigación de deserción estudiantil, la regresión logística es una técnica estadística fundamental utilizada para modelar la probabilidad de un evento binario, como la deserción o la permanencia de un estudiante. Los principios básicos de esta técnica son los siguientes:

Naturaleza del Modelo: A diferencia de la regresión lineal, la regresión logística está diseñada para variables dependientes dicotómicas (binarias), donde la variable de interés toma dos valores, generalmente codificados como 0 o 1. En este caso, el 1 representa la ocurrencia de un evento, como la deserción, y el 0 la no ocurrencia.

Función Sigmoide (Logit): La regresión logística se basa en la función logística, también conocida como la función sigmoide, que convierte cualquier valor real en un rango entre 0 y 1. Esto permite modelar la probabilidad de que ocurra un evento, proporcionando un ajuste adecuado a problemas donde la relación entre las variables no es necesariamente lineal.

Odds y Log-Odds: El modelo de regresión logística estima la probabilidad de ocurrencia del evento en términos de odds (probabilidades). El odds se define como la razón

entre la probabilidad de que el evento ocurra y la probabilidad de que no ocurra. La regresión logística modela el logaritmo de las odds (log-odds) como una función lineal de las variables predictoras.

Coefficientes de Regresión: Los coeficientes estimados en el modelo de regresión logística representan el cambio en el log-odds de la probabilidad del evento por cada unidad de cambio en las variables independientes. Un coeficiente positivo implica un aumento en las probabilidades del evento, mientras que un coeficiente negativo indica una disminución.

Estadística de Verosimilitud: El ajuste del modelo se evalúa mediante la función de verosimilitud, que mide la probabilidad de observar los datos dados los parámetros estimados. La regresión logística busca maximizar esta función para encontrar los coeficientes que mejor ajusten el modelo a los datos.

Métricas de Desempeño: Una vez entrenado el modelo, su desempeño se evalúa mediante métricas como precisión, recall, y F1-score. Estas métricas permiten medir la capacidad del modelo para predecir correctamente las ocurrencias y no ocurrencias del evento. Adicionalmente, se utilizan técnicas como la matriz de confusión y el área bajo la curva ROC (Receiver Operating Characteristic) para evaluar la calidad de las predicciones.

Interpretación de Resultados: En la regresión logística, los coeficientes deben interpretarse en términos de probabilidades. Por ejemplo, un coeficiente de 0.5 indica que, por cada unidad de incremento en la variable predictora, las probabilidades del evento se multiplican por el exponente de 0.5, lo que equivale a aproximadamente un 65% de aumento en las odds.

CAPÍTULO 3

DISEÑO METODOLÓGICO

3.1 Enfoque de la Investigación

El enfoque de la investigación es cuantitativo, ya que se busca desarrollar un modelo matemático basado en técnicas estadísticas para estimar la tasa de deserción estudiantil en el Instituto Tecnológico Superior Cotacachi. Este enfoque permite recolectar y analizar datos numéricos con el fin de identificar patrones y relaciones entre variables socioeconómicas, académicas y demográficas que influyen en la deserción estudiantil.

3.2 Diseño de la Investigación

El diseño de la investigación es no experimental y transversal, debido a que los datos se analizan en un único momento en el tiempo. Se emplea un diseño correlacional para examinar la relación entre las variables que inciden en la deserción y un diseño predictivo para el desarrollo del modelo matemático.

Es así como según Valderrama Mendoza & Jaimes Velasquez (2019)“Las variables estudiadas no se manipulan, sino que se observan en su contexto natural, permitiendo establecer inferencias sobre las relaciones observadas sin intervención directa” (p.67).

3.3 Tipo de Investigación

Se pueden establecer varios factores que permiten tipificar la investigación, detallados a continuación.

Aplicada: existen varios trabajos similares al tema que se pretende investigar, por lo que, los mismos sirven de base para enriquecer el presente estudio, permitiendo reestructurar

y mejorar las investigaciones anteriores, constituyéndose en un estudio aplicado. (Niño Rojas, 2011).

Documental: las fuentes secundarias utilizadas en la descripción teórica, legal y referencial de la presente investigación permiten que sea de tipo documental a fin de analizar y utilizar como base de estudio la información encontrada en la revisión bibliográfica. (Bernal, 2010).

No Experimental: las variables en este estudio no serán intervenidas, se estudia el problema de deserción en sus condiciones normales, es decir, sin manipular ninguno de los factores, por ello el estudio es de tipo no experimental. (Bernal, 2010).

3.4 Nivel de Investigación

El nivel de la investigación es exploratorio y descriptivo. Exploratorio al examinar variables que pueden no haber sido estudiadas en profundidad previamente en el contexto específico del ISTCO. Descriptivo porque se busca describir las características y factores que contribuyen a la deserción estudiantil.

3.5 Técnicas e Instrumentos de Recolección de Datos

La recolección de datos es una fase crucial que proporciona la base para el análisis y modelado. Para esta investigación se realiza lo siguiente:

- Fuentes de datos: se utilizan datos históricos de los estudiantes del Instituto Tecnológico Superior Cotacachi, que incluyen variables socioeconómicas, académicas y demográficas.
- Instrumentos de recolección: se emplean bases de datos internas del instituto.

- Periodo de recolección: la recolección abarca los datos de los últimos tres años académicos para asegurar una muestra representativa.

3.6 Técnicas para el Procesamiento e Interpretación de Datos

Figura 2

Proceso metodológico



Fuente: Elaboración Propia

El procesamiento e interpretación de los datos (figura 2) se lleva a cabo mediante:

Recolección de datos

Se recolectaron 2,282 registros de matrículas correspondientes a estudiantes de los últimos cinco periodos académicos (primer periodo de 2022 al primer periodo de 2024) obtenidos del departamento de la base de datos del ITSCO separados por cada semestre, incluyendo variables categorizadas en socioeconómicas, académicas entre otras.

Depuración de datos

Para garantizar la calidad y consistencia de los datos, se implementaron las siguientes acciones:

- Eliminación de redundancias: se depuraron registros repetidos, resultando en un conjunto final de 1,385 registros únicos.
- Escalamiento de variables: las variables numéricas se transformaron para mantener un rango uniforme, utilizando técnicas como la normalización (escalamiento a rango $[0,1)$).
- Manejo de datos faltantes: se aplicaron estrategias de imputación de valores (media para variables continuas y moda para categóricas) en casos con proporciones bajas de datos ausentes. Para variables con alta cantidad de datos faltantes, se excluyeron del análisis.
- Transformación de variables categóricas: variables cualitativas fueron codificadas mediante métodos como one-hot encoding para su incorporación en el modelo.

Análisis descriptivo

En esta etapa, se exploraron las propiedades de las variables utilizando técnicas estadísticas como análisis de distribuciones, detección de valores atípicos y cálculo de medidas de tendencia central y dispersión. De igual manera se clasificaron las variables según su naturaleza (numérica o categórica).

Modelado predictivo

Se implementó un modelo de regresión logística binaria, el cual estima la probabilidad de deserción estudiantil (1: deserta, 0: no deserta).

- La selección de variables independientes se realizó con base en análisis de correlación y pruebas de significancia estadística.
- Se empleó una división de datos para entrenamiento y validación (70% - 30% respectivamente) y se ejecutó un procedimiento iterativo para optimizar los coeficientes del modelo.
- El modelo integró predictores como nivel de ingresos, promedio académico, y frecuencia de matrícula.

Validación del modelo

El modelo fue evaluado utilizando métricas estadísticas robustas:

- Precisión global: tasa de predicciones correctas.
- Sensibilidad y especificidad: habilidad para identificar correctamente casos de deserción y no deserción.
- Curva ROC y AUC: indicadores del rendimiento global del modelo.
- Precisión promedio con validación cruzada.
- Ajustes iterativos se realizaron para mejorar la robustez del modelo y minimizar errores en las predicciones.

Propuesta de Estrategias

A partir de los resultados del modelo, se diseñaron estrategias enfocadas en los factores críticos detectados:

- Apoyo académico: programas de tutorías dirigidas a estudiantes con bajo rendimiento.
- Asistencia económica: becas y financiamiento para estudiantes con condiciones económicas desfavorables.
- Fortalecimiento administrativo: optimización de los procesos de registro y comunicación institucional. Estas estrategias buscan prevenir la deserción mediante intervenciones focalizadas, basadas en el análisis predictivo.

3.7 Población y Muestra

3.7.1 Población

La población de estudio corresponde al periodo comprendido entre el primer semestre de 2022 y el primer semestre de 2024, conformada por 2,282 registros de matrícula correspondientes a diversos programas educativos.

3.7.2 Tamaño de la Muestra

La muestra de estudio se derivó de la depuración de la población inicial de 2,282 registros de matrículas, obteniendo 1,385 registros únicos que representan la base consolidada para el análisis.

La representatividad y fiabilidad de la muestra se asegura mediante la inclusión de registros de estudiantes de diversos programas académicos y niveles socioeconómicos,

distribuidos uniformemente a lo largo de los cinco periodos analizados. Este enfoque minimiza posibles sesgos asegurando la validez estadística y precisión del modelo predictivo utilizado en el estudio.

CAPÍTULO 4

ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS

4.1 Análisis Descriptivo de los Resultados

El análisis realizado en la base de datos, que comprende 2,281 registros de matrículas de estudiantes de los últimos cinco periodos (primer periodo de 2022 hasta el primer periodo de 2024), permite entender las variables más relevantes en la predicción de la deserción estudiantil. El preprocesamiento de la información incluye la agrupación por estudiante, la eliminación de datos de identificación personal y la categorización de variables ordinales y binarias. Además, se normalizaron variables cuantitativas como ingresos económicos, porcentaje de discapacidad y número de integrantes del núcleo familiar.

Tabla 2

Categorización de variables del modelo

Variable	Categorías	Valor
Sexo	Mujer	0
	Hombre	1
Género	Femenino	0
	Masculino	1
Estado Civil	Casado	0
	Divorciado	1
	Soltero	2
	Unión de hecho	3
	Viudo	4
Etnia	Mestizo	0
	Indígena	1
	Negro	2
	Mulato	3

	Afroecuatoriano	4
	Montubio	5
	Blanco	6
Pueblo indígena	No aplica	0
	Otavalo	1
	Kichwa	2
	Puruhá	3
	Awá	4
	Natabuela	5
	Karanki	6
	Kayampi	7
	Tipo de Sangre	A RH (+)
O RH (+)		1
O RH (-)		2
B RH (+)		3
AB RH (-)		4
AB RH (+)		5
A RH (-)		6
B RH (-)		7
Discapacidad		No
	Sí	1
Porcentaje de discapacidad	No aplica	0
	38%	1
	70%	2
	0%	3
	45%	4
	(valores referencia)	
Tipo de discapacidad	No aplica	0
	Auditiva	1
	Intelectual	2
	Visual	3

	Física	4
Nacionalidad	Ecuador	0
	Italia	1
	(valores referencia)	
Provincia nacimiento	Bolívar	0
	Carchi	1
	Imbabura	2
	(valores referencia)	
Cantón nacimiento	Echeandía	0
	Montúfar	1
	Tulcán	2
	(valores referencia)	
Provincia residencia	Imbabura	0
	Carchi	1
	Pichincha	2
	(valores referencia)	
Cantón residencia	Cotacachi	0
	Ibarra	1
	Otavalo	2
	(valores referencia)	
Tipo de colegio	Fiscal	0
	Particular	1
	(valores referencia)	2
Modalidad de estudios	Presencial	0
	Dual	1
Jornada de estudios	Vespertina	0
	Intensiva	1
	Matutina	2
Nivel	Primero	0
	Segundo	1
	Tercero	2
	Cuarto	3

	Quinto	4
Repetición de curso	No	0
	Sí	1
Pérdida de gratuidad	No	0
	Sí	1
Percepción de dificultad académica	No	0
	Sí	1
Dedicación al estudio	Trabaja y estudia	0
	Solo estudia	1
Destino del ingreso estudiantil	No aplica	0
	Mantener hogar	1
	Financiar estudios	2
	Gastos personales	3
Recibe bono familiar	No	0
	Sí	1
Nivel de educación del padre	No aplica	0
	Centro de alfabetización	1
	Cardín de infantes	2
Nivel de educación de la madre	Educación básica	3
	(valores referencia)	6
Ingreso económico del hogar	386	0.08
	1,000	0.2
	(valores referencia)	
Número de integrantes de la familia	1	0
	18	1
Edad	18	0
	59	1

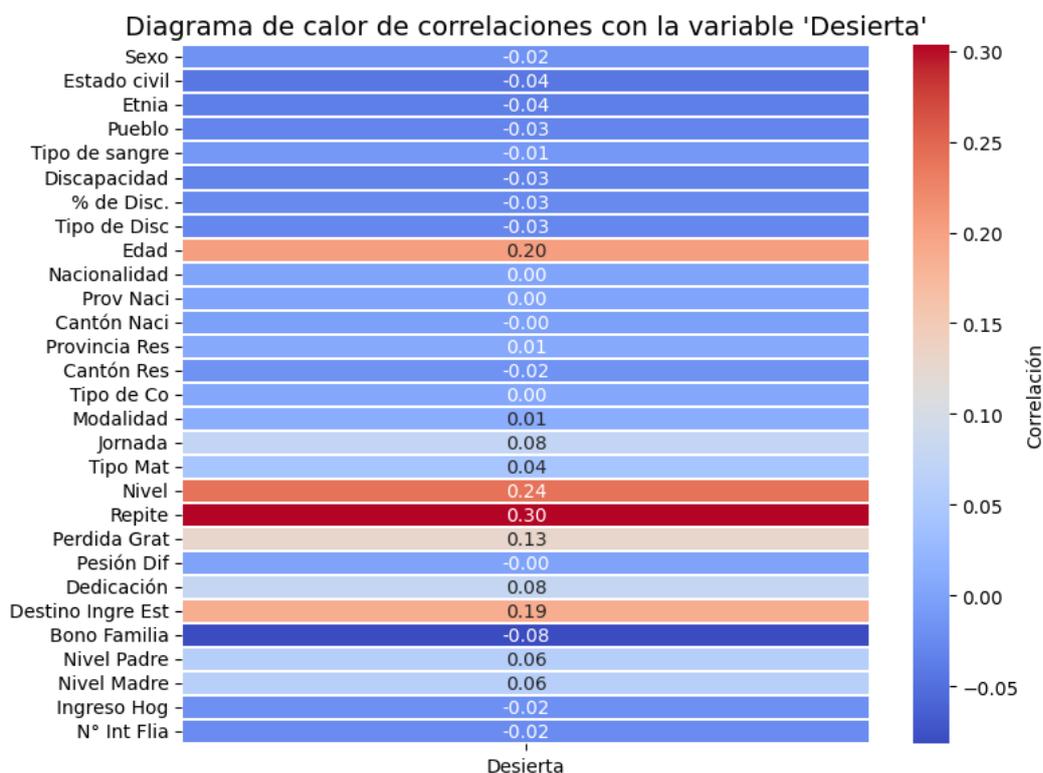
Fuente: Elaboración Propia

Este cuadro resume las variables con las categorías y sus valores correspondientes, algunas categorías tienen múltiples valores, como "Porcentaje de Discapacidad" e "Ingreso Económico," por lo que los detalles completos se han mantenido para referencia.

4.2 Análisis Correlacional

Figura 3

Mapa de calor



Fuente: Elaboración Propia

Los resultados presentados en la matriz de correlación inicial donde el coeficiente "r" indica la relación entre las variables socioeconómicas, académicas y la deserción estudiantil, muestran que mayor influencia positiva en el fenómeno de estudio presentan la repetición de materias ($r=0.303$), el nivel educativo ($r=0.240$), la edad normalizada ($r=0.201$) y la pérdida de

gratuidad ($r=0.128$). En cambio, el sexo ($r=-0.018$), los ingresos familiares ($r=-0.019$) y el nivel de instrucción de los padres ($r=-0.060$ y $r=-0.061$) presentan correlaciones débiles, indicando una influencia mínima.

4.3 Modelo de Regresión Logística

Tabla 3

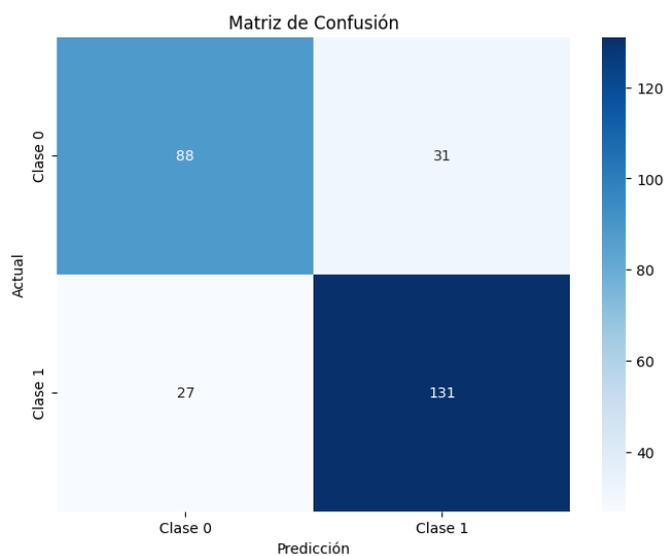
Resumen de métricas del modelo de regresión logística

Class	Precision	Recall	F1-Score	Support
0	0.77	0.74	0.75	119
1	0.81	0.83	0.82	158
accuracy			0.79	277
macro avg	0.79	0.78	0.79	277
weighted avg	0.79	0.79	0.79	277

Fuente: Elaboración propia

Figura 4

Matriz de Confusión y Desempeño del Modelo

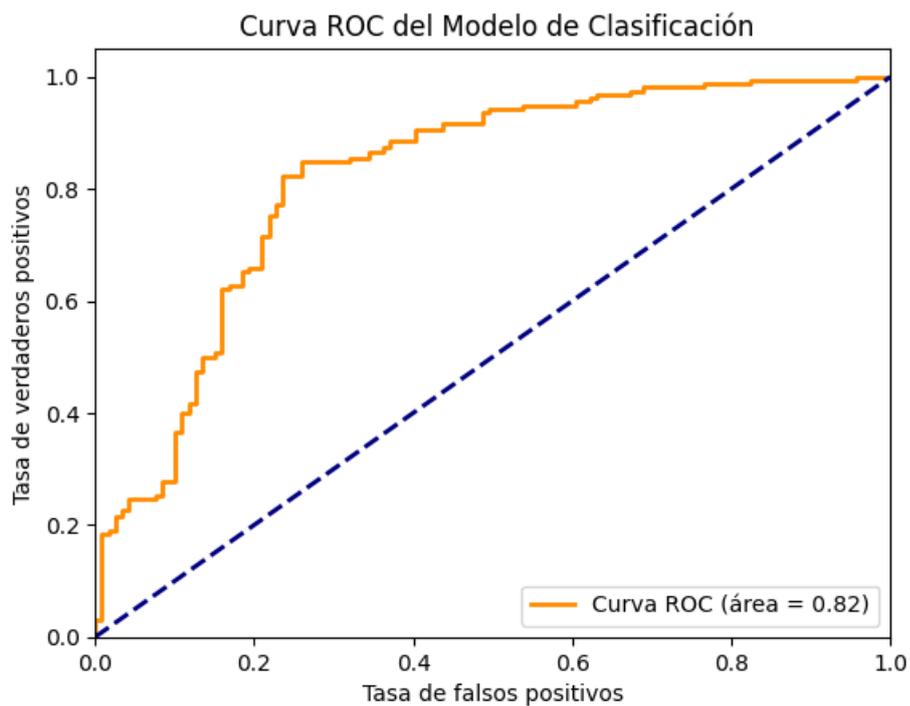


Fuente: Elaboración Propia

4.4 Validación Cruzada

Figura 5

Curva ROC del Modelo de Clasificación



Fuente: Elaboración Propia

Tabla 4

Validación Cruzada

	Valor
Precisión con validación cruzada	0.7342960288808664
Área curva ROC	0.82

Fuente: Elaboración propia

El modelo de regresión logística desarrollado para predecir la deserción estudiantil en el Instituto Superior Cotacachi alcanza una precisión global del 79% al clasificar 277 casos,

identificando correctamente el 83% de las deserciones (clase 1) y el 74% de los casos de no deserción (clase 0). Para la clase 1, el modelo destaca con una precisión del 81%, un recall del 83% y un F1-score del 82%, lo que demuestra su eficacia en la detección de deserciones. En la clase 0, aunque el rendimiento es ligeramente inferior, la precisión del 77%, el recall del 74% y el F1-score del 75% reflejan un desempeño consistente. Los promedios macro y ponderados de las métricas alcanzan el 79%, evidenciando un balance razonable entre las dos clases. La matriz de confusión muestra 31 falsos positivos (intervenciones innecesarias) y 27 falsos negativos (oportunidades perdidas de intervención), destacando la necesidad de priorizar la reducción de estos últimos para mitigar el impacto directo en las tasas de abandono.

Al comparar las métricas del modelo con los resultados obtenidos mediante validación cruzada, se observa una precisión promedio más baja de 73.4%, lo que sugiere variabilidad en el rendimiento del modelo al evaluarlo en diferentes subconjuntos de datos. Esta discrepancia destaca la importancia de evaluar la robustez del modelo frente a datos no vistos. Adicionalmente, el área bajo la curva ROC (0.82) reafirma la capacidad discriminativa del modelo, indicando que puede distinguir de manera efectiva entre estudiantes que desertan y los que no. Estos resultados, en conjunto, reflejan un desempeño satisfactorio, pero también señalan oportunidades de mejora, como ajustar el umbral de clasificación o considerar variables adicionales para fortalecer la capacidad predictiva y reducir errores críticos.

Tabla 5

Coefficientes de regresión logística

	coeficiente	Error estándar	z	P> z
Const	-4.4285	0.639	-6.935	0
Sexo	-0.2753	0.171	-1.607	0.108

Estado civil	0.1934	0.144	1.339	0.181
Etnia	0.0177	0.106	0.168	0.867
Pueblo	0.0168	0.055	0.306	0.76
Tipo sangre	-0.1019	0.085	-1.201	0.23
Discapacidad	-1.0591	1.389	-0.763	0.446
% discapacidad	0.5449	2.026	0.269	0.788
Tipo disc	0.018	0.413	0.044	0.965
Edad	3.8731	0.689	5.621	0
Nacionalidad	-0.0215	0.337	-0.064	0.949
Prov. Nacimiento	-0.2613	0.171	-1.532	0.126
Can. Nacimiento	0.0751	0.057	1.32	0.187
Prov. Residencia	0.0602	0.216	0.279	0.781
Can Residencia	-0.0792	0.059	-1.336	0.182
Tipo de colegio	0.0716	0.089	0.803	0.422
Modalidad	0.4508	0.17	2.646	0.008
Jornada	0.2191	0.097	2.251	0.024
Tip Map	0.2039	0.374	0.546	0.585
Nivel	0.4644	0.055	8.408	0
Repite	1.8693	0.185	10.098	0
Perdida Gratuidad	1.1955	0.285	4.188	0
Pesión Dif	-0.0161	0.659	-0.024	0.981
Dedicación	-0.5268	0.183	-2.875	0.004
Dest Ingre Est	0.4347	0.083	5.257	0
Bono Familia	-0.4229	0.233	-1.818	0.069
Nivel Padre	0.0265	0.052	0.509	0.611
Nivel Madre	0.1099	0.053	2.078	0.038
Ing Hog	-0.7269	1.135	-0.641	0.522
N° Int Flia	0.0119	0.038	0.312	0.755

Fuente: Elaboración propia

El análisis de los coeficientes de la regresión logística indica que las variables con mayor impacto sobre la probabilidad de deserción son: Edad, Repetición de asignaturas,

Pérdida de gratuidad y la Modalidad de estudio, todas con una asociación positiva y estadísticamente significativa con el abandono. Además, la dedicación al estudio muestra un coeficiente negativo y significativo lo que sugiere que solo estudiar sin otras ocupaciones disminuye el riesgo de deserción. El nivel educativo de la madre también se asocia positivamente con la retención, mientras que los canales de ingreso al estudio aumentan la probabilidad de deserción. En contraste, variables como Sexo, Estado civil, Etnia y otros factores socioeconómicos no presentan significancia estadística en este modelo, lo que sugiere que su influencia en la deserción es limitada en este contexto particular. Estos factores no parecen ser determinantes en la decisión de los estudiantes de abandonar los estudios en el Instituto Tecnológico Superior Cotacachi.

A continuación se detalla un resumen con la interpretación de valores de los coeficientes del modelo de regresión logística y su vínculo con el diseño de estrategias para reducir la deserción estudiantil en el ITSCO.

Tabla 6

Resumen de métricas del modelo

Variable	Coeficiente	Valor p	Interpretación	Estrategia Propuesta
Edad	3.87	0	Estudiantes con mayor edad tienen mayor probabilidad de desertar.	Diseñar programas de apoyo específicos para estudiantes con responsabilidades.
Nivel	0.46	0	Estudiantes en niveles avanzados tienen mayor probabilidad de desertar.	Implementar tutorías y orientación en los últimos semestres.
Repite	1.87	0	Repetir materias es un factor crítico para la deserción.	Crear programas de recuperación y refuerzo académico temprano.

Pérdida de gratuidad	1.19	0	Perder la gratuidad educativa aumenta el riesgo de deserción.	Asegurar que los estudiantes conozcan y mantengan los requisitos de gratuidad.
Dedicación	-0.52	0.004	Dedicación exclusiva a los estudios reduce la probabilidad de deserción.	Apoyar económicamente a estudiantes con trabajos para que puedan estudiar más.
Destino del ingreso estudiantil	0.43	0	Estudiantes que financian sus estudios tienen mayor probabilidad de desertar.	Diseñar programas de políticas de apoyo y talleres de educación financiera.

Fuente: Elaboración propia

Al realizar el modelado matemático se obtienen los valores que permiten estructurar la ecuación para el logaritmo de las probabilidades en la regresión logística.

Ecuación para el logaritmo de las probabilidades en la regresión logística.

$$\log\left(\frac{p}{1-p}\right) = -0.1019 \cdot \text{TipSangre} - 1.0591 \cdot \text{Discap} + 0.5449 \cdot \text{PorDiscap} + 0.0180 \cdot \text{TipDisc} + 3.8731 \cdot \text{Edad} - 0.0215 \cdot \text{Nacionalidad} - 0.2613 \cdot \text{ProvNaci} + 0.0751 \cdot \text{CanNaci} + 0.0602 \cdot \text{ProvRes} - 0.0792 \cdot \text{CanRes} + 0.0716 \cdot \text{TipColegio} + 0.4508 \cdot \text{Modalidad} + 0.2191 \cdot \text{Jornada} + 0.2039 \cdot \text{TipMap} + 0.4644 \cdot \text{Nivel} + 1.8693 \cdot \text{Repite} + 1.1955 \cdot \text{PerdiGrat} - 0.0161 \cdot \text{PesiDif} - 0.5268 \cdot \text{Dedicacion} + 0.4347 \cdot \text{DestIngreEst} - 0.4229 \cdot \text{BonoFamilia} + 0.0265 \cdot \text{NivelPadre} + 0.1099 \cdot \text{NivelMadre} - 0.7269 \cdot \text{IngHog} + 0.0119 \cdot \text{NumIntFam} - 4.4285 - 0.2753 \cdot \text{Sexo} + 0.1934 \cdot \text{Estciv} + 0.0177 \cdot \text{Etnia} + 0.0168 \cdot \text{Pueblo}$$

Tabla 7**Resultados de la regresión logística**

Optimización exitosa			
Valor de la función	0.546899		
Número de iteraciones	6		
Parámetro	Valor	Parámetro	Valor
Variable dependiente	Desierta	Número de Observaciones	1108
Modelo	Logit	Grados de Libertad Residual	1078
Método	MLE	Grados de Libertad del Modelo	29
Fecha	Lun, 30 Sep 2024	Pseudo R-cuadrado	0.2094
Hora	23:55:37	Log-Likelihood	-605.96
Convergió	Verdadero	LL-Null	-766.49
Tipo de covarianza	No robusto	Valor p de LLR	5.43E-51

Fuente: Elaboración Propia

El modelo de regresión logística utilizado para estimar la deserción estudiantil ha convergió satisfactoriamente tras seis iteraciones, obteniendo una log-verosimilitud de -605.96. El pseudo R-cuadrado de 0.2094 sugiere que el modelo captura aproximadamente el 21% de la varianza en la variable dependiente, lo que implica una capacidad predictiva moderada. Este valor indica que, aunque el modelo es adecuado, podrían existir otros factores no incluidos que influyan en la deserción.

El estadístico de la razón de verosimilitud (LLR) muestra una significancia estadística muy alta ($p = 5.426e-51$), lo que confirma que los predictores seleccionados tienen un efecto significativo sobre la probabilidad de deserción estudiantil. En resumen, el modelo presenta un ajuste estadísticamente robusto, pero sugiere la necesidad de explorar otras variables adicionales para optimizar la capacidad predictiva en el contexto estudiado.

4.5 Discusión de los Resultados

El modelo para la predicción de la deserción estudiantil en el Instituto Tecnológico Superior Cotacachi evidencia la pertinencia de ciertos factores como predictores clave ha demostrado ser adecuado para identificar variables críticas que influyen en la deserción, tales como la edad, la repetición de asignaturas, la pérdida de gratuidad, y la modalidad de estudio, todas ellas con un impacto significativo en la probabilidad de abandono.

El rendimiento del modelo, con una precisión del 79% y un pseudo R-cuadrado de 0.2094, señala una capacidad predictiva moderada, lo que implica que, aunque se ha capturado una parte significativa de la variabilidad en la deserción, aún quedan factores no modelados que podrían explicar mejor el fenómeno. La convergencia del modelo en seis iteraciones y el valor altamente significativo del estadístico de verosimilitud (LLR p-value de 5.426e-51) confirman que el conjunto de predictores utilizados en el análisis tiene una influencia significativa sobre la variable dependiente. Sin embargo, estos resultados también reflejan la necesidad de incorporar variables adicionales que podrían estar relacionadas con factores psicológicos, pedagógicos o de infraestructura que impactan la retención estudiantil.

Consideraciones Finales

El análisis subraya la importancia de focalizar políticas institucionales que aborden las causas subyacentes identificadas, como el apoyo a estudiantes que repiten materias, la mejora en modalidades educativas y la preservación de beneficios financieros. En conclusión, aunque el modelo muestra un rendimiento robusto dentro de sus límites actuales, su capacidad predictiva podría incrementarse con la inclusión de nuevas variables que aborden otros aspectos críticos en la deserción.

CAPÍTULO 5

MARCO PROPOSITIVO

5.1 Planificación de la Actividad Preventiva

En base a los hallazgos del estudio de deserción estudiantil en el Instituto Tecnológico Superior Cotacachi, se plantea una propuesta integral para mitigar el problema de deserción, utilizando enfoques basados en los factores clave identificados en el análisis de datos. Esta propuesta está orientada a la intervención preventiva, tomando en cuenta los aspectos académicos, financieros y psicosociales que se identificaron como críticos en el modelo predictivo desarrollado. El enfoque de la solución propuesta combina la implementación de políticas institucionales y herramientas tecnológicas para mejorar la retención estudiantil.

5.1.1 *Actividades Preventivas*

5.1.1.1 Programa de seguimiento académico

- **Objetivo:** reducir la deserción mediante un programa de seguimiento académico temprano, con especial énfasis en los estudiantes que repiten asignaturas o están en riesgo de perder beneficios económicos.

Según los resultados obtenidos, se recomienda implementar un sistema de seguimiento personalizado para los estudiantes que, de acuerdo con el modelo predictivo, presentan una mayor probabilidad de desertar. Este programa incluirá:

- **Tutorías académicas:** asignación de tutores a estudiantes en riesgo, con seguimiento mensual de su progreso académico y asistencia a clases.

- **Plan de recuperación de asignaturas:** establecer un esquema de recuperación de materias que permita a los estudiantes que repiten asignaturas avanzar en sus estudios sin afectar su carga académica.
- **Plataforma de alertas tempranas:** desarrollo de una herramienta tecnológica para identificar a tiempo estudiantes en riesgo, basada en su rendimiento académico, asistencia y dedicación al estudio.

5.1.1.2 Programa de apoyo financiero y orientación socioeconómica

- **Objetivo:** Mantener la estabilidad económica de los estudiantes mediante un programa de orientación y apoyo financiero.

El análisis mostró que la pérdida de gratuidad es un factor significativo en la deserción, por lo tanto, se propone la creación de un programa de apoyo financiero que incluya:

- **Becas y subsidios:** reforzar el sistema de becas y crear mecanismos de apoyo adicional para aquellos estudiantes que pierdan su gratuidad.
- **Asesoría financiera:** implementar un programa de orientación para ayudar a los estudiantes y sus familias a planificar mejor los costos relacionados con la educación, buscando también alternativas de financiamiento externo o apoyo gubernamental.
- **Talleres de planificación financiera:** ofrecer talleres dirigidos a los estudiantes para mejorar sus habilidades de gestión financiera personal, asegurando que puedan manejar de forma efectiva los recursos disponibles para su educación.

5.1.1.3 Adaptación de modalidades educativas y flexibilidad curricular

- **Objetivo:** facilitar el acceso a diferentes modalidades de estudio que se adapten a las necesidades de los estudiantes, reduciendo así la deserción vinculada a incompatibilidades con el formato educativo actual.

A partir del coeficiente significativo de la variable Modalidad, se propone revisar y adaptar las modalidades de estudio existentes, ofreciendo mayor flexibilidad curricular. Esto se logrará mediante:

- **Aumento de la oferta en modalidades semipresenciales y virtuales:** incrementar la disponibilidad de cursos en modalidades no presenciales para los estudiantes que tienen dificultades para asistir a clases presenciales debido a su situación laboral o personal con la autorización respectiva de los organismos institucionales como el Consejo de Educación Superior (CES).
- **Flexibilidad en la carga académica:** permitir que los estudiantes puedan ajustar su carga académica de acuerdo con sus necesidades, favoreciendo una mayor retención a largo plazo.
- **Revisión de la estructura del programa académico:** implementar un proceso continuo de evaluación y ajuste de los contenidos curriculares para asegurarse de que sean relevantes, actuales y aplicables al entorno laboral de los estudiantes.

5.1.1.4 Programa de fortalecimiento psicosocial

- **Objetivo:** mejorar el bienestar emocional y psicosocial de los estudiantes, asegurando su permanencia en el instituto.

Dado que muchos estudiantes desertan debido a problemas emocionales o psicosociales, se propone establecer un programa de apoyo integral que incluya:

- **Orientación psicológica:** ofrecer sesiones de asesoramiento psicológico gratuitas para estudiantes en riesgo, enfocadas en el manejo del estrés, la ansiedad y la adaptación al entorno educativo.
- **Grupos de apoyo entre pares:** crear grupos de acompañamiento entre estudiantes que estén enfrentando situaciones similares para promover un ambiente de apoyo mutuo y disminuir el aislamiento.
- **Charlas y talleres sobre bienestar emocional:** desarrollar un programa de charlas y talleres enfocados en la salud mental y el manejo del estrés académico, facilitando herramientas para mejorar su resiliencia y bienestar.

5.1.1.5 Evaluación y seguimiento continuo

Para garantizar que las propuestas tengan el impacto deseado, se propone establecer un sistema de evaluación periódica del programa preventivo basado en indicadores de éxito métricos para cada acción tanto académica, socioeconómica y psicosocial para asegurar que las estrategias implementadas sean efectivas y se ajusten en tiempo real según los resultados obtenidos.

De esta manera se engloba en un sistema ordenado de tablas la idea básica del sistema de seguimiento que se adapta a espacio y tiempo según la necesidad, vinculado, extrayendo actividades de retención o modificando los indicadores.

5.1.2 Sistema de Evaluación

Tabla 8

Acciones académicas

Programa/Acción	Indicador de éxito	Meta	Frecuencia de evaluación	Método de medición
Tutorías Académicas	Porcentaje de estudiantes en riesgo atendidos por tutores	90% de estudiantes en riesgo	Trimestral	Registro de asistencia y reportes de progreso académico de tutores
	Tasa de mejora en el rendimiento académico de estudiantes en riesgo	Incremento del 20% en promedio	Semestral	Comparación de calificaciones antes y después del programa
Plan de Recuperación de Asignaturas	Número de asignaturas recuperadas exitosamente	80% de asignaturas recuperadas	Semestral	Registro de aprobaciones en asignaturas de recuperación
	Tasa de retención de estudiantes que recuperaron asignaturas	85% de retención	Anual	Análisis de datos de matrícula

Fuente: Elaboración Propia

Para la evaluación de las acciones preventivas psicosociales el método de medición tiene un carácter cualitativo y cuantitativo reflejado en la flexibilidad de los registros de bienestar estudiantil.

Tabla 9

Acciones psicosociales

Programa/Acción	Indicador de éxito	Meta	Frecuencia de evaluación	Método de medición
Plataforma de Alertas Tempranas	Tiempo promedio de respuesta a alertas de riesgo	≤ 7 días desde la alerta generada	Mensual	Logro de actividad de la plataforma
	Porcentaje de intervenciones exitosas tras alerta	75% de estudiantes intervenidos	Semestral	Análisis de casos gestionados
Modalidades Flexibles de Estudio	Porcentaje de estudiantes inscritos en modalidades no presenciales	Incremento del 10% anual	Anual	Análisis de matrícula en modalidades semipresenciales y virtuales
	Tasa de retención en modalidades flexibles	$\geq 85\%$ de retención	Anual	Registro de matrícula y deserción
Talleres Psicosociales	Número de talleres realizados por semestre	≥ 3 talleres por semestre	Semestral	Registro de actividades realizadas
	Participación promedio por taller	$\geq 70\%$ de asistencia esperada	Semestral	Registro de asistencia
	Cambio percibido en el bienestar emocional de estudiantes	$\geq 75\%$ de estudiantes reportan mejora	Anual	Encuestas pre y post-talleres
Orientación Psicológica	Número de sesiones psicológicas realizadas	Incremento del 20% anual	Trimestral	Registro de citas y reportes de seguimiento
	Tasa de estudiantes que continúan después de intervención psicológica	$\geq 80\%$ de retención	Anual	Análisis de datos de matrícula

Fuente: Elaboración Propia

Tabla 10

Acciones socioeconómicas

Programa/Acción	Indicador de éxito	Meta	Frecuencia de evaluación	Método de medición
Becas y Subsidios	Número de estudiantes beneficiados con apoyo financiero	Incremento del 15% anual	Anual	Reportes financieros y listado de beneficiarios
Becas y Subsidios	Tasa de retención de estudiantes beneficiados	90% de retención	Anual	Análisis de datos de matrícula
Asesoría Financiera	Porcentaje de estudiantes asistiendo a sesiones de asesoría	70% de estudiantes en riesgo económico	Trimestral	Registro de asistencia y encuestas de satisfacción
Asesoría Financiera	Satisfacción de estudiantes con las asesorías	≥80% de satisfacción	Trimestral	Encuestas posteriores a las sesiones

Fuente: Elaboración Propia

Tabla 11

Evaluación general

Programa/Acción	Indicador de éxito	Meta	Frecuencia de evaluación	Método de medición
Evaluación General	Tasa global de retención estudiantil	Incremento del 15% en tres años	Anual	Análisis comparativo de tasas de deserción por periodo
	Índice de satisfacción estudiantil general	≥85% de satisfacción	Anual	Encuestas institucionales

Fuente: Elaboración Propia

5.2 Impacto y Beneficios Esperados

La implementación de esta propuesta permitirá mejorar significativamente las tasas de retención en el Instituto Tecnológico Superior Cotacachi. A través del seguimiento académico

personalizado, el apoyo financiero, la flexibilidad curricular y el fortalecimiento psicosocial, se logrará una disminución en los índices de deserción, especialmente entre los estudiantes más vulnerables. Estas acciones posicionarán a la institución como un referente en la implementación de políticas de retención educativa y apoyo integral al estudiante, logrando excelencia académica.

El plan preventivo propuesto tiene el potencial de reducir significativamente la deserción estudiantil en el Instituto Tecnológico Superior Cotacachi al abordar los principales factores identificados en este estudio. A través de un enfoque integral que incluye la identificación temprana de estudiantes en riesgo, el fortalecimiento del apoyo académico y socioeconómico, y la implementación de un sistema de monitoreo y evaluación constante, se espera que la institución logre mejorar la retención y aumentar el éxito académico con un ambiente educativo inclusivo y flexible que promueva el éxito académico a largo plazo de sus estudiantes.

Conclusiones

El modelo matemático basado en técnicas de análisis multivariado, específicamente la regresión logística, cumplió su objetivo de estimar la tasa de deserción estudiantil en el Instituto Tecnológico Superior Cotacachi, demostrando una capacidad predictiva adecuada con una exactitud del 79%. Este resultado confirma la viabilidad del uso de técnicas estadísticas para abordar problemas educativos complejos.

Los análisis realizados revelaron que las variables como la edad, la repetición de asignaturas y la pérdida de gratuidad son predictores significativos de la deserción, lo que refuerza la relevancia de factores socioeconómicos y académicos en la permanencia de los estudiantes. Por otro lado, variables como el sexo y el estado civil no mostraron significancia estadística en este contexto, indicando la necesidad de refinar los modelos predictivos mediante la incorporación de factores adicionales que puedan capturar mejor la dinámica de la deserción.

La metodología aplicada, mediante la combinación del análisis descriptivo inicial con la validación del modelo predictivo mediante métricas de desempeño (precisión, sensibilidad y F1-score), resultó robusta para la identificación de patrones de deserción. Los resultados de la matriz de confusión evidenciaron un adecuado equilibrio entre sensibilidad y especificidad, lo que confirma la capacidad del modelo para clasificar de manera efectiva a los estudiantes en riesgo y a aquellos con mayor probabilidad de continuar en la institución.

El modelo predictivo desarrollado permitió clasificar con alta precisión a los estudiantes con mayor probabilidad de deserción, al igual que generar insumos técnicos para el diseño de intervenciones específicas. En particular, los resultados evidencian que reforzar el apoyo a estudiantes afectados por la pérdida de gratuidad y establecer un monitoreo focalizado para

quienes presentan repetición de asignaturas podría contribuir significativamente a la reducción de la tasa de deserción, con un impacto estimado del 10 al 15 % en los próximos periodos académicos.

En el estudio realizado en el Instituto Tecnológico Superior Cotacachi se logró diseñar un modelo matemático basado en regresión logística para predecir la deserción estudiantil, cumpliendo con el objetivo general propuesto de igual manera, el análisis descriptivo reveló la influencia de variables claves, evidenciando la relevancia de factores académicos y socioeconómicos en la deserción. El modelo predictivo demostró un desempeño confiable, validándose como una herramienta eficaz para estimar la probabilidad de deserción. De igual manera, la identificación de aspectos críticos evidenciados brinda bases sólidas para diseñar estrategias efectivas que reduzcan el riesgo de abandono. Estos hallazgos subrayan la utilidad de las técnicas analíticas para orientar decisiones institucionales y fortalecer la permanencia estudiantil.

Recomendaciones

Se recomienda desarrollar un sistema automatizado que utilice análisis predictivos para identificar señales tempranas de riesgo académico, priorizando métricas clave como el rendimiento académico, la asistencia y la repetición de asignaturas, permitiendo intervenciones oportunas y específicas de mayor impacto en la reducción de la deserción.

Para mitigar la deserción por factores económicos, es esencial desarrollar un programa institucional que incluya la provisión de ayudas financieras, becas condicionadas al rendimiento académico, y planes de pago flexibles para estudiantes con dificultades económicas. De igual manera es pertinente la adopción de políticas de retención más adaptativas, como la reestructuración de la carga académica o la oferta de modalidades de estudio más flexibles para aquellos con responsabilidades familiares o laborales.

El desarrollo de un sistema integral de soporte académico, que incluya tutorías personalizadas, talleres sobre estrategias de aprendizaje, y programas de mentoría que promuevan el desarrollo de habilidades transversales es fundamental. Paralelamente, se debe implementar un sistema automatizado de monitoreo del rendimiento académico, apoyado en análisis predictivos, que permita la detección temprana de señales de riesgo en los estudiantes, facilitando intervenciones oportunas y dirigidas

Se recomienda explorar enfoques analíticos avanzados, como modelos híbridos que combinan técnicas de regresión con algoritmos de inteligencia artificial, incluyendo árboles de decisión, redes neuronales y métodos ensemble. Estas herramientas permiten identificar relaciones no lineales y patrones complejos en la deserción, mejorando la precisión predictiva

y la personalización de las estrategias de intervención mediante la segmentación efectiva de grupos de riesgo.

Referencias Bibliográficas

- Azoumana, K. (2013). Análisis de la deserción estudiantil en la universidad simón bolívar, facultad ingeniería de sistemas, con técnicas de minería de datos. *Pensamiento Americano*, 41-51.
- Bernal, C. (2010). *Metodología de la investigación*. Bogotá: Pearson. Obtenido de <https://anyflip.com/vede/ohla/basic>
- Blomhøj, M. (2008). Modelización matemática: Una teoría para la práctica. *Revista de Educación Matemática*, 21.
- Chitarroni, H. (2002). *La regresión logística*. Buenos Aires: IDICSO.
- Consejo de Educación Superior (CES). (2013). *Resolución No. 048/2013*. Resolución oficial de organismo regulador.
- Espejo U, R. (2012). Factores sociales asociados al logro y abandono escolar en contextos rurales del departamento de La Paz. *Temas Sociales*, 111-138.
- Gaviria Hincapie, J. (2024). *Reconocimiento de patrones de deserción estudiantil mediante técnicas de analítica de datos, en el contexto de la educación por ciclos propedéuticos*. Medellín: Universidad de Medellín.
- Kelleher, J., Namee, B. M., & D'Arcy, A. (2015). *Fundamentals of Machine Learning for Predictive Data Analytics, Algorithms, Worked Examples, and Case Studies*. England : The MIT Press.

- Khalilian, M., Mustapha, N., Sulaiman, M., & Mamat, A. (2011). Intrusion detection system with data mining approach: A review. *Global Journal of Computer Science and Technology*. Obtenido de <https://api.semanticscholar.org/CorpusID:67080525>
- Lázaro Alvarez, N., Callejas Carrión, Z., & Griol Barres, D. (2020). Factores que inciden en la deserción estudiantil en carreras de perfil Ingeniería Informática. *Revista Fuentes*, 22(1), 105-126. doi:10.12795/revistafuentes.2020.v22.i1.09
- Lázaro Alvarez, N., Callejas Carrión, z., & Griol Barres, D. (2021). Utilización del software SPSS para identificar factores predictivos de deserción estudiantil. *LUZ*, 21(1), 38-50. Obtenido de <https://luz.uho.edu.cu/index.php/luz/article/view/1155>
- Ministerio de Educación del Ecuador. (2019). *Acuerdo Ministerial No. 140-2019*.
- Niño Rojas, V. (2011). *Metodología de la Investigación: Diseño y ejecución*. Bogotá: Ediciones de la U.
- Organización de las Naciones Unidas . (2015). *Agenda 2030 para el Desarrollo Sostenible*.
- Organización de las naciones unidas. (1948). *Declaración Universal de Derechos Humanos*. Asamblea General de las Naciones Unidas.
- Perez Lopez , C., & Santin Gonzalez, D. (2007). *Minería de datos. Técnicas y herramientas*. Madrid: Ediciones Paraninfo, S.A.
- Picardo Joao, O., Escobar, J., & Balmore Pacheco , R. (2004). *Diccionario enciclopédico de ciencias de la educación*. San Salvador: Centro de Investigación Educativa ; Colegio García Flamenco.

- Quintero Tangarife, Y. A. (2002). *Diseño de un modelo predictivo para generar alertas tempranas de deserción universitaria en los programas de pregrado presenciales de la Facultad de Ingeniería de la Universidad de Antioquia*. Medellín: Universidad de Antioquia.
- República del Ecuador. (2008). *Constitución de la República del Ecuador*. Montecristi: Registro Oficial.
- República del Ecuador. (2010 (actualizada 2018)). *Ley Orgánica de Educación Superior (LOES)*. Registro Oficial.
- República del Ecuador. (2018). *Decreto Ejecutivo No. 357*. Quito: Registro Oficial.
- Rivera Vergaray, K. (2021). Predictive model for the early detection of students with high risk of academic dropout. *Innovation and Software.*, 2, 6-13. doi:<https://doi.org/10.48168/innosoft.s6.a40>
- Teleamazonas. (9 de febrero de 2023). Obtenido de <https://www.teleamazonas.com/estudio-senescyt-desercion-universitaria-ecuador/>
- Valderrama Mendoza , S., & Jaimes Velasquez, C. (2019). *El desarrollo de la tesis. Descriptiva - comparativa, correlacional y cuasiexperimental*. Lima: San Marcos.
- Vásquez, J. (2016.). *Modelo predictivo para estimar la deserción de estudiantes en una institución de educación superior*. Santiago. Obtenido de <https://repositorio.uchile.cl/handle/2250/144169>

Vinueza López, C. N. (2021). *Diseño de un modelo matemático para estimar la deserción estudiantil mediante técnicas de análisis multivariado en una institución de educación superior tecnológica*. Ambato: Universidad Técnica de Ambato.

Young Chung, J., & Lee, S. (2019). Dropout early warning systems for high school students using machine learning. *Children and Youth Services Review*, 96, 346-353. Obtenido de www.sciencedirect.com/science/article/pii/S0190740918309721

Anexos

Anexo A. Base de datos



INSTITUTO SUPERIOR TECNOLÓGICO COTACACHI

ESTUDIANTES MATRICULADOS 2022_1

1. Tipo de Documento de Identificación	2. Número de Documento de Identificación	3. Primer apellido del estudiante	4. Segundo apellido del estudiante	5. Primer nombre del estudiante	6. Segundo nombre del estudiante	7. Sexo del estudiante	8. Género del estudiante	9. Estado civil	10. Etnia	11. Pueblo	12. Tipo de sangre del estudiante	13. Tiene discapacidad	14. Porcentaje de discapacidad
CÉDULA	1005406317	CADENA	CORAL	NAYDELIN	ANAHI	MUJER	FEMENINO	SOLTERO	MESTIZO	NO APLICA	ORH(+)	NO	NO APLICA
CÉDULA	1004169148	MORAN	QUINCHIGUANGO	ROSA	BELEN	MUJER	FEMENINO	CASADO	INDIGENA	OTAVALO	ORH(+)	NO	NO APLICA
CÉDULA	1003888539	CAMPUES	ANPRANGO	ANDREA	MARCELA	MUJER	FEMENINO	SOLTERO	MESTIZO	NO APLICA	ORH(+)	NO	NO APLICA
CÉDULA	1004034813	CARRERA	VILLEGAS	MARIA	GABRIELA	MUJER	FEMENINO	SOLTERO	MESTIZO	NO APLICA	ORH(+)	SI	98.00
CÉDULA	1003700877	RAMOS	MAIGUA	SONIA	MARIBEL	MUJER	FEMENINO	DIVORCIADO	INDIGENA	OTAVALO	ORH(+)	NO	NO APLICA
CÉDULA	1002530663	CAHUASQUI	CASTRO	MARTHA	FABIOLA	MUJER	FEMENINO	CASADO	MESTIZO	NO APLICA	ORH(+)	NO	NO APLICA
CÉDULA	1004024905	YASELGA	FLORES	SONIA	ELIZABETH	MUJER	FEMENINO	SOLTERO	MESTIZO	NO APLICA	ORH(+)	NO	NO APLICA
CÉDULA	1003874714	ARTES	MORALES	KORY	KORAIMA	MUJER	FEMENINO	SOLTERO	MESTIZO	NO APLICA	ARH(+)	NO	NO APLICA
CÉDULA	1005078140	ECHAVEERRIA	CIFUENTES	DAYANA	MICHELE	MUJER	FEMENINO	SOLTERO	MESTIZO	NO APLICA	ORH(+)	NO	NO APLICA
CÉDULA	1002381018	MARVAEZ	CAZAR	SANDRA	ELIZABETH	MUJER	FEMENINO	CASADO	MESTIZO	NO APLICA	ORH(+)	NO	NO APLICA
CÉDULA	1725425290	FLORES	MALES	VIVIANA	MARISOL	MUJER	FEMENINO	SOLTERO	MESTIZO	NO APLICA	ORH(+)	NO	NO APLICA
CÉDULA	1004089726	LIBIDIA	NAVARRO	STEEVEN	ISRAEL	HOMBRE	MASCULINO	SOLTERO	MESTIZO	NO APLICA	ORH(+)	NO	NO APLICA



INSTITUTO SUPERIOR TECNOLÓGICO COTACACHI

ESTUDIANTES MATRICULADOS 2024_1

49. Sexo razón de beca	50. Valor del monto de la beca	51. Porcentaje de la beca que cubre el valor del arancel	52. Porcentaje de la beca que cubre la manutención	53. Tipo de financiamiento de la beca	54. Valor del monto de la ayuda económica	55. Valor del monto de crédito educativo	56. Ha participado durante el periodo de su	57. Alcance del proyecto de vinculación con la sociedad	58. Correo electrónico del estudiante
NO APLICA	NO APLICA	NO APLICA	NO APLICA	FONDOS PROPIOS	NO APLICA	NO APLICA	NO	NO APLICA	gladysmaria1839@gmail.com
NO APLICA	NO APLICA	NO APLICA	NO APLICA	FONDOS PROPIOS	NO APLICA	NO APLICA	SI	PARROQUIAL	abigailquinchano2000@gmail.com
NO APLICA	NO APLICA	NO APLICA	NO APLICA	FONDOS PROPIOS	NO APLICA	NO APLICA	NO	NO APLICA	liss.abril1234@gmail.com
NO APLICA	NO APLICA	NO APLICA	NO APLICA	FONDOS PROPIOS	NO APLICA	NO APLICA	NO	NO APLICA	achinasandy85@gmail.com
NO APLICA	NO APLICA	NO APLICA	NO APLICA	FONDOS PROPIOS	NO APLICA	NO APLICA	NO	NO APLICA	valentinamorillo17@gmail.com
NO APLICA	NO APLICA	NO APLICA	NO APLICA	FONDOS PROPIOS	NO APLICA	NO APLICA	SI	PARROQUIAL	carolinatabango823@gmail.com
NO APLICA	NO APLICA	NO APLICA	NO APLICA	FONDOS PROPIOS	NO APLICA	NO APLICA	SI	CANTONAL	estefy.beja22@gmail.com
NO APLICA	NO APLICA	NO APLICA	NO APLICA	FONDOS PROPIOS	NO APLICA	NO APLICA	NO		aguirreisabel945@gmail.com
NO APLICA	NO APLICA	NO APLICA	NO APLICA	FONDOS PROPIOS	NO APLICA	NO APLICA	SI	CANTONAL	miryamlanchimba123@gmail.com
NO APLICA	NO APLICA	NO APLICA	NO APLICA	FONDOS PROPIOS	NO APLICA	NO APLICA	SI	CANTONAL	carmitapatty2001@gmail.com
NO APLICA	NO APLICA	NO APLICA	NO APLICA	FONDOS PROPIOS	NO APLICA	NO APLICA	SI	PARROQUIAL	katherinviema2001@gmail.com
NO APLICA	NO APLICA	NO APLICA	NO APLICA	FONDOS PROPIOS	NO APLICA	NO APLICA	NO	NO APLICA	daya2003flores25@gmail.com
NO APLICA	NO APLICA	NO APLICA	NO APLICA	FONDOS PROPIOS	NO APLICA	NO APLICA	NO		anfuerez14@gmail.com
NO APLICA	NO APLICA	NO APLICA	NO APLICA	FONDOS PROPIOS	NO APLICA	NO APLICA	NO	NO APLICA	fernandacastrourbano@gmail.com

Anexo B. Depuración y limpieza

Cedula	Sexo	Genero	Estciv	Etnia	Pueblo	TipSangre	Discapacidad	PorDiscap	TipDisc	Fascimicato	Edad	Nacionalidad	ProvNaci
0202064242	MUJER	FEMENINO	CASADO	MESTIZO	NO APLICA	ARH(+)	NO	0	NO APLICA	01/07/1983	41.00	ECUADOR	BOLIVAR
0401194428	MUJER	FEMENINO	DIVORCIADO	MESTIZO	NO APLICA	ORH(+)	NO	0	NO APLICA	31/07/1975	49.00	ECUADOR	CARCHI
0401194428	MUJER	FEMENINO	DIVORCIADO	MESTIZO	NO APLICA	ORH(+)	NO	0	NO APLICA	31/07/1975	49.00	ECUADOR	CARCHI
0401387469	MUJER	FEMENINO	SOLTERO	MESTIZO	NO APLICA	ORH(+)	NO	0	NO APLICA	09/12/1981	42.00	ECUADOR	CARCHI
0401422369	MUJER	FEMENINO	CASADO	MESTIZO	NO APLICA	ORH(+)	NO	0	NO APLICA	30/06/1982	42.00	ECUADOR	CARCHI
0401422369	MUJER	FEMENINO	CASADO	MESTIZO	NO APLICA	ORH(+)	NO	0	NO APLICA	30/06/1982	42.00	ECUADOR	CARCHI
0401526769	MUJER	FEMENINO	SOLTERO	INDIGENA	OTAVALO	ORH(+)	SI	38.00	DISCAPACIDAD D.AJUDITIVA	25/04/1994	30.00	ECUADOR	CARCHI
0401584966	MUJER	FEMENINO	SOLTERO	MESTIZO	NO APLICA	ORH(+)	NO	0	NO APLICA	08/07/1990	34.00	ECUADOR	CARCHI
0401589734	MUJER	FEMENINO	SOLTERO	MESTIZO	NO APLICA	ORH(+)	NO	0	NO APLICA	30/09/1990	34.00	ECUADOR	IMBABURA
0401625413	MUJER	FEMENINO	SOLTERO	MESTIZO	NO APLICA	ORH(+)	NO	0	NO APLICA	28/12/2002	21.00	ECUADOR	CARCHI
0401625413	MUJER	FEMENINO	SOLTERO	MESTIZO	NO APLICA	ORH(+)	NO	0	NO APLICA	28/12/2002	21.00	ECUADOR	CARCHI
0401683099	MUJER	FEMENINO	CASADO	MESTIZO	NO APLICA	ORH(+)	NO	0	NO APLICA	01/03/1991	33.00	ECUADOR	IMBABURA
0401728993	MUJER	FEMENINO	CASADO	MESTIZO	NO APLICA	ORH(+)	NO	0	NO APLICA	22/11/1993	31.00	ECUADOR	CARCHI
0401728993	MUJER	FEMENINO	CASADO	MESTIZO	NO APLICA	ORH(+)	NO	0	NO APLICA	22/11/1993	31.00	ECUADOR	CARCHI
0401763248	MUJER	FEMENINO	SOLTERO	MESTIZO	NO APLICA	ARH(+)	NO	0	NO APLICA	08/11/2003	21.00	ECUADOR	CARCHI
0401763248	MUJER	FEMENINO	SOLTERO	MESTIZO	NO APLICA	ARH(+)	NO	0	NO APLICA	08/11/2003	21.00	ECUADOR	CARCHI

A	B
1	
2	
3	Etiquetas de fila Cuenta de ProvNaci
4	26
5	AZUAY 2
6	BOLIVAR 4
7	CARCHI 78
8	CHIMBORAZO 1
9	COTOPAXI 11
10	EL ORO 12
11	ESMERALDAS 19
12	GUAYAS 10
13	IMBABURA 2232
14	LOJA 3
15	LOS RIOS 2
16	MANABI 15
17	PASTAZA 1
18	PICHINCHA 442
19	SANTA ELENA 1
20	SANTO DOMINGO DE LOS TSACHILAS 12
21	SUCUMBIOS 10
22	Total general 2881
23	
24	
25	
26	
27	
28	
29	
30	

Campos de tabla dinámica

Seleccionar campos para agregar al informe:

Buscar

- Cedula
- Sexo
- Genero
- Estciv
- Etnia
- Pueblo
- TipSangre
- Discapacidad
- PorDiscap

Arrastrar campos entre las áreas siguientes:

Filtros	Columnas
<ul style="list-style-type: none"> Provincia 	<ul style="list-style-type: none"> Cuenta de ProvNaci

Anexo C. Normalización

Dedicacion	DestIngreE	DestIngreEst_cod	BonoFamili	BonoFamili	NivelPadre	NivelPadre	NivelMadre	NivelMadre	IngHog	IngHog_no	NumIntFam	Desierta	
1	PARA MANTE	1	NO		0 EDUCACIÓN		3 EDUCACIÓN		3	386	0.06	4	1
1	PARA MANTE	1	NO		0 PRIMARIA		4 PRIMARIA		4	500	0.08	4	1
1	PARA MANTE	1	NO		0 PRIMARIA		4 PRIMARIA		4	200	0.02	3	1
1	PARA MANTE	1	NO		0 SUPERIOR U		8 SECUNDARI		5	1000	0.18	4	1
0	FINANCIAR S	2	NO		0 PRIMARIA		4 EDUCACIÓN		3	460	0.07	3	1
1	PARA MANTE	1	NO		0 SECUNDARI		5 SECUNDARI		5	450	0.07	4	1
0	NO APLICA	0	NO		0 SECUNDARI		5 SECUNDARI		5	2000	0.39	6	1
0	NO APLICA	0	NO		0 EDUCACIÓN		3 PRIMARIA		4	350	0.05	3	1
0	FINANCIAR S	2	NO		0 EDUCACIÓN		3 EDUCACIÓN		3	450	0.07	3	1
0	FINANCIAR S	2	NO		0 PRIMARIA		4 SECUNDARI		5	500	0.08	4	1
0	NO APLICA	0	NO		0 PRIMARIA		4 PRIMARIA		4	450	0.07	6	1
1	FINANCIAR S	2	NO		0 PRIMARIA		4 SECUNDARI		5	140	0.01	5	1
0	FINANCIAR S	2	NO		0 EDUCACIÓN		6 SECUNDARI		5	1200	0.22	4	1
0	NO APLICA	0	NO		0 SECUNDARI		5 SUPERIOR U		8	500	0.08	2	1
1	FINANCIAR S	2	NO		0 SECUNDARI		5 SECUNDARI		5	500	0.08	3	1
0	NO APLICA	0	NO		0 SECUNDARI		5 EDUCACIÓN		6	600	0.10	9	1
0	FINANCIAR S	2	NO		0 PRIMARIA		4 SECUNDARI		5	500	0.08	5	1
0	FINANCIAR S	2	NO		0 EDUCACIÓN		3 EDUCACIÓN		3	460	0.07	4	1
0	NO APLICA	0	NO		0 PRIMARIA		4 PRIMARIA		4	250	0.03	3	1
0	NO APLICA	0	NO		0 SUPERIOR N		7 SECUNDARI		5	300	0.04	3	1
0	GASTOS PEF	3	NO		0 SUPERIOR N		7 SUPERIOR N		7	400	0.06	7	1
0	FINANCIAR S	2	NO		0 SUPERIOR N		7 EDUCACIÓN		3	1100	0.20	5	1
0	NO APLICA	0	NO		0 SUPERIOR N		7 PRIMARIA		4	900	0.16	6	1
0	NO APLICA	0	NO		0 EDUCACIÓN		3 SECUNDARI		5	300	0.04	4	1
1	GASTOS PEF	3	NO		0 PRIMARIA		4 PRIMARIA		4	200	0.02	4	1
1	FINANCIAR S	2	NO		0 SECUNDARI		5 EDUCACIÓN		6	750	0.13	7	1
1	PARA MANTE	1	NO		0 SECUNDARI		5 SECUNDARI		5	340	0.05	3	1
1	FINANCIAR S	2	NO		0 SECUNDARI		5 EDUCACIÓN		3	500	0.08	6	1
1	PARA MANTE	1	NO		0 SECUNDARI		5 CENTRO DE		1	380	0.06	3	1
1	FINANCIAR S	2	SI		1 EDUCACIÓN		3 EDUCACIÓN		3	1000	0.18	5	1

A	B	C	D	E	F	G	H	I	J	K	L	M	N
Se	Estciv_c	Etnia_c	Pueblo_c	TipSangre_c	Discap_c	PorDiscap_t	TipDisc_c	Edad_nor	Nacionalidad_c	ProvNaci_c	CanNaci_c	Resideni	ProvRes_co
0	0	0	0	0	0	0.00	0	0.56	0	0	0	0	1
0	1	0	0	1	0	0.00	0	0.76	0	3	0	0	1
0	2	0	0	1	0	0.00	0	0.59	0	3	0	0	1
0	0	0	0	1	0	0.00	0	0.59	0	3	0	0	1
0	2	1	1	1	1	0.45	1	0.29	0	3	0	0	1
0	2	0	0	1	0	0.00	0	0.39	0	3	0	0	1
0	2	0	0	1	0	0.00	0	0.37	0	1	0	0	1
0	2	0	0	1	0	0.00	0	0.07	0	3	0	0	1
0	0	0	0	1	0	0.00	0	0.37	0	1	4	0	1
0	0	0	0	1	0	0.00	0	0.29	0	3	0	0	1
0	2	0	0	0	0	0.00	0	0.05	0	3	0	0	1
1	2	0	0	2	0	0.00	0	0.12	0	0	0	0	1
0	2	0	0	3	0	0.00	0	0.07	0	3	0	0	3
0	2	0	0	0	0	0.00	0	0.07	0	3	0	0	3
0	2	0	0	1	0	0.00	0	0.34	0	1	2	0	1
0	2	0	0	0	0	0.00	0	0.07	0	3	0	0	3
1	2	0	0	0	0	0.00	0	0.10	0	3	0	0	1
0	2	0	0	1	0	0.00	0	0.24	0	3	0	0	3
0	2	0	0	1	0	0.00	0	0.07	0	3	0	0	3
0	2	0	0	1	0	0.00	0	0.17	0	3	0	0	3
0	2	0	0	1	0	0.00	0	0.10	0	3	0	0	3
0	2	0	0	2	0	0.00	0	0.10	0	3	0	0	1
0	2	0	0	1	0	0.00	0	0.07	0	3	0	0	3
1	2	0	0	1	0	0.00	0	0.12	0	3	0	0	3
0	2	0	0	1	0	0.00	0	0.20	0	3	0	0	1
0	2	0	0	3	0	0.00	0	0.02	0	3	0	0	1
1	2	0	0	1	0	0.00	0	0.12	0	3	0	0	3
0	2	0	0	1	0	0.00	0	0.10	0	1	2	0	1
0	2	0	0	1	0	0.00	0	0.10	0	3	0	0	3
0	2	0	0	3	0	0.00	0	0.05	0	3	0	0	3

Anexo D. Programación

```
[ ] !pip install scikit-learn
```

Requirement already satisfied: scikit-learn in /usr/local/lib/python3.10/dist-packages (1.5.2)
Requirement already satisfied: numpy>=1.19.5 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.26.4)
Requirement already satisfied: scipy>=1.6.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.13.1)
Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.4.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (3.5.0)

Regresión Logística

```
[ ] import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
```

```
[ ] #Carga de datos
df = pd.read_csv('/content/DatosDepuradosFinal_1.csv')
df.head()

#df = pd.DataFrame(data, columns=columns)
print(df)
```

```
df.head()
```

	Sexo	Estciv_cod	Etnia_cod	Pueblo_cod	TipSangre_cod	Discap_cod	PorDiscap_nor	TipDisc_cod	Edad_nor
0	0	0	0	0	0	0	0.00	0	0.56
1	0	1	0	0	1	0	0.00	0	0.76
2	0	2	0	0	1	0	0.00	0	0.59
3	0	0	0	0	1	0	0.00	0	0.59
4	0	2	1	1	1	1	0.45	1	0.29

5 rows × 10 columns

```
corr=df.corr()
print(corr)
```

BonoFamilia_cod	-0.033972	-0.011517	0.083920	0.070687	-0.028562
NivelPadre_cod	0.058884	0.046925	-0.156121	-0.121011	-0.023488
NivelMadre_cod	0.015104	0.069757	-0.206470	-0.215633	0.025389
IngHog_nor	0.094211	-0.016986	-0.081504	-0.063730	-0.037953
NumIntFam	-0.004707	0.113147	0.110650	0.158595	-0.039572
Desierta	-0.018284	-0.043490	-0.036036	-0.029989	-0.011245

Anexo E. Correlación

```
X = df.drop('Desierta', axis=1)
y = df['Desierta']

[ ] # prompt: correlation of df

# Calcular la correlación de cada variable independiente con la variable dependiente
correlation_with_y = X.corrwith(y)

# Convertir el resultado a DataFrame para una visualización más clara
correlation_with_y_df = pd.DataFrame(correlation_with_y, columns=['Desierta'])

# Asegurarse de que 'Desierta' esté en la primera fila del DataFrame
#correlation_with_y_df = correlation_with_y_df.append(pd.DataFrame({'Desierta': [y.corr(y)], index=['Desierta']}))

# Mostrar el resultado
print(correlation_with_y_df)
```

Anexo F. Diagrama de calor

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

# Datos de las correlaciones obtenidas
data = {
    'Variable': ['Sexo', 'Estado civil', 'Etnia', 'Pueblo', 'Tipo de sangre', 'Discapacidad', '% de Disc.',
                'Tipo de Disc', 'Edad', 'Nacionalidad', 'Prov Naci', 'Cantón Naci', 'Provincia Res',
                'Cantón Res', 'Tipo de Co', 'Modalidad', 'Jornada', 'Tipo Mat', 'Nivel',
                'Repite', 'Perdida Grat', 'Pesión Dif', 'Dedicación', 'Destino Ingre Est', 'Bono Familia',
                'Nivel Padre', 'Nivel Madre', 'Ingreso Hog', 'Nº Int Flia'],
    'Desierta': [-0.018284, -0.043490, -0.036036, -0.029989, -0.011245, -0.031043, -0.025617, -0.027750,
                0.201170, 0.001386, 0.001790, -0.002204, 0.007070, -0.017321, 0.004392, 0.012062, 0.075480,
                0.044335, 0.240644, 0.303248, 0.128675, -0.000353, 0.078729, 0.186072, -0.081740, 0.060425,
                0.061388, -0.019523, -0.023908]
}

# Crear un DataFrame
df_corr = pd.DataFrame(data)

# Configuración para el heatmap
plt.figure(figsize=(12, 10))
heatmap_data = df_corr.set_index('Variable') # Usar las variables como índice
sns.heatmap(heatmap_data, annot=True, cmap="coolwarm", fmt=".2f", cbar_kws={'label': 'Correlación'})

# Personalización del gráfico
plt.title("Diagrama de calor de correlaciones con la variable 'Desierta'", fontsize=16)
plt.xlabel("Variable Dependiente: Desierta")
plt.ylabel("Variables Independientes")
plt.tight_layout()

# Mostrar el gráfico
plt.show()
```

Anexo G. Matriz de confusión

```
[ ] from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.pipeline import make_pipeline
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report

# Crear un pipeline que incluya el escalado de datos y el modelo de regresión logística
pipeline = make_pipeline(
    StandardScaler(), # Escalar las características
    LogisticRegression(solver='liblinear', max_iter=500) # Modelo de regresión logística
)

# Ajustar el modelo usando el pipeline
pipeline.fit(X_train, y_train)

# Realizar predicciones en el conjunto de prueba
predicciones = pipeline.predict(X_test)

# Calcular la precisión del modelo
precision = accuracy_score(y_test, predicciones)
print(f'Precisión del modelo: {precision}')

# Mostrar la matriz de confusión
matriz_confusion = confusion_matrix(y_test, predicciones)
print('Matriz de Confusión:')
print(matriz_confusion)

# Mostrar el informe de clasificación
report = classification_report(y_test, predicciones)
print('Informe de clasificación:')
print(report)

# Configurar la visualización
plt.figure(figsize=(8, 6))
sns.heatmap(matriz_confusion, annot=True, fmt='d', cmap='Blues',
            xticklabels=['No deserta', 'Deserta'], yticklabels=['No deserta', 'Deserta'])
plt.xlabel('Predicción')
plt.ylabel('Actual')
plt.title('Matriz de Confusión')
plt.show()
```

Anexo H. Modelo de regresión

```
[ ] # Obtener el modelo de regresión logística del pipeline
    modelo_logistico = pipeline.named_steps['logisticregression']

    # Extraer coeficientes e intercepto
    coeficientes = modelo_logistico.coef_
    intercepto = modelo_logistico.intercept_

    # Imprimir los coeficientes y el intercepto
    print("Coeficientes:", coeficientes)
    print("Intercepto:", intercepto)

    # Importar statsmodels y train_test_split una sola vez
    import statsmodels.api as sm
    from sklearn.model_selection import train_test_split
```

Coeficientes: [[-0.11858929 0.10932288 0.01278788 0.02255362 -0.0855227 -0.14659836
 0.03738101 0.00427404 0.54267487 -0.00420109 -0.13583942 0.11852649
 0.02342068 -0.11550706 0.05559776 0.20304835 0.16301707 0.04513862
 0.67701103 0.82099455 0.32562297 -0.00189349 -0.24875521 0.44862818
 -0.1287842 0.04354413 0.17541695 -0.05293619 0.02268395]]
 Intercepto: [0.13971552]

Anexo I. Ajuste del modelo

```
[ ] import statsmodels.api as sm
    from sklearn.model_selection import train_test_split

    X_train_const = sm.add_constant(X_train)

    # Ajustamos el modelo de regresión logística con los datos de entrenamiento
    modelo_glm = sm.Logit(y_train, X_train_const).fit()

    # Imprimimos el resumen del modelo para obtener coeficientes, significancia, y otros detalles
    print(modelo_glm.summary())

    # Para acceder a los coeficientes y sus valores p
    coeficientes = modelo_glm.params
    valores_p = modelo_glm.pvalues
    print("Coeficientes:", coeficientes)
    print("Valores p:", valores_p)

    # Opcionalmente, puedes usar el modelo ajustado para hacer predicciones en X_test
    X_test_const = sm.add_constant(X_test)
    predicciones = modelo_glm.predict(X_test_const)

    # Si necesitas convertir predicciones a 0 o 1 (clasificación binaria)
    predicciones_binarias = [1 if x > 0.5 else 0 for x in predicciones]
```