

UNEMI

UNIVERSIDAD ESTATAL DE MILAGRO

REPÚBLICA DEL ECUADOR

UNIVERSIDAD ESTATAL DE MILAGRO

VICERRECTORADO DE INVESTIGACIÓN Y POSGRADO

PROYECTO DE INVESTIGACIÓN APLICADA Y/O DE DESARROLLO
PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

MAGÍSTER EN INTELIGENCIA ARTIFICIAL EN EDUCACIÓN

TEMA:

Modelo Predictivo de IA Enfocado en la Detección Temprana del Abandono Escolar,
Vinculando la Variable Socioemocional en estudiantes de Bachillerato en la
Educación Ecuatoriana.

Autor:

Vanesa Susana Sagñay Cholca
Jimmy Javier Montaguano Jimenez

Tutor: Magister Beatriz Annabell Loor Avila

Milagro, 2026

Derechos de autor

Sr. Dr.

Fabricio Guevara Viejó

Rector de la Universidad Estatal de Milagro

Presente.

Yo, **Vanesa Susana Sagñay Cholca** en calidad de autor y titular de los derechos morales y patrimoniales de este informe de investigación, mediante el presente documento, libre y voluntariamente cedo los derechos de Autor de este proyecto de desarrollo, que fue realizada como requisito previo para la obtención de mi Grado, de **Magíster en Inteligencia Artificial en Educación**, como aporte a la Línea de Proyecto de Investigación de conformidad con el Art. 114 del Código Orgánico de la Economía Social de los Conocimientos, Creatividad e Innovación, concedo a favor de la Universidad Estatal de Milagro una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos. Conservo a mi favor todos los derechos de autor sobre la obra, establecidos en la normativa citada.

Así mismo, autorizo a la Universidad Estatal de Milagro para que realice la digitalización y publicación de este Proyecto de Investigación en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

El autor declara que la obra objeto de la presente autorización es original en su forma de expresión y no infringe el derecho de autor de terceros, asumiendo la responsabilidad por cualquier reclamación que pudiera presentarse por esta causa y liberando a la Universidad de toda responsabilidad.

Milagro, **3 de Marzo del 2026**



Vanesa Susana Sagñay Cholca

1752702678

Derechos de autor

Sr. Dr.

Fabricio Guevara Viejó

Rector de la Universidad Estatal de Milagro

Presente.

Yo, **Jimmy Javier Montaguano** en calidad de autor y titular de los derechos morales y patrimoniales de este informe de investigación, mediante el presente documento, libre y voluntariamente cedo los derechos de Autor de este proyecto de desarrollo, que fue realizada como requisito previo para la obtención de mi Grado, de **Magíster en Inteligencia Artificial en Educación**, como aporte a la Línea de Proyecto de Investigación de conformidad con el Art. 114 del Código Orgánico de la Economía Social de los Conocimientos, Creatividad e Innovación, concedo a favor de la Universidad Estatal de Milagro una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos. Conservo a mi favor todos los derechos de autor sobre la obra, establecidos en la normativa citada.

Así mismo, autorizo a la Universidad Estatal de Milagro para que realice la digitalización y publicación de este Proyecto de Investigación en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

El autor declara que la obra objeto de la presente autorización es original en su forma de expresión y no infringe el derecho de autor de terceros, asumiendo la responsabilidad por cualquier reclamación que pudiera presentarse por esta causa y liberando a la Universidad de toda responsabilidad.

Milagro, **3 de Marzo del 2026**



Validar únicamente en FirmaEC.
Firmado electrónicamente por:
**JIMMY JAVIER
MONTAGUANO JIMENEZ**

Jimmy Javier Montaguano

CI

Aprobación del tutor del Trabajo de Titulación

Yo, **Beatriz Loor** en mi calidad de director del trabajo de titulación, elaborado por **Vanesa Susana Sagñay Cholca y Jimmy Javier Montaguano**, cuyo tema es **Modelo Predictivo de IA Enfocado en la Detección Temprana del Abandono Escolar, Vinculando la Variable Socioemocional en estudiantes de Bachillerato en la Educación Ecuatoriana**, que aporta a la Línea de Investigación **Proyecto de Investigación**, previo a la obtención del Grado Maestría en Inteligencia Artificial en educación. Trabajo de titulación que consiste en una propuesta innovadora que contiene, como mínimo, una investigación exploratoria y diagnóstica, base conceptual, conclusiones y fuentes de consulta, considero que el mismo reúne los requisitos y méritos necesarios para ser sometido a la evaluación por parte del tribunal calificador que se designe, por lo que lo APRUEBO, a fin de que el trabajo sea habilitado para continuar con el proceso de titulación de la alternativa de Informe de Investigación de la Universidad Estatal de Milagro.

Milagro, **22 de Marzo del 2026**



Beatriz Annabell Loor Avila

Aprobación del tribunal calificador



FACULTAD DE POSGRADO ACTA DE SUSTENTACIÓN MAESTRÍA EN INTELIGENCIA ARTIFICIAL PARA LA EDUCACIÓN

En la Facultad de Posgrado de la Universidad Estatal de Milagro, a los veinticuatro días del mes de abril del dos mil veintiseis, siendo las 11:00 horas, de forma VIRTUAL comparece el/la maestrante, SOC. MONTAGUANO JIMÉNEZ JIMMY JAVIER, a defender el Trabajo de Titulación denominado " **MODELO PREDICTIVO DE IA ENFOCADO EN LA DETECCIÓN TEMPRANA DEL ABANDONO ESCOLAR, VINCULANDO VARIABLE SOCIOEMOCIONAL EN EDUCACIÓN GENERAL BÁSICA SUPERIOR** ", ante el Tribunal de Calificación integrado por: Msc MONTERO REYES YILENA, Presidente(a), Mgtr., ZALDUMBIDE LOPEZ VERONICA JANETH en calidad de Vocal; y, FRIAS SERRANO LUIS FELIPE que actúa como Secretario/a.

Una vez defendido el trabajo de titulación; examinado por los integrantes del Tribunal de Calificación, escuchada la defensa y las preguntas formuladas sobre el contenido del mismo al maestrante compareciente, durante el tiempo reglamentario, obtuvo las siguientes calificaciones:

TRABAJO DE TITULACIÓN	55.00
DEFENSA ORAL	36.67
PROMEDIO	91.67
EQUIVALENTE	MUY BUENO

Para constancia de lo actuado firman en unidad de acto el Tribunal de Calificación, siendo las 12:00 horas.



Msc MONTERO REYES YILENA
PRESIDENTE/A DEL TRIBUNAL



Mgtr., ZALDUMBIDE LOPEZ VERONICA JANETH
VOCAL



FRIAS SERRANO LUIS FELIPE
SECRETARIO/A DEL TRIBUNAL



SOC. MONTAGUANO JIMÉNEZ JIMMY JAVIER
MAGISTER

FACULTAD DE POSGRADO
ACTA DE SUSTENTACIÓN
MAESTRÍA EN INTELIGENCIA ARTIFICIAL PARA LA EDUCACIÓN

En la Facultad de Posgrado de la Universidad Estatal de Milagro, a los veinticuatro días del mes de abril del dos mil veintiseis, siendo las 11:00 horas, de forma VIRTUAL comparece el/la maestrante, LCDA. SAGRAY CHOLCA VANESA SUSANA, a defender Trabajo de Titulación denominado " **MODELO PREDICTIVO DE IA ENFOCADO EN LA DETECCIÓN TEMPRANA DE ABANDONO ESCOLAR, VINCULANDO VARIABLE SOCIOEMOCIONAL EN EDUCACIÓN GENERAL BÁSICA SUPERIOR** ", ante el Tribunal de Calificación integrado por: Msc MONTERO REYES YILENA, Presidente(a), Mgtr., ZALDUMBIDE LOPEZ VERONICA JANETH en calidad de Vocal; y, FRIAS SERRANO LUIS FELIPE que actúa como Secretario/a.

Una vez defendido el trabajo de titulación; examinado por los integrantes del Tribunal de Calificación, escuchada la defensa y las preguntas formuladas sobre el contenido del mismo al maestrante compareciente, durante el tiempo reglamentario, obtuvo las siguientes calificaciones:

TRABAJO DE TITULACIÓN	55.00
DEFENSA ORAL	38.67
PROMEDIO	93.67
EQUIVALENTE	MUY BUENO

Para constancia de lo actuado firman en unidad de acto el Tribunal de Calificación, siendo las 12:00 horas.



UNIVERSIDAD ESTATAL DE MILAGRO
FACULTAD DE POSGRADO
YILENA MONTERO REYES

Msc MONTERO REYES YILENA
PRESIDENTE/A DEL TRIBUNAL



UNIVERSIDAD ESTATAL DE MILAGRO
FACULTAD DE POSGRADO
VERONICA JANETH ZALDUMBIDE LOPEZ

Mgtr., ZALDUMBIDE LOPEZ VERONICA JANETH
VOCAL



UNIVERSIDAD ESTATAL DE MILAGRO
FACULTAD DE POSGRADO
LUIS FELIPE FRIAS SERRANO

FRIAS SERRANO LUIS FELIPE
SECRETARIO/A DEL TRIBUNAL



Vanesa Susana Sagray Cholca
Time Stamp
Integrity Data

LCDA. SAGRAY CHOLCA VANESA SUSANA
MAGISTER

DEDICATORIA

Este presente trabajo lo dedico al cielo, quienes desde allá han visto todos y cada uno de mis sueños logrados. Y agradecida por la bendición de tener a mi familia que me apoya en cada uno de mis momentos, y me muestra su amor incondicional.

Vanesa Sagñay

AGRADECIMIENTOS

Agradezco a Dios por su amor y las bendiciones que ha derramado sobre mí, porque sin él no viviría estos momentos. A mi familia por el amor y el apoyo en cada uno de los momentos de mi vida, porque sin su ayuda no estaría logrando mis objetivos.

Vanesa Sagñay

ÍNDICE DE CONTENIDOS

Resumen	xvii
Abstract	xviii
INTRODUCCIÓN	1
CAPÍTULO I	3
PROBLEMA DE INVESTIGACIÓN	3
1.1. Planteamiento del problema	5
1.2. Formulación del problema.	7
1.3. Preguntas de investigación.	8
1.3.1. Objetivo general	8
1.3.2. Objetivos específicos	8
1.4. Justificación.	8
1.5. Formulación de hipótesis y determinación de variables.	8
1.5.1. Hipótesis general	9
1.5.2. Hipótesis variables	9
1.5.3. Variable dependiente	9
1.5.4. Variables independientes	9
1.5.5. Variables moderadoras y de intervención	9
CAPÍTULO II	10
MARCO TEÓRICO REFERENCIAL	10
2.1. Antecedentes de la investigación	10
2.1.1. Antecedentes internacionales	12
2.1.2. Antecedentes nacionales	16
2.2. Estado del arte	18
2.2.1. Dimensiones socioeconómica y estructural	24
2.2.2. Dimensión académica e institucional	24
2.2.3. Dimensión socioemocional y salud mental	25
2.2.4. Sistemas de alerta temprana y abandono escolar	26
2.3. Bases teóricas conceptuales de la investigación	28
2.3.1. Teoría Ecológica y Multidimensionalidad del Riesgo	28
2.3.2. Constructos socioemocionales y resiliencia educativa	29
2.3.3. Inteligencia Artificial y Aprendizaje Basado en Datos (ABD)	29
2.3.4. Teorías del aprendizaje en la era digital	30
2.3.5. IA Explicable (XAI) y ética algorítmica	30
2.4. Bases legales que fundamentan el estudio	32
CAPÍTULO III	38
MARCO METODOLÓGICO	38
3.1. Paradigma, Enfoque, alcance, modalidad, tipo de estudio y diseño de investigación	38
3.1.1. Paradigma	40

3.1.2. Enfoque	41
3.1.3. Alcance	44
3.1.4. Modalidad	46
3.1.5. Tipo de estudio	46
3.1.6. Diseño de la investigación	48
3.2. Matriz de Operacionalización de variables	48
3.4. Técnicas e instrumentos de recolección de información.	51
3.4.1. Formato del instrumento de recolección de datos	52
3.5. Técnica de análisis de datos	53
CAPÍTULO IV	56
ANÁLISIS E INTERPRETACIÓN DE RESULTADOS	56
4.1. Análisis de resultados.	56
4.2. Discusión de los Resultados/Triangulación (por cada ítem)	79
CAPÍTULO V	88
5.1. Conclusiones	90
5.2. Recomendaciones	92
Referencias	94
Anexo	98
Anexo 1: Formato de Validación de Instrumentos de Recolección de Datos	

Índice de tablas

Tabla 1 Criterios a evaluar en el estudio	11
Tabla 2. Tasa de abandono nacional.	21
Tabla 3. Deserción escolar en América Latina.	23
Tabla 4. Dimensión de riesgo en la incidencia del abandono académico	26
Tabla 5. Niveles de alerta.	27
Tabla 6. Marco conceptual de IA y psicopedagogía educativa.	32
Tabla 7. Métricas de calidad del sistema	39
Tabla 8. Estrategia metodológica: enfoque mixto e IA predictiva	43
Tabla 9. Análisis del proyecto: IA para la prevención del abandono escolar	45
Tabla 10. Resumen técnico del diseño	47
Tabla 11. Operacionalización de las variables en el estudio	49
Tabla 12. Cuestionario de diagnóstico socioemocional	52
Tabla 13. Rango de Alfa de Cronbach	53
Tabla 14. Tipo de variable y función del modelo	53
Tabla 15. Distribución de estudiantes según el curso	58
Tabla 16. Diversidad de género de los participantes	58
Tabla 17. Estadística descriptiva de comprensión de contenidos	59
Tabla 18. Análisis de la variable cumplimiento de tareas	60
Tabla 19. Análisis de la variable rendimiento y esfuerzo	61
Tabla 20. Distribución de la carga académica	62
Tabla 21. Análisis de la variable apoyo docente	63
Tabla 22. Análisis de la variable agotamiento emocional	64
Tabla 23. Niveles de ansiedad y estrés percibido	65
Tabla 24. Desmotivación académica	66
Tabla 25. Capacidad para afrontar retos académicos	68
Tabla 26. Sentido de pertenencia institucional	69
Tabla 27. Apoyo emocional por parte del docente	70
Tabla 28. Relación con amigos en el colegio	71
Tabla 29. Análisis de ausentismo	72
Tabla 30. Importancia percibida del bachillerato	73
Tabla 31. Situación familiar percibida	74
Tabla 32. Motivos laborales	75
Tabla 33. Motivos económicos	76
Tabla 34. Motivo de salud	77
Tabla 35. Responsabilidad familiar	78

Índice de ilustraciones

Ilustración 1. IA contra el abandono escolar	49
Ilustración 2. Procesamiento de datos y evaluación del modelo predictivo	56
Ilustración 3. Distribución de estudiantes según el curso	58
Ilustración 4. Diversidad de género de los participantes	59
Ilustración 5. Estadística descriptiva de comprensión de contenidos	60
Ilustración 6. Análisis de la variable cumplimiento de tareas	61
Ilustración 7. Análisis de la variable rendimiento y esfuerzo	62
Ilustración 8. Distribución de la carga académica	63
Ilustración 9. Análisis de la variable apoyo docente	64
Ilustración 10. Análisis de la variable agotamiento emocional	65
Ilustración 11. Niveles de ansiedad y estrés percibido	66
Ilustración 12. Desmotivación académica	67
Ilustración 13. Capacidad para afrontar retos académicos	68
Ilustración 14. Sentido de pertenencia institucional	69
Ilustración 15. Apoyo emocional por parte del docente	70
Ilustración 16. Relación con amigos en el colegio	71
Ilustración 17. Análisis de ausentismo	72
Ilustración 18. Importancia percibida del bachillerato	73
Ilustración 19. Situación familiar percibida	74
Ilustración 20. Motivos laborales	75
Ilustración 21. Motivos económicos	76
Ilustración 22. Motivo de salud	77
Ilustración 23. Responsabilidad familiar	78

Resumen

Investigaciones previas en abandono escolar han presentado la problemática en los últimos años de la Educación Superior, encontrando a los modelos tradicionales predictivos como herramientas limitadas al omitir variables críticas de importancia ante el fenómeno. El objetivo de este estudio se enfocó en el desarrollo de un modelo predictivo que detecta de manera temprana posibles riesgos de abandono a través de las funciones que contiene la Inteligencia Artificial para identificar por primera vez dimensiones socio emocionales, económicas y académicas que afectan la continuidad académica. Desde un análisis numérico y una profundidad descriptiva – asociativa, se compilaron datos de 102 estudiantes mediante instrumentos diseñados para calcular los factores de riesgo en el entorno escolar como la motivación y la autogestión. Por otro lado, en el procesamiento y análisis de datos se aplicó la herramienta Orange Data Mining que integra el algoritmo XGBoost de aprendizaje supervisado. En los resultados obtenidos se evidencia la capacidad predictiva que tiene el sistema para identificar las principales dimensiones de abandono en donde se destaca la dimensión académica con un 17,91% y la situación laboral con la salud en un 62,32%. Se establece por consiguiente que, al incluir variables socioemocionales en herramientas de IA, su valor incrementa debido a la precisión de detectar posibles riesgos de abandono. Este factor permite una intervención orientada al sentido de permanencia y bienestar estudiantil, lo que transforma una gestión en déficit en una atención preventiva. Por ende, el modelo proporciona tableros estratégicos de control que facilitan la elaboración de estrategias pedagógicas dirigidas a estudiantes con niveles críticos de abandono con el objetivo de mitigar la exclusión educativa.

Palabras clave: Inteligencia Artificial, abandono escolar, variables socioemocionales, modelo predictivo, XGBoost, Educación Superior

Abstract

Previous research on school dropout within Higher Education has highlighted the limitations of traditional predictive models, which frequently overlook critical subjective variables. This study aimed to develop an Artificial Intelligence (AI) early-warning predictive model to identify, through a multidimensional approach, the socio-emotional, economic, and academic factors jeopardizing educational continuity. Based on a numerical analysis and a descriptive-associative depth, data from 102 students were compiled using instruments designed to calculate risk factors in the school environment, such as motivation and self-management. Data processing was executed via **Orange Data Mining**, implementing the **XGBoost** supervised learning algorithm. The findings demonstrate a robust predictive capacity, revealing that the academic dimension accounts for a **17.91%** incidence rate, while the intersection of health and labor status reaches **62.32%**. Consequently, it is established that incorporating socio-emotional variables into AI frameworks significantly enhances diagnostic precision. Furthermore, the model provides strategic control dashboards that shift school management from a reactive stance to a preventive and ethical intervention, effectively mitigating educational exclusion.

Keywords: Artificial Intelligence, School dropout, Socio-emotional variables, Predictive model, XGBoost, Higher Education.

INTRODUCCIÓN

En el contexto actual, la educación presenta desafíos cruciales en la permanencia académica, debido a la ausencia de estrategias de acompañamiento por parte del DECE hacia grupos críticos. Por ende, las instituciones educativas, en los últimos años de Educación Básica Superior presentan indicadores estadísticos negativos que fracturan el desarrollo integral de los estudiantes. Este fenómeno, al interrumpir la trayectoria académica de los estudiantes, limita las oportunidades futuras y genera una desigualdad social de carácter permanente o difícil de romper (Castillo & Santilla-Lima, 2023). A nivel nacional, el abandono escolar ha sido tratado desde el análisis de variables socioeconómicas y académicas que el estudiante presenta después de dejar el sistema educativo.

El desarrollo de nuevas tecnologías (IA) que integran mecanismos de aprendizaje automático «Machine Learning» aportan una oportunidad para transformar este enfoque en un modelo predictivo que responde a las necesidades específicas de cada estudiante (Quiroz, 2025). Al tener una base sólida para el procesamiento de datos, permite identificar de manera temprana los principales factores de riesgo que presenta el estudiantado para facilitar las intervenciones por parte del DECE. Así, el estudio se enfoca en el concepto sobre los modelos predictivos actuales al ser deficientes por no incorporar dimensiones humanas del estudiante, lo que limita en la precisión de alertas tempranas ante la deserción escolar. Por esta razón, la propuesta integra variables socioemocionales como factores decisivos para mejorar la precisión en la detección temprana en grupos críticos.

La importancia de este estudio traspasa el enfoque administrativo y estadístico, al incorporar herramientas tecnológicas capaces de codificar datos socioemocionales en valores significativos con el propósito de capturar la realidad del estudiante. Al proporcionar un sistema de inteligencia artificial con la capacidad de traducir variables socioemocionales, los docentes adquieren información predictiva que permite tomar decisiones en base a las características individuales de cada estudiante (Gairín & Olmos, 2022). De esta manera, la investigación fortalece el sistema educativo en la retención escolar y en promover un ambiente sano en el transcurso de su preparación profesional.

Introducción a los capítulos

- **Capítulo I:** Examina el problema a investigar en base a un análisis en modelos predictivos deficientes utilizados en la detección de abandono escolar

en la Educación Básica Superior. Se describen los objetivos e hipótesis que guían este estudio en la necesidad de incorporar variables socioemocionales en modelos predictivos.

- **Capítulo II:** Representa el marco teórico referencial, apartado en el que se estudia los argumentos de la Inteligencia Artificial dirigida al contexto educativo, basados en estudios sobre el abandono escolar y el impacto de factores socioemocionales, económicos y familiares en el desempeño académico.
- **Capítulo III:** Explica el diseño metodológico, con un enfoque cuantitativo, la investigación describe la población de 102 estudiantes como el proceso de recolección de datos mediante el uso del programa Orange Data Mining que integra el algoritmo XGBoost.
- **Capítulo IV:** Se describe el análisis e interpretación de datos obtenidos mediante el programa. En este apartado se examinan las capacidades del sistema al integrar nuevas variables para identificar los factores críticos que afectan en la continuidad académica.
- **Capítulo V:** Dispone de conclusiones y recomendaciones relacionadas a la investigación, en el cual se destaca la confiabilidad de implementar herramientas tecnológicas avanzadas para detectar posibles riesgos de abandono con el fin de mitigar la exclusión escolar dentro del sistema educativo.

CAPÍTULO I

PROBLEMA DE INVESTIGACIÓN

1.1. Planteamiento del problema

El abandono escolar constituye uno de los desafíos más críticos del sistema educativo ecuatoriano, especialmente en los niveles de bachillerato. Según datos del Ministerio de Educación (2024), este fenómeno no solo interrumpe la trayectoria académica, sino que restringe el desarrollo integral y las futuras oportunidades económicas de los jóvenes. En tal sentido, investigaciones desarrolladas por Pachay y Rodríguez (2021) donde sostienen “que este fenómeno es el resultado de una concurrencia de vulnerabilidades, las cuales se vinculan con los factores económicos y familiares, con dimensiones críticas de salud mental”. De esta manera, variables como la ansiedad y el estrés actúan de manera silenciosa en la desvinculación formativa del estudiante antes de culminar su formación académica.

En base a la información detallada por algunos autores, mencionan que, al restar la resiliencia psicológica, intensifica la falta de conectividad con el aislamiento lo que, genera una ruptura en la vida del estudiante. De esta forma, los alumnos con mayor vulnerabilidad perciben el abandono escolar como una respuesta rápida con relación a su malestar emocional para evitar sentir la presión escolar causada por la carga académica. Según Castillo y Santillán-Lima (2023) señala que el abandono escolar en Ecuador se abordó a través de herramientas tecnológicas, a pesar de tener un vacío en la implementación de datos cualitativos. Sin embargo, para los investigadores, los índices de abandono escolar han reflejado cambios estadísticos en factores de riesgo como personales y familiares los cuales permanecen con un peso significativo en el proceso de la formación académica. Desde este punto, se busca implementar modelos que analicen cifras significativas con el propósito de capturar la realidad socioemocional del estudiante para detectar de manera temprana posibles riesgos de abandono.

Por consiguiente, la dependencia institucional y de convivencia ejerce una influencia importante en la permanencia estudiantil a través de la calidad del clima escolar y las diferentes estrategias pedagógicas usadas. Un ambiente educativo percibido como excluyente o poco empático acelera la desconexión emocional del adolescente, quien deja de hallar sentido a su asistencia diaria. Según Carrillo y González, (2023), señalan que cuando las

metodologías docentes no se implementan a las realidades diversas del alumnado, esto genera un sentimiento de indiferencia. Así, al tener una desconexión psicológica y académica por parte del alumno frente a los protocolos de convivencia, se obtiene una respuesta negativa en el centro educativo al transformar el espacio de estudio en un lugar de frustración.

En base a las teorías expuestas, las causas del abandono en los niveles educativos del bachillerato ecuatoriano se muestran mediante tres dimensiones críticas que trabajan de forma sistematizada. Para comenzar, la dimensión socioeconómica se presenta por medio de la limitación presupuestaria en los hogares junto a la presión por obtener una integración laboral prematura por lo que estos factores obligan al estudiante a enfocarse en la subsistencia antes de su culminación profesional. Según datos de UNICEF (2024), esta vulnerabilidad económica se ha agudizado en las zonas periféricas, donde el costo de oportunidad de estudiar resulta insostenible para familias de escasos recursos.

En la actualidad, las instituciones educativas en Ecuador emplean estrategias de seguimiento tradicionales, tales como el monitoreo de asistencia, seguimiento de planificaciones, informes disciplinarios y el control de rendimientos académicos. Sin embargo, herramientas tecnológicas que contienen un carácter reactivo y tardío, su enfoque se dirige en el riesgo que es visible en el estudiante después de presentar signos críticos de ausentismo y bajos niveles de rendimiento escolar. Según el Ministerio de Educación (2024), la brecha para la detección de posibles riesgos de abandono se ve limitada en la respuesta preventiva por medio del Departamento de Consejería Estudiantil (DECE) y en la coordinación académica, lo que ocasiona una gestión de crisis en lugar de combatir el aumento de este fenómeno. En consecuencia, al tener un límite en herramientas tecnológicas proactivas impiden que el personal académico no tenga la información adecuada para entender la complejidad del entorno estudiante lo que provoca que las intervenciones sean lentas al momento de tener un abandono irreversible.

De esta manera, es necesario elaborar una propuesta tecnológica que identifique los principales factores de riesgo de abandono escolar en relación a los datos académicos y socioemocionales, con el objetivo de producir alertas tempranas que favorezcan a las estrategias pedagógicas y a la orientación estudiantil. Según Gonzales et al., (2023), el estudio supera la limitación al sugerir un modelo predictivo que abarque el procesamiento de datos cualitativos para generar una mayor eficiencia en la detección temprana. Así, al responder esta brecha, el proyecto busca elevar la gestión educativa mediante la

implementación de una herramienta tecnológica que emite una alerta temprana en la gestión educativa y de esta manera el personal del DECE tome decisiones pedagógicas en base a la evidencia reflejada por el sistema.

1.2. Formulación del problema.

El sistema educativo en el nivel bachillerato general unificado y técnico enfrenta actualmente desafíos estructurales en su proceso. Por tal razón, se evidencia una limitada capacidad institucional para identificar de manera temprana el riesgo de abandono escolar. Este fenómeno se ve intensificado por la crisis actual del país que involucra la precarización laboral, la inseguridad y la brecha en el acceso a insumos tecnológicos que impactan directamente en el entorno familiar (Aguilera, 2022). Dado que, este factor genera una presión indirecta hacia estudiantes de último año escolar, optan por abandonar las instalaciones educativas lo que provoca un corte en su desarrollo académico y profesional.

Según datos oficiales del Ministerio de Educación, la Zona 9 (Distrito Metropolitano de Quito, 2024), la presencia de abandono escolar correspondiente al periodo 2022-2023 se refleja en un 1,72%, cifra que evidencia que más de mil jóvenes no culminan sus estudios debido a la presión económica y social. Cabe mencionar que, el problema radica en el sistema actual al no contar con herramientas analíticas predictivas que procesen el impacto de la crisis económica en el desempeño académico del estudiante. Por otra parte, en el estudio realizado por Pinales y Cedeño (2025), el reclutamiento de adolescentes en bandas criminales, se ve reflejado en los niveles de pobreza y la falta de conexión psicológica en el sistema educativo.

La problemática se define por la interacción de variables críticas donde los factores académicos, socioeconómicos y socioemocionales actúan como variables independientes (causas). Por el contrario, el abandono escolar se establece como la variable dependiente (efecto), siendo el fenómeno que se pretende predecir y mitigar (Mercado & De la Cruz, 2024). El vacío actual se encuentra en la ausencia de un modelo que integre estos datos de manera sistemática, ya que la mayoría de rasgos o reportes escolares son descriptivos y no predictivos, lo que impide una gestión proactiva para el hallazgo breve del riesgo de abandono temprano.

Desde este punto de vista, la investigación se localiza geográficamente en instituciones que abarcan la Zona 9 (Pichincha), con el objetivo de analizar a estudiantes ubicados en los últimos años lectivos. Es importante señalar que, este estudio se dirige en el desarrollo y la implementación de un modelo predictivo respaldado en la Inteligencia Artificial cuyos sistemas avanzados permiten gestionar las variables cuantitativas y cualitativas del estudiante como lo hace el programa Orange Data Mining. Por esta razón, al innovar el sistema educativo con nuevas tecnologías, se fortalece la orientación del personal estudiantil para determinar de manera precisa respuestas tempranas y estrategias preventivas que reduzcan el impacto del abandono escolar dentro de la región (Gairín & Olmos, 2022).

1.3. Preguntas de investigación.

¿Cómo puede un modelo predictivo de inteligencia artificial, detectar el riesgo de abandono escolar, integrando variables académicas y socioemocionales en estudiantes de bachillerato en el contexto de la educación ecuatoriana?

1.3.1. Objetivo general

Desarrollar un modelo de Inteligencia Artificial para predecir de manera temprana el riesgo de abandono escolar en bachillerato, integrando variables socioemocionales, con el fin de fundamentar estrategias de intervención mediante Orange Data Mining.

1.3.2. Objetivos específicos

- Identificar las variables socioemocionales y académicas más influyentes en el riesgo de abandono escolar.
- Evaluar mediante la observación sistemática, la interacción entre estudiantes y docentes de alto riesgo, con el fin de comprender los factores contextuales y relacionales asociados a la predicción.
- Diseñar estrategias pedagógicas y de acompañamiento socioemocional basadas en las alertas generadas por el modelo predictivo, para fortalecer la retención escolar en los estudiantes de bachillerato.

1.4. Justificación.

La educación en Ecuador se enfrenta a un gran desafío estructural significativo: el sistema actual no logra identificar de manera temprana el riesgo de abandono escolar en el bachillerato. Este fenómeno no es un evento aislado, sino un proceso que se va acumulando poco a poco, donde factores económicos, familiares y, especialmente, socioemocionales debilitan la permanencia estudiantil. Según Zamora Cusme et al., (2023), factores como el estrés, la desmotivación y las relaciones interpersonales conflictivas inciden directamente en la tasa de abandono, agravados por la crisis actual, marcada por la precariedad laboral, económica y la creciente inseguridad, exige un cambio de paradigma donde la prevención deje de ser una respuesta administrativa de papeleo y se convierta en un enfoque científico fundamentado en datos reales del momento.

Al respecto, estudios recientes destacan la inseguridad nacional como un factor que agrava la vulnerabilidad en las instalaciones educativas. Según, Parrales y Cedeño (2025), al tener una mayor presencia en el reclutamiento de adolescentes en grupos delictivos, este factor se relaciona con la falta de apoyo por parte del docente y en la estabilidad socioemocional del alumno. Al ubicarse en zonas afectadas de pobreza y escasez de oportunidades, los estudiantes de bachillerato tienden a ser más propensos a optar por abandonar sus estudios debido a la falta de redes de apoyo que prevé el riesgo, ya que se verán atraídos por promesas de protección o ingresos rápidos. Desde una perspectiva geoespacial, en Ecuador se evidencian problemas en el desempeño académico junto con una crisis que vulnera la protección social e impacta en las vidas de los adolescentes lo cual, el sistema educativo busca herramientas innovadoras dirigidas en el monitoreo constante de datos mixtos y así tener una intervención personalizada para cada estudiante en peligro.

Al implementar modelos predictivos orientados en Inteligencia Artificial (IA) evoluciona la gestión educativa dando paso a la elaboración de soluciones innovadoras y esenciales para mitigar este fenómeno. Sin embargo, herramientas que usan la IA, procesan datos del rendimiento académico, narraciones y dimensiones socioafectivas de los estudiantes de una forma rápida y precisa lo que lo posiciona como un sistema eficaz en comparación a los métodos tradicionales. Según Zabala Riveras (2025), modelos como NUMELA, integran la IA en técnicas avanzadas que logran anticipar el riesgo de abandono de manera exacta, por lo que contribuye al personal educativo en el desarrollo de estrategias focalizadas y oportunas ante el abandono escolar. De este modo, las herramientas que contienen sistemas de IA en su

programación se convierten en un punto clave para responder las brechas sobre la detección y optimización de los recursos institucionales.

La presente investigación se justifica en la Unidad Educativa "Nelson Torres" al proponer un modelo que integra herramientas como Orange Data Mining, y chatbots para el análisis de datos mediante procesamiento de lenguaje natural. Al capturar variables socioemocionales que suelen ser invisibles para los mecanismos convencionales, este estudio ofrece un enfoque integral y humano a la predicción del abandono escolar (Eager & Brunton, 2023). Su valor social y académico radica en ofrecer una ruta metodológica clara para que otras instituciones ecuatorianas utilicen tecnologías emergentes en la prevención del abandono escolar. Con ello, no solo se busca mejorar indicadores institucionales, sino salvaguardar el futuro y las oportunidades de los estudiantes.

1.5. Formulación de hipótesis y determinación de variables.

1.5.1. Hipótesis general

El modelo predictivo de Inteligencia Artificial (IA), desarrollado mediante Orange Data Mining, que integra variables académicas (ej. calificaciones, asistencias) y socioemocionales (ej. autoconcepto, resiliencia), será capaz de predecir el riesgo de abandono escolar en estudiantes de bachillerato.

1.5.2. Hipótesis variables

- Factores predictivos: Las variables socioemocionales especialmente las relacionadas con la motivación y la autogestión emocional tendrán una asociación más fuerte y significativa con el riesgo de abandono escolar en los estudiantes de bachillerato, que las variables académicas tradicionales.
- Aporte de la IA: Al incorporar técnicas de Procesamiento de Lenguaje Natural (PLN), asistidas por “Orange Data Mining” garantiza un análisis en las dimensiones subjetivas lo que, mejora de manera significativa el rendimiento predictivo dentro del modelo de IA en comparación a modelos que se basan en la detección de valores cualitativos.
- Efectividad de la intervención: En la identificación temprana de estudiantes con niveles críticos facilita al modelo predictivo generar información relevante de cada estudiante sobre su estado psicoeducativa y de cómo el docente debe interactuar con

el estudiante para disminuir el índice de "Alto Riesgo" a una permanencia escolar en los siguientes años lectivos.

1.5.3. Variable dependiente

Es el resultado que el modelo busca anticipar, el abandono escolar en los estudiantes de bachillerato.

1.5.4. Variables independientes

El modelo utilizará factores que se le proporcionará para estimar la probabilidad de la variable dependiente. La dimensión académica, tanto como su promedio trimestral, número de materias reprobadas e historial de disciplina. La socioeconómica tanto para la situación laboral del representante como el acceso que pueda tener a conectividad o dispositivos electrónicos. La dimensión socioemocional para conocer las escalas de estrés/ansiedades recolectadas vía DECE. Y el clima escolar para conocer su convivencia y percepción de seguridad en el aula, y de existir algún reporte de conflictos entre pares.

1.5.5. Variables moderadoras y de intervención

Estas variables pueden afectar la relación entre las variables predictoras y el abandono escolar, y son clave para la fase cualitativa/propositiva. El modelo predictivo de IA, que se usará como herramienta tecnológica que procesa las variables independientes para calcular la probabilidad de ocurrencia de la variable dependiente.

CAPÍTULO II

MARCO TEÓRICO REFERENCIAL

2.1. Antecedentes de la investigación

El fenómeno del «abandono escolar» en el nivel de bachillerato representa uno de los nudos críticos por resolver para las políticas públicas educativas en América Latina, y adquiere una crisis inminente para el contexto ecuatoriano actual. En este sentido, el modelo de gestión para la permanencia se ha basado en modelos estadísticos descriptivos y reactivos, los cuales impedían la intervención de cualquier guía preventiva al no poder registrar la desvinculación temprana. Por ello, enfoques proactivos fundamentados en Inteligencia Artificial (IA) y en Analítica de Aprendizaje (Learning Analytics) han marcado un cambio en el paradigma de la gestión sobre la retención estudiantil (Bolaño-García & Duarte-Acosta, 2024).

La comprensión del «abandono escolar» ha evolucionado en tres etapas que las diferencian en la literatura científica reciente. En primer lugar, el conocimiento dominaba en una demografía estática que estudia las variables como el ingreso familiar, la ubicación geográfica y la formación académica de los padres. Estos datos al permitir identificar grupos de riesgo, carecían de capacidad para predecir el momento exacto del abandono. La segunda etapa, el auge del «Big Data» y el «Machine Learning», movió las trayectorias del aprendizaje digital (Barra & Pilicita, 2025). Por ende, el conocimiento elevó la comprensión del «compromiso» mediante el análisis de la interacción en plataformas virtuales. Es decir, la era de los modelos de clasificación binaria alcanzan altas precisiones técnicas que a menudo operan como «cajas negras», ofreciendo predicciones certeras.

Por último, en la tercera etapa el conocimiento actual reconoce que el rendimiento académico es solo la superficie de un ecosistema emocional complejo. Dado que, al evolucionar el conocimiento nos lleva ahora a integrar el Procesamiento de Lenguaje Natural (NLP) para capturar la subjetividad del alumno. Sin embargo, la trayectoria de nuestra investigación se justifica no solo en predecir el abandono mediante datos numéricos, sino en buscar «humanizar» la IA integrando a Orange Data Mining para interpretar las narrativas estudiantiles, cerrando así el ciclo evolutivo del conocimiento en este campo (Golec & Hachaj, 2025). Al dar un cambio en el marco técnico y epistemológico, trasciende el

conocimiento del abandono a un acontecimiento ocasional que comprende un proceso multicausal y dinámico en la inclusión digital.

Esta sección aborda la trayectoria del conocimiento desde la cuantificación del rendimiento académico hasta la codificación de la subjetividad del estudiante. Para garantizar que los antecedentes seleccionados no constituyan unidades aisladas, se han aplicado criterios de selección estratégicos que aseguran un hilo conductor coherente con los objetivos de la presente investigación. Así, se prioriza en esta investigación que, al ser analizadas en conjunto, revelan las limitaciones de los modelos actuales y, por ende, la necesidad de nuestra intervención.

Tabla 1 Criterios a evaluar en el estudio

Criterio de Selección	Requisito Técnico y Académico	Propósito en el Informe
Vínculo Temático Estricto	Estudios que abordan el «abandono escolar» mediante IA, Learning Analytics y «análisis socioemocional»	Asegurar que los hallazgos previos sean directamente aplicables a las «variables del estudio» (Barra & Pilicita, 2025)
Vigencia Cronológica	Prioridad de investigaciones publicadas entre 2023 y 2025.	Capturar la «vanguardia en algoritmos» «XGBoost, LLMs» que evolucionan rápidamente (Barra & Pilicita, 2025).
Calidad de la Fuente	Artículos en revistas y tesis doctorales de alto impacto.	Asegurar el rendimiento de las «métricas de precisión» metodológica (Quiroz, 2025).
Diversidad Geográfica	Inclusión de informes globales y locales en relación con el estudio.	Contrastar la eficacia de los modelos en entornos de «alta» y «baja digitalización» (Soriano-Sánchez & Jiménez-Vázquez, 2023).
Aporte Metodológico	Selección de estudios que utilicen «modelos de ensamble», procesamiento de «lenguaje natural» (NLP) o IA explicable (XAI).	Justificar la elección de herramientas avanzadas frente a «métodos estadísticos tradicionales» (Barra & Pilicita, 2025).

Nota: Elaboración propia basada en Inteligencia artificial generativa (IA Gen) en la transformación digital de la educación superior una revisión sistemática de literatura (2025).

Los criterios escogidos a evaluar permiten que cada antecedente contribuya al conocimiento que se reconstruye en una narrativa evolutiva. En primer lugar, se realizó un criterio coherente y adecuado junto a la relación lógica entre las variables del tema a estudiar, priorizando investigaciones que aborden la relación entre modelos de aprendizaje automático (Machine Learning) y el riesgo de abandono (Golec & Hachaj, 2025). Los estudios seleccionados no solo presentarán resultados de precisión algorítmica, sino que detallarán las dimensiones académicas, demográficas y socioemocionales utilizadas para alimentar el

algoritmo del sistema. En segundo lugar, se analizó información de métodos y tecnologías de estudios actuales, limitando la búsqueda a publicaciones realizadas entre 2020 y 2025.

Este parámetro es importante para la arquitectura de los modelos de IA, dado que los modelos de lenguaje de gran escala (LLM), han experimentado cambios significativos en un periodo corto. Al usar estudios previos a 2020 sus metodologías carecen de innovación tecnológica frente al nivel de innovación propuesto. En tercer lugar, se trazó un enfoque en la diversidad geográfica y contextual en la organización de la literatura que transita de lo global a lo nacional. Los aportes internacionales (Canadá, China y Europa) fueron seleccionados por su estabilidad técnica y validación de algoritmos en plataformas digitales (Carrillo & Gonzáles, 2023). Por el contrario, los antecedentes nacionales se escogieron por su capacidad de ilustrar la forma en que la vulnerabilidad se presenta en el bachillerato fiscal. De esta manera, permite verificar el desempeño del sistema con la realidad de las instituciones de la Zona 9 y Manabí.

Por último, en el contexto de la calidad metodológica de cada investigación se realizará un análisis de las herramientas utilizadas para la recolección de datos y de su validación. Por lo tanto, al seleccionar artículos de revistas y tesis de posgrado, garantiza que la investigación contenga información confiable y basada en conocimientos aplicados (Carrillo & Gonzáles, 2023). Así, esta investigación responde al sistema de ordenamiento y contextual diseñada para fortalecer el marco teórico-referencial del estudio. El siguiente apartado examina de manera exhaustiva los antecedentes investigativos que sustentan la necesidad de integrar variables socioemocionales, a través del uso de modelos de analítica predictiva, contrastando la evidencia global y la realidad del Ecuador.

2.1.1. Antecedentes internacionales

El análisis de la «deserción escolar» a nivel internacional ha evolucionado desde aquel puramente demográfico hasta la comprensión de las trayectorias de aprendizaje mediante el uso de algoritmos. A continuación, se presentan los estudios de vanguardia en la identificación de riesgos académicos mediante el procesamiento de datos complejos.

“Modelo basado en aprendizaje automático para la predicción del abandono escolar en la formación en línea”

Referencia: Zerkouk, M., et al. (2024).

El objetivo central de este estudio consistió en abordar la problemática del abandono estudiantil en la modalidad de educación a distancia, una de las áreas con mayores tasas de deserción a nivel global. El estudio buscó diseñar un modelo de clasificación binaria capaz de anticipar los niveles de actividad y el riesgo de abandono de los estudiantes, realizando una detección temprana fundamentada en el compromiso y los logros académicos actuales. Así, la investigación partió de la premisa que, «la identificación oportuna de factores críticos de éxito e interacción dentro de entornos de aprendizaje virtuales permite a las instituciones implementar intervenciones personalizadas antes de que la desvinculación sea irreversible».

El estudio se apoya metodológicamente en un conjunto de datos exhaustivo proveniente de una Plataforma de Aprendizaje a Distancia en Canadá, el cual integraba 49 características (*features*) que abarcaban desde dimensiones sociodemográficas hasta datos de comportamiento y rendimiento académico. La ejecución técnica se realizó plenamente en lenguaje Python. Por lo tanto, el plan sistemático del modelo incluyó un sistema perfilado de estudiantes por medio del algoritmo «XGBoost» (eXtreme Gradient Boosting), el cual se escogió debido a la robustez para tratar métodos de ensamble incluyendo la capacidad intrínseca para seleccionar las características con mayor relación predictiva (Golec & Hachaj, 2025). Así mismo, se utilizó el algoritmo de aprendizaje automático «machine learning» para validar la eficacia del modelo.

Los resultados de este estudio, demostraron que el uso de técnicas avanzadas de aprendizaje automático permite alcanzar una precisión (*accuracy*) del 82% en datos no vistos del año académico posterior. A través de matrices de correlación y técnicas de importancia de características, el modelo reveló que el comportamiento y la interacción dentro de la plataforma (como los clics en el entorno virtual de aprendizaje) poseen un peso predictivo significativamente mayor que las variables demográficas aisladas (Hernández, 2023). En consecuencia, el estudio concluyó que estos modelos son herramientas vitales para la personalización de la enseñanza y la creación de sistemas de alerta temprana efectivos.

Bajo este contexto, el aporte de esta investigación radica en la validación del algoritmo XGBoost como una herramienta de alta precisión para el análisis de trayectorias estudiantiles. Según Basantes et al., (2024), el presente estudio en el bachillerato ecuatoriano se diferencia al trasladar este enfoque de la educación superior virtual a la educación media presencial en un contexto de vulnerabilidad socioeconómica. Mientras que, el estudio de

Zerkouk (2024) se centra en el comportamiento dentro de una plataforma digital. De ahí que, el estudio propuesto integra el procesamiento de lenguaje natural (NLP) a través de Orange Data Mining para analizar la subjetividad socioemocional y las narrativas de los estudiantes, factores que en el Ecuador actual son determinantes debido a la crisis de inseguridad y precariedad laboral que no están presentes en el «dataset» canadiense.

Por otra parte, se menciona otro tipo de antecedente que radica en el análisis enfocado en la Inteligencia Artificial, donde se desarrollan las variables socioemocionales desde la literatura científica existente. Se presenta a continuación el estudio analizado:

“Análisis de aprendizaje impulsado por IA para el desarrollo metacognitivo y socioemocional: una revisión sistemática”

Referencia: Pacheco et al., (2025).

Esta investigación tuvo como objetivo explorar el estado del arte de la Analítica de Aprendizaje (*Learning Analytics*) potenciada por IA, centrando su atención en cómo estas tecnologías contribuyen al desarrollo de competencias metacognitivas y socioemocionales. De este modo, los autores buscaron identificar las brechas existentes en la investigación actual, la cual tradicionalmente se ha sesgado hacia indicadores cuantitativos de rendimiento académico, descuidando el bienestar integral y el desarrollo holístico del estudiante. A su vez, la metodología siguió rigurosamente las directrices del marco PRISMA para garantizar una revisión sistemática organizada.

El proceso incluyó la identificación de 6,782 documentos en la base de datos *Scopus* publicados entre 2013 y 2023, de los cuales se seleccionó una muestra probabilística de 346 documentos para un cribado inicial y finalmente 161 artículos para una lectura profunda. De este modo, el análisis de datos combinó enfoques cuantitativos (análisis de frecuencias de tipos de analítica y niveles educativos) y cualitativos (análisis de contenido sobre implicaciones pedagógicas guiado por los constructos de Aprendizaje Autorregulado y Aprendizaje Social y Emocional) (Infante & Araya, 2020). Los resultados revelaron una marcada desproporción en las áreas de aplicación de la analítica de aprendizaje. La analítica predictiva domina el campo con un 46% de presencia, seguida de la prescriptiva con un 28%. Sin embargo, la analítica socioafectiva se encuentra subrepresentada con apenas un 10%, lo que genera preocupaciones críticas sobre la capacidad de los sistemas actuales para apoyar competencias como la autoconciencia emocional y el aprendizaje colaborativo.

En este sentido, el estudio subraya que la IA tiene un potencial creciente en la detección de estados afectivos y comportamientos autorreguladores, permitiendo intervenciones oportunas siempre que se diseñen con transparencia y sensibilidad ética. De este modo, el aporte de este estudio radica en la revisión sistemática que provee la justificación teórica y científica para la necesidad de «humanizar» la IA educativa. Así, el aporte principal consiste en la identificación del vacío en la analítica socioafectiva (10%). De este modo, la presente investigación se diferencia al posicionarse precisamente en este nicho desatendido, utilizando Orange Data Mining no solo para predecir si un estudiante abandonará el aula, sino para codificar y ponderar patrones socioemocionales complejos que suelen ser invisibles para los modelos cuantitativos tradicionales de predicción (Carrillo & Gonzáles, 2023).

Como antecedente final en el ámbito global, se presenta a continuación un estudio enfocado en la predicción del abandono escolar mediante el análisis de la Inteligencia Artificial, que a su vez es fundamentado por datos académicos, socioeconómicos, y demográficos. Se presenta a continuación la investigación:

“Predicción de la deserción estudiantil mediante aprendizaje conjunto con análisis de IA explicable basado en SHAP”

Referencia: Liu, Z et al., (2025).

El propósito de este estudio fue el desarrollo de «EASE-Predict» (Ensemble-SHAP Explainable Student Prediction), en donde se explica el marco de aprendizaje por conjuntos diseñado que predice los resultados académicos de estudiantes para entender el por qué mediante valores «SHAP» (Shapley Additive Explanations) (Hernández, 2023). El objetivo se centraba en resolver las «cajas negras» de la IA tradicional, con el fin de proporcionar al grupo de docentes y administradores información fundamentada sobre el por qué un estudiante es clasificado en riesgo. Metodológicamente, evaluaron cinco algoritmos de aprendizaje automático (*Random Forest*, *Gradient Boosting*, *Extra Trees*, *Regresión Logística* y *SVM*) sobre un «dataset» de 4.424 estudiantes con 36 características que incluían datos académicos, socioeconómicos y demográficos (Golec & Hachaj, 2025).

Desarrollaron modelos de votación y apilamiento (*stacking*) para mejorar la precisión. La validación la realizaron mediante pruebas estadísticas de McNemar y validación cruzada para asegurar la estabilidad del modelo. En los resultados obtenidos por el modelo

«EASE-Predict» se visualizó una precisión de 77,4%, cuyo porcentaje mejora estadísticamente los modelos individuales (Barra & Pilicita, 2025). En consecuencia, el análisis «SHAP» explicó que al culminar las materias curriculares del segundo semestre refleja un 60% de influencia predictiva, seguida con el pago de matrícula de 35% con la disponibilidad de obtener becas de un 12%. El estudio concluyó que, es primordial combinar la precisión con la interpretabilidad con el propósito de que las instituciones educativas integren estas herramientas tecnológicas con integridad.

El aporte de este estudio radica en la introducción de la «IA explicable» en el ámbito educativo, permitiendo que la predicción no sea un número frío sino una ruta de intervención. Sin embargo, el presente estudio es necesario porque adapta la necesidad de explicabilidad al contexto del bachillerato público en Ecuador, donde las variables de mayor peso no son necesariamente el «pago de la matrícula» o las «becas» (como en el estudio internacional de educación superior) (Castillo & Santilla-Lima, 2023). Sino más bien, factores como la desmotivación intrínseca, la falta de apoyo en el hogar y la presión del entorno social e inseguridad (Nuñez, 2023). En otras palabras, la diferenciación radica en el uso de Orange Data Mining para generar explicaciones cualitativas sobre el estado emocional del alumno, complementando los datos numéricos con un análisis profundo del entorno del estudiante.

2.1.2. Antecedentes nacionales

En Ecuador, la investigación sobre la «deserción escolar» ha cobrado una relevancia crítica debido al impacto de la pandemia y la situación de seguridad nacional. Los estudios nacionales están transitando de la descripción del fenómeno hacia la aplicación de herramientas tecnológicas avanzadas. A continuación, se presentan los estudios de vanguardia en la identificación de riesgos académicos mediante el procesamiento de datos complejos.

“Factores de deserción escolar de los estudiantes de la Unidad Educativa Leónidas Plaza km. 20”

Referencia: Maldonado, R., & Soledispa, S. (2024).

El objetivo de esta investigación consistió en identificar y analizar los factores que inciden en la «deserción escolar» de estudiantes de bachillerato en una zona rural de Manabí, marcada por actividades económicas familiares como la agricultura y la pesca. El estudio

buscó categorizar las causas del abandono en dimensiones individuales, familiares y escolares para comprender la complejidad del fenómeno en contextos de vulnerabilidad. La metodología se fundamentó en una revisión teórica y el análisis comparativo de registros escolares oficiales del periodo 2023-2024. Se centró en una población de bachillerato, utilizando un enfoque descriptivo para agrupar diez factores críticos identificados en tres dimensiones principales. No se utilizaron algoritmos de IA, más bien su análisis se realizó mediante herramientas estadísticas tradicionales para determinar tendencias y proporciones. De este modo, los resultados revelaron una tasa de deserción del 9,43% en la unidad educativa analizada.

Las causas principales detectadas fueron el desinterés académico y el compromiso adolescente, tales como la necesidad de trabajo prematuro o responsabilidades familiares. El estudio destacó que el primer año de bachillerato es el momento de mayor criticidad para el abandono, debido a la transición educativa y los cambios propios de la etapa del desarrollo. Según Aguilar et al., (2025), es importante la información que entrega los modelos, ya que, aportan una visión realista y necesaria en base al contexto ecuatoriano sobre los niveles de bachillerato al ser la parte más vulnerable del sistema educativo. Por otro lado, existe una diferencia con el presente estudio al radicar el paso de la descripción a la predicción. Sin embargo, para Maldonado y Soledispa (2024) explican el por qué los estudiantes dejan sus estudios, y propone un modelo proactivo que utilice la IA para detectar factores de riesgo antes de que ocurra el abandono. El uso de Orange Data Mining para procesar datos socioemocionales recolectados por el DECE permitirá automatizar la detección que en el estudio de Manabí se realizó de forma manual y retrospectiva.

El uso de Orange Data Mining para procesar datos socioemocionales recolectados por el DECE permitirá automatizar la detección que en el estudio de Manabí se realizó de forma manual y retrospectiva.

Con el propósito de analizar el modelo predictivo desarrollado en esta institución educativa, se establecerán características y funciones indispensables en la prevención del abandono escolar. Se presenta a continuación la siguiente investigación:

“NUMELA: Un modelo predictivo para prevenir el abandono escolar mediante Inteligencia Artificial y Neuromarketing”

Referencia: Zabala Riveras, J. (2025).

El estudio tuvo como propósito central el desarrollo del modelo NUMELA, el cual integra técnicas de Inteligencia Artificial para predecir comportamientos estudiantiles y optimizar los recursos de gestión educativa, buscando crear un sistema más inclusivo y eficiente. De este modo, el investigador buscó validar los modelos predictivos avanzados incluyendo redes neuronales artificiales con el fin de superar los métodos tradicionales. Por ello, la metodología usada fue de tipo hermenéutico, basándose en datos informativos que tiene las instituciones educativas respecto a su rendimiento, factores socioeconómicas y conductuales. Así, el estudio se posicionó como una referencia en el uso de «neuromarketing» aplicado a la educación para entender la fidelización y permanencia del alumno en el sistema. En consecuencia, los resultados indicaron que las redes neuronales poseen una capacidad superior para reconocer patrones de riesgo no lineales en comparación con los modelos estadísticos básicos.

Además, se determinó que factores como el entorno familiar y la asistencia escolar tienen una incidencia crítica, pero también se identificaron barreras como la brecha digital y la falta de formación docente como obstáculos para la implementación real de estos modelos en las escuelas ecuatorianas. Consecuentemente, el aporte de este estudio radica en la introducción de una visión transdisciplinaria «IA y Neuromarketing» para la retención. Así, la presente investigación es necesaria porque toma la base tecnológica de NUMELA y la aplica de forma práctica en instituciones de la Zona 9 utilizando Orange Data Mining, una herramienta de procesamiento de lenguaje natural que facilita la interacción humana (Hernández, 2023). A diferencia de NUMELA, que se basa en simulaciones y redes neuronales densas de difícil interpretación para el docente común, este estudio propone una interfaz más accesible y centrada en variables socioemocionales capturadas directamente de la interacción docente-estudiante, alineándose con la realidad de los Departamentos de Consejería Estudiantil.

2.2. Estado del arte

En Ecuador, la desescolarización en la Educación General Básica (EGB) Superior equivale a un sesgo de resultados entre la inversión estatal en la formación académica y la desvinculación del estudiante de su sistema educativo (Castillo & Santilla-Lima, 2023) . Por consiguiente, al quebrantar su trayectoria en la Educación Media Superior (EMS), se causa una pérdida social y económica que prolonga ciclos de exclusión, limitando a los jóvenes su capacidad para integrarse a una ciudadanía productiva y segura. Según Concepción et al.,

(2024), el «abandono» a nivel bachillerato actúa como un multiplicador de desigualdades, que limita de manera severa el desarrollo integral de los jóvenes y sus futuras oportunidades de «inserción laboral calificada».

El historial de la tasa de abandono educativo en Ecuador está marcado por las volubles políticas públicas contemporáneas y las crisis disruptivas que implican una multiplicidad de variables (macroeconómicas, demográficas, seguridad) que convergen en los últimos diez años. De hecho, en el año lectivo 2010-2012 los registros administrativos del Ministerio de Educación (2024) contabilizaron una tasa de abandono estudiantil de 229.620 jóvenes, cifra que motivó una reestructuración de los programas de inclusión educativa, factor que logró estabilizar el régimen institucional por debajo de 100.000 casos, a partir del año 2018-2019 (Ministerio de Educación del Ecuador, 2024). Así, por medio de políticas públicas centralizadas se logró una reducción paulatina de abandono escolar en la educación secundaria, pasando de un 6,8% en 2011 a un 4,3% en 2018 (Castillo & Santilla-Lima, 2023).

Es decir, en siete años se reflejó un descenso de 2,5 puntos porcentuales, gracias a la eficacia de las intervenciones gubernamentales en dimensiones como infraestructura, calidad curricular, apoyo socioeconómico y marcos legales inclusivos (Castillo & Santilla-Lima, 2023). Así mismo, para comprender el impacto de la reducción del abandono escolar (de 6,8% al 4,3%), resulta primordial señalar la naturaleza multidimensional del fenómeno en el contexto andino. Por un lado, esta cifra evidencia el bajo desarrollo de habilidades tecnológicas en los estudiantes, que a futuro los dirige hacia la precariedad laboral, vinculada con la informalidad económica por falta de competitividad (Concepción, et al., 2024). Por el otro, se observa la fractura del desarrollo integral, que interrumpe la formación de habilidades blandas, como la resiliencia psicológica necesaria para la vida adulta.

Respecto a la crisis sanitaria derivada del COVID-19, que erosionó la vulnerabilidad preexistente, al transformar las brechas digitales de acceso, en una desconexión sistémica del estudiante con el entorno escolar-familiar. Dado que, el cierre prolongado de las instituciones, obligó a una transición de emergencia hacia la educación telemática, la cual desnudó una infraestructura digital insuficiente. A modo de ejemplo, en la Zona 9, el 37% de los hogares carecía de dispositivos adecuados para el aprendizaje sincrónico (Pachay-López & Rodríguez-Gámez, 2021). Este hecho no solo interrumpió la adquisición de competencias curriculares, sino que generó una demanda de aprendizaje, que los modelos tradicionales no logró condensar (Carrillo & Gonzáles, 2023). Según UNICEF (2024), se entiende por

«brecha digital de acceso» a la falta de insumos tecnológicos y conectividad que durante el aislamiento debilitaron la resiliencia del sistema.

A su vez, el impacto de la pandemia trascendió la educación para incidir en la salud mental y la estabilidad financiera de los núcleos familiares. Este acontecimiento reflejó un aumento tanto en la tasa de desempleo como en el encarecimiento económico, que forzó a muchos estudiantes de Bachillerato a priorizar su subsistencia sobre la escolaridad (INEC, 2025). Además, el aislamiento social debilitó la resiliencia psicológica, al incrementar los índices de ansiedad y desmotivación intrínseca durante las cuarentenas (UNICEF, 2024). Por ese motivo, para que un Sistema de Alerta Temprana (SAT) sea efectivo en la actualidad, debe considerar el periodo de pandemia como un punto de inflexión estadístico, donde la «deserción» no fue un evento lineal, sino una fractura multifactorial impulsada por la falta de soporte socioemocional y la carencia estructural.

De este modo, la ausencia de un entorno institucional protector incrementa la permeabilidad hacia redes delictivas que incrementa el abandono escolar. En el estudio realizado por Tomalá, et al., (2025) señalan que, el reclutamiento por bandas criminales es una amenaza creciente, que advierte un desplazamiento geográfico del riesgo. Si bien, esta crisis se presenta más crítica en la Costa, actualmente ya se visualiza con fuerza en las instituciones educativas de la Sierra (Zona 9). Por esa razón, la fortaleza del entorno institucional representa la última línea de defensa para que el estudiante no sucumba a las presiones externas de subsistencia y supervivencia (Parrales & Cedeño, 2025).

En este sentido, el sistema educativo es objetivo de las fallas estructurales del Estado en la captación y permanencia de estudiantes, cuyas condiciones de vulnerabilidad los obliga a dar preferencia a la estabilidad económica sobre su preparación académica (García y otros, 2022). De la misma manera, la presencia de trabajo no adecuado para los padres de familia, actúa como el principal factor externo que obliga a los estudiantes a priorizar la supervivencia inmediata sobre la formación académica. En otras palabras, la educación no es solo un derecho primordial para el buen vivir de los estudiantes, sino también una oportunidad para vincularse en favor de una inserción laboral digna (Vélez-Miranda, et al., 2020).

Así, el punto crucial normativo aparece en la Constitución de la República (2008), que garantiza la educación como una prioridad de la política pública y de la inversión estatal, salvaguardando la gratuidad hasta los estudios superiores. Así mismo, en una declaración

solemne de la Ley Orgánica de Educación Intercultural (2015) se indica el andamiaje legal necesario para implementar políticas de «retención». Esta medida consigna al Estado la creación de mecanismos para asegurar la permanencia y el éxito escolar de todos los estudiantes, especialmente de aquellos en situaciones de vulnerabilidad incluidas las personas con discapacidad.

Al emplear mecanismos tradicionales en los centros de monitoreo sobre la asistencia y los promedios de cada estudiante, resulta insuficiente para detectar el problema a resolver, después de que el abandono escolar tomo un peso irreversible. Así, al existir una brecha en la detección crítica, los Departamentos de Consejería Estudiantil (DECE) no logran cauterizar el fenómeno que aumenta de manera progresiva dado que su enfoque se dirige al área administrativa en lugar de una protección holística. Por ende, el rol del DECE evoluciona en su labor de consejería al encontrar respuestas frente a un sistema que colapsa debido a la complejidad en la gestión institucional. Por consiguiente, la investigación contemporánea sugiere que, la persistencia del abandono en instituciones fiscales, que concentran el 83,5% de los casos nacionales, evidencia una vulnerabilidad estructural que demanda herramientas de analítica avanzada para una intervención proactiva (Gobierno de la República del Ecuador, 2024).

Tabla 2. Tasa de abandono nacional.

Variable temporal	Tasa de abandono nacional (%)	Casos de abandono (Absoluto)	Contexto sociopolítico
2011-2012	6,80%	229.620	Expansión del «sistema público» (Gobierno de la República del Ecuador, 2024).
2013-2014	5,22%	N/A	«Reformas curriculares» iniciales (Gobierno de la República del Ecuador, 2024)
2017-2018	4,30%	N/A	Programas de «inclusión» (Gobierno de la República del Ecuador, 2024).
2019-2020	1,73%	N/A	Inicio de la «pandemia» (Gobierno de la República del Ecuador, 2024).
2023-2024	1,75%	72.644	Crisis de «seguridad» y «trabajo informal» (Gobierno de la República del Ecuador, 2024).

Nota. Elaboración propia basada en el Gobierno de la República del Ecuador (2024).

En este escenario, el tránsito hacia un modelo predictivo basado en IA no es solo una innovación tecnológica, sino una necesidad operativa para el Departamento del DECE y las

autoridades educativas. Dado que, la estabilidad institucional exige una transición técnica hacia la prevención del analfabetismo, cuya labor orientadora se fundamente en datos y no solo en la intuición. Al respecto, una base institucional sólida puede gestionar la dimensión más profunda y subjetiva del estudiante: su mundo socioemocional. Así, la estabilidad emocional se configura en el ancla para responder, resistir, superar las adversidades y alcanzar el éxito académico. De igual forma, los factores psicológicos suelen actuar como promotores silenciosos del abandono escolar, lo que conlleva a una falla para los reportes administrativos tradicionales.

Por consiguiente, un estudiante sin habilidades de autorregulación optará por el abandono como la única solución inmediata ante el malestar familiar insostenible. De acuerdo con Pachay y Rodríguez (2021), existen variables críticas que deben ser codificadas con rigor científico por la falta de apoyo; factores como el estrés y la ansiedad derivados de la presión académica y la inestabilidad familiar, tienden a erosionar la salud mental. A su vez, la desmotivación intrínseca genera la pérdida del sentido personal respecto al logro educativo y la visión de futuro. De manera similar la falta de apoyo familiar y la baja resiliencia psicológica, no solo fomentan la ausencia de una red de validación que valore el esfuerzo académico del estudiante, sino también la incapacidad de adaptarse frente a crisis, intensificada por el trabajo infantil y la brecha tecnológica de acceso.

Por ese motivo, la complejidad de las variables subjetivas requiere de herramientas tecnológicas avanzadas que las vuelva medibles en términos prácticos, especialmente aquello que, la analítica tradicional no alcanza a capturar como la desmotivación e idealización. Por ello, la integración de herramientas de vanguardia resulta necesaria para transformar estos estados cualitativos en indicadores de riesgo accionables. De ahí que, la literatura especializada ha transformado las definiciones teórico-administrativas del abandono hacia conceptos prácticos que evidencian la ruptura del vínculo entre el estudiante y su institución. Vincent Tinto, en su modelo clásico de integración, postula que la persistencia estudiantil es una función del equilibrio entre la integración académica y la calidad de las interacciones con pares y docentes (Infante & Araya, 2020).

Por ello, cuando los compromisos iniciales del estudiante se ven perjudicados por experiencias negativas en cualquiera de estos dos ámbitos, el individuo entra en un proceso de «alejamiento voluntario o forzoso». El modelo psicológico de Bean y Eaton (2023) señala que, el análisis de los procesos de autoevaluación del estudiante debe incluir de manera

holística aspectos como: la autoeficacia, la actitud y el afrontamiento conductual. Desde esta perspectiva, la decisión de abandonar la escuela surge de una incapacidad percibida para manejar el estrés académico y social, lo que genera una pérdida de confianza y motivación. En relación a la EGB Superior, este fenómeno se ve agravado por los cambios neurobiológicos de la adolescencia, donde el sistema límbico presenta un desarrollo más acelerado que la corteza prefrontal, responsable del control de impulsos y la visión a largo plazo (Vélez-Miranda y otros, 2020).

A continuación, se presentan los determinantes de la deserción escolar según la literatura contemporánea en América Latina (Nuñez, 2023):

Tabla 3. Deserción escolar en América Latina.

Categoría de factor	Variables específicas	Impacto en la trayectoria escolar
Endógenos	Calificaciones, asistencia, clima escolar, «metodologías docentes» (Carrillo & Gonzáles, 2023).	Afectan el sentido de competencia y «pertenencia académica».
Exógenos	Pobreza, trabajo voluntario, incerteza, estructura familiar (Zabala, 2025).	Produce barreras y estrés ambiental.
Psico biosociales	Autoconcepto, ansiedad, resiliencia, «consumo de sustancias» (Castillo & Santilla-Lima, 2023).	Determinan la capacidad de afrontamiento ante «desafíos escolares».
Contextuales	Migración, reclutamiento por bandas, brecha digital (Hernández, 2023).	Altera el entorno y la percepción de seguridad dentro del área de estudio.

Nota. Elaboración propia basada en estrategias pedagógicas implementadas por los docentes del grado noveno de la institución técnico nacional de comercio durante el aislamiento obligatorio por el Covid 19 (2020).

En consecuencia, el clima escolar y la pedagogía empática son factores determinantes de la retención escolar, donde el centro educativo debe dejar de ser una estructura rígida para convertirse en un ecosistema, que dote de significado la permanencia del estudiante. Es decir, se requiere contrastar las dos realidades institucionales: excelencia académica y protección integral del alumnado, para así generar un sentido de pertenencia genuino, como un ancla emocional (Hernández, 2023). Por lo tanto, resulta imperativo eliminar aquellas metodologías que transforman el aula en un espacio de frustración y desapego. Con el objetivo, de mejorar la comprensión del abandono escolar a través de una disección técnica de las dimensiones que interactúan de forma sistémica tanto en el aula como en la familia.

En resumen, investigaciones enfocadas en el comportamiento estudiantil dentro y fuera de las instituciones educativa, identifica a la deserción escolar no solo como la carencia económica dentro del hogar, sino como se ancla a la estabilidad emocional del estudiante y a la percepción que tiene en el ambiente institucional. Así, el debilitamiento en la conexión

psicológica intensifica la ruptura en la vida profesional del estudiante que percibe el abandono como una solución inmediata.

2.2.1. Dimensiones socioeconómica y estructural

La dimensión socioeconómica en Ecuador evidencia que, la precariedad económica en el 70% de la población, se convierte en una incidencia interdisciplinar de los hogares, factor que genera una presión sostenida hacia una inserción laboral prematura. En base al análisis realizado por la UNICEF (2024), esta vulnerabilidad se agudiza en las zonas periféricas de Ecuador, en donde el costo de oportunidad para estudiar resulta insostenible para familias de escasos recursos. Por consiguiente, la falta de liquidez para insumos tecnológicos, materiales educativos o transporte se convierte en la primera barrera que fractura el proceso educativo. Además, durante la pandemia se examinó que, 150.000 niños y adolescentes dejaron de estudiar debido a la crisis económica que azotó a sus familias, así como por la falta de acceso a dispositivos de conectividad, factor que incidió al 37% (Bolaño-García & Duarte-Acosta, 2024).

En consecuencia, este subnivel actúa como un punto crucial hacia el Bachillerato, donde se manifiesta con mayor fuerza el impacto del rezago escolar, junto con la desmotivación intrínseca. En relación con datos de la región, a partir de los 13 años se observa un incremento sostenido en la deserción, que arroja a los estudiantes hacia trabajos informales que ponen en riesgo su vida e integridad. Según la Comisión Económica Para América Latina – CEPAL (2020), a nivel estructural, la brecha de alfabetismo digital y la carencia de suministros de conectividad en zonas rurales y periurbanas de la Zona 9 actúan como una barrera técnica que anula la eficacia de cualquier modelo pedagógico híbrido. Es decir, existe una relación directa en la integración de dichas variables en un modelo de IA, para obtener un análisis sólido de pobreza y accesibilidad geográfica en la calidad educativa.

2.2.2. Dimensión académica e institucional

Frente a la dimensión académica, habitualmente centrada en el monitoreo de calificaciones y asistencia, se ha visto insuficiente como único predictor. Aunque, factores como el bajo rendimiento y el ausentismo son indicadores de riesgo, en ocasiones son signos de una desconexión más profunda con el currículo. Según Concepción et al., (2024), En virtud de un ambiente educativo percibido como excluyente y poco empático, las metodologías docentes no logran adaptarse a las realidades diversas del alumnado, lo que

acelera la desconexión emocional del adolescente. Es decir, se genera un sentimiento de indiferencia y falta de pertenencia a través de protocolos negativos de convivencia que transforman el centro educativo en un espacio de frustración, en vez de un lugar seguro para la enseñanza. Sin embargo, desde el punto de vista del desarrollo cognitivo, la juventud adquiere las capacidades de razonamiento abstracto y formal, mientras su pensamiento sigue estando fuertemente influido por el egocentrismo y la impulsividad propias de su edad (Maldonado & Soledispa, 2024).

Así, la capacidad que presenta una institución para mitigar el abandono escolar depende de la robustez de sus sistemas de gestión de datos, donde la interacción entre el rendimiento académico y el compromiso estudiantil actúa como un indicador temprano de vulnerabilidad (Ministerio de Educación del Ecuador, 2024). En este aspecto, la reestructuración en la gestión educativa exige que las instituciones educativas trasciendan el registro administrativo pasivo hacia un modelo de gobernanza de datos activa. En conclusión, la dimensión institucional debe integrar variables de clima escolar y soporte pedagógico como valores críticos en los algoritmos de detección. Por ende, las intervenciones del DECE se vuelven proactivas y fundamentadas en evidencia empírica, reduciendo así la brecha entre la detección del riesgo y la ejecución de la política de permanencia (Ministerio de Educación del Ecuador, 2024).

2.2.3. Dimensión socioemocional y salud mental

La dimensión socioemocional emerge como el factor con mayor peso predictivo en estudios recientes. Por medio de variables como el estrés, la ansiedad, la desmotivación intrínseca y la falta de apoyo en el hogar actúan como promotores silenciosos y predictores tempranos del abandono. En cambio, la estabilidad emocional del estudiante es un pilar de la permanencia. Según López et al., (2025), la enseñanza secundaria sugiere que las variables relacionadas con la motivación y la autogestión emocional guardan una asociación más fuerte con el riesgo de abandono, que las variables académicas tradicionales. Con este fin, la capacidad de identificar patrones de riesgo integrando datos subjetivos se vuelve el centro de la innovación tecnológica en la gestión educativa.

Lamentablemente, la gestión escolar se ha limitado a monitorear datos cuantitativos, que presentan pasividad cuando existe una ausencia masiva en la institución por parte del alumnado, así como cuando no aprueban sus materias. Por lo tanto, los estados emocionales negativos como la ansiedad, baja autoestima y desmotivación actúan como «predictores

líderes» que preceden al colapso académico, lo que disocia el proceso académico llevándolo al colapso. Además, desde un enfoque neuro-pedagógico, la exposición prolongada a factores estresores, inhibe las funciones ejecutivas del estudiante, impactando directamente en su capacidad de persistencia cognitiva (Barra & Pilicita, 2025). De acuerdo con esto, la integración de métricas psicométricas dentro del modelo de IA, resulta necesaria para medir el bienestar subjetivo y resiliencia, factores usados para identificar patrones de «fatiga académica».

En este contexto, la literatura contemporánea sugiere que, la salud mental no debe abordarse como un factor aislado, sino como una variable moderadora, ya que, en contextos de vulnerabilidad socioeconómica tiende a acelerar la desvinculación escolar. Según Zamora et al., (2023), si no se cuenta con un soporte institucional preventivo y tecnológicamente asistido. A continuación, se presentan las incidencias más comunes de deserción escolar a nivel país:

Tabla 4. Dimensión de riesgo en la incidencia del abandono académico

Dimensión de riesgo	Factores clave identificados	Incidencia en el abandono (%)
Socioeconómica	Inestabilidad financiera, «trabajo infantil/juvenil» (Carrillo & Gonzáles, 2023).	21,0%
Tecnológica	«Falta de conectividad» (Internet), carencia de dispositivos (Carrillo & Gonzáles, 2023).	68,0%
Personal/familiar	«Falta de apoyo familiar», embarazo adolescente (Infante & Araya, 2020).	Variable
Académica	Reprobación recurrente, «desajuste curricular» (Golec & Hachaj, 2025).	Moderada
Socioemocional	Estrés, ansiedad, «baja resiliencia» (Infante & Araya, 2020).	Alta (silenciosa)

Nota. Elaboración propia.

La «salud emocional» al representar una ausencia en estudios académicos enfocado en estudiantes, se considera un factor aislado en las instituciones educativas que, ante la vulnerabilidad socioeconómica, aceleran la desvinculación escolar. Así, al transformar la gestión educativa en la capacidad de inspeccionar el desempeño académico a un sistema preventivo que identifique patrones subjetivos de riesgo. Por último, la «persistencia académica» se apoya de la mitigación de los estímulos psicológicos que bloquean el funcionamiento cognitivo del estudiante.

2.2.4 Sistemas de alerta temprana y abandono escolar

El abandono escolar no se clasifica como un evento apartado, sino como un proceso gradual de desvinculación hacia las actividades académicas. En teoría, se manifiesta como una «renuncia definitiva a la formación académica formal» antes de obtener la certificación correspondiente a los distintos niveles educativos (Hernández, 2023). Desde una perspectiva cronológica, la literatura distingue tres estados de deserción académica: en primera instancia, cuando el estudiante deserta antes de iniciar o después de iniciar el nivel básico. En segundo lugar, ocurre durante el nivel medio de educación.

En tercer lugar, se presenta en los niveles finales del bachillerato, a menudo vinculada a la falta de proyección profesional, crisis económica y apoyo familiar. A continuación, se detallan los factores críticos que deben considerarse al desarrollar un sistema de alerta temprana basado en IA para el Bachillerato:

- Preprocesamiento de datos. Es esencial el manejo de datos por medio de técnicas como SMOTE para mejorar la recuperación de los estudiantes que realmente van a desertar (Bolaño-García & Duarte-Acosta, 2024).
- Integración de datos multidimensionales. Los modelos más robustos combinan datos académicos, administrativos y personales (Hernández, 2023).
- Temporalidad de la intervención. Para la detección de intervenciones oportunas es vital un análisis de anticipación de 2 o 3 semestres (Golec & Hachaj, 2025).

Mientras tanto, el Ministerio de Educación determina una ruta de alerta temprana que debe ser alimentada por el sistema de la IA la cual categoriza el riesgo basándose en el desempeño y la asistencia. La siguiente tabla muestra se basa en la integración de los niveles de alerta de la Guía informativa de Riesgos Psicosociales (2023) con las capacidades de detección de la IA.

Tabla 5. Niveles de alerta.

Nivel de alerta (MINEDUC)	Indicadores tradicionales	Detección predictiva con IA (propuesto)	Acción institucional
Normal	Asistencia > 90%, Promedio > 7/10.	Sentimiento positivo en bitácoras; alta «motivación intrínseca».	Monitoreo estándar (Zabala, 2025).

Nivel de alerta (MINEDUC)	Indicadores tradicionales	Detección predictiva con IA (propuesto)	Acción institucional
Amarilla	Faltas esporádicas, notas en proceso (EP).	Detección de ansiedad y «micro-caídas» en el compromiso.	Entrevista preventiva y «refuerzo pedagógico» (Pachay-López & Rodríguez-Gámez, 2021).
Naranja	Ausentismo recurrente, notas iniciadas (I).	Desmotivación externa detectada por «NLP»; estrés familiar.	Plan de acompañamiento integral «DECE» (Infante & Araya, 2020).
Roja	Abandono de hecho, inasistencia masiva.	Probabilidad de «deserción» > 80% basada en datos «multimodales».	Protocolo de protección de derechos y reinserción (Barra & Pilicita, 2025).

Nota. Elaboración propia.

En conclusión, en esta etapa el modelo de la IA permite identificar una serie de patrones, tales como la crisis de vinculación institucional y la inestabilidad emocional subyacente. Que, en este contexto estudiantil, se estima la probabilidad de que el abandono sea definitivo.

2.3. Bases teóricas conceptuales de la investigación

La construcción de un modelo predictivo para la retención escolar exige un marco conceptual que integre la teoría educativa con los paradigmas emergentes de la ciencia de datos. A continuación, se detallan los ejes teóricos que fundamentan la presente investigación:

2.3.1. Teoría Ecológica y Multidimensionalidad del Riesgo

La investigación se fundamenta en la Teoría Ecológica de los Sistemas, la cual sugiere que el éxito al lado del fracaso del estudiante es el resultado de la interacción entre múltiples entornos ecológicos (Carrillo & González, 2023). Bajo este enfoque, los factores de riesgo se agrupan en cinco categorías críticas:

- Demográficos. Edad, género y composición familiar.
- Socioeconómicos. Nivel de ingresos, becas, estabilidad laboral del representante y acceso a conectividad.
- Institucionales. Calidad de la infraestructura, clima escolar, servicios de orientación y relación docente-estudiante.

- Académicos. Rendimiento (notas), tasa de aprobación de asignaturas, asistencia y antecedentes de reprobación.
- Personales/Socioemocionales. Autoconcepto, resiliencia, salud mental, embarazo adolescente y motivación intrínseca.

Por consiguiente, al incorporar la IA “Orange Data Mining” en este marco ecológico, por primera vez, se cuantifica la interdependencia de estos sistemas mediante el procesamiento de grandes volúmenes de datos. Al contrario de los modelos estadísticos tradicionales, el uso de redes neuronales y algoritmos de ensamble permite identificar el descenso en la motivación detectada por NLP (Hernández, 2023). Según King et al., (2024) esta capacidad de análisis multivariado consigue recodificar la Teoría Ecológica de un modelo descriptivo a una herramienta predictiva de alta precisión, facilitando intervenciones que no solo atienden al estudiante, sino que consideran la complejidad de su entorno sistémico.

2.3.2. Constructos socioemocionales y resiliencia educativa

La Inteligencia Emocional (IE) y la autorregulación son el cimiento de nuevos conceptos para la presente investigación. De esta manera, al combinar las habilidades psicológicas con las competencias transversales, se reconoce la capacidad de identificar, comprender y gestionar las emociones propias y ajenas dentro de un entorno determinado (López y otros, 2025). En el contexto del bachillerato, un debilitamiento en la capacidad de adaptación ante situaciones adversas actúa como un predictor de «desenganche emocional». Por ese motivo, los sistemas de IA actuales buscan modelar estos constructos mediante el análisis de la Inteligencia Emocional Artificial, el cual evalúa el lenguaje que refleja el estado interno del alumno (Soriano-Sánchez & Jiménez-Vázquez, 2023).

A gran escala, el desafío de incluir la variable socioemocional radica en que, gran parte de esta información reside en datos no estructurados sobre sus logros académicos en lenguaje natural o entrevistas cualitativas. Para abordar esto, se propone una Arquitectura Híbrida Multimodal que combina (Barra & Pilicita, 2025):

- Módulo predictivo de datos estructurados (XGBoost). Analiza de manera cuantitativa la trayectoria académica y socioeconómica (Parrales & Cedeño, 2025).

- Módulo de Procesamiento de Lenguaje Natural (NLP). Consiste en modelos como BERT o LLMs que permiten decodificar la carga emocional de los textos (Infante & Araya, 2020).

En resumen, la combinación de estos datos se realiza en una capa de procesamiento que pondera las señales cualitativas de estrés o desmotivación, junto con el rendimiento académico, puesto que genera perfiles de riesgo holísticos significativos para reducir falsos negativos.

2.3.3. Inteligencia Artificial y Aprendizaje Basado en Datos (ABD)

El marco técnico se basa en el Aprendizaje Basado en Datos (ABD), una estrategia que utiliza la analítica para mejorar la intervención pedagógica y la toma de decisiones gerenciales (Bolaño-García & Duarte-Acosta, 2024). Dentro de este paradigma, se destacan tres conceptos fundamentales:

- Machine Learning (Aprendizaje Automático). Algoritmos que identifican patrones en datos históricos para clasificar nuevos casos (ej. desertor vs. no desertor) (Barra & Pilicita, 2025).
- Procesamiento de Lenguaje Natural (NLP). Rama de la IA que permite a las máquinas interpretar y procesar el lenguaje humano, facilitando el análisis de «datos no estructurados», tales como comentarios de docentes o diarios de estudiantes (Carrillo & Gonzáles, 2023).
- Transformers y LLMs (Large Language Models). Arquitecturas como Orange Data Mining utilizan mecanismos de «atención» para asignar pesos dinámicos a diferentes partes de una secuencia de texto, permitiendo una comprensión contextual profunda de los sentimientos y riesgos del alumno.

Recapitulando, al introducir un modelo predictivo robusto, se coloca en la Minería de Datos Educativos (EDM), un campo interdisciplinario que utiliza técnicas de aprendizaje automático para extraer patrones de comportamiento en entornos escolares (Nuñez, 2023). Por el contrario, los modelos estadísticos tradicionales quedan inactivos por su veracidad ante problemas educativos, lo que permite a la IA procesar relaciones no lineales entre variables y manejar grandes volúmenes de datos con una precisión superior al 85%. Por ello, el algoritmo

de elección para datos tabulares estructurados en el rendimiento académico es el XGBoost (Extreme Gradient Boosting). Dado que, permite utilizar una serie de datos a través de árboles de decisión que aprenden de manera secuencial, donde cada nuevo árbol intenta corregir los errores de los anteriores, optimizando la precisión mediante el gradiente descendente (Barra & Pilicita, 2025).

2.3.4. Teorías del aprendizaje en la era digital

A través de los cambios en los modelos de aprendizaje, la aparición de los modelos de lenguaje de gran escala (LLM), específicamente la familia Gemini (Google), ha transformado la capacidad de procesar las dimensiones subjetivas del estudiante (Golec & Hachaj, 2025). Dado que, Orange Data Mining ofrece ventanas de contexto masivas de hasta 1 millón de *tokens* que permiten analizar el historial completo de un estudiante. Esta herramienta marca un paso en la enseñanza por la elaboración de nuevas metodologías estratégicas dirigidas a los docentes (Quiroz, 2025). Una de las habilidades técnicas que lidera este modelo, es la generación de «*Structured Outputs*» que permite al modelo predictivo enviar reportes no estructurados a la API de Orange Data Mining y recibir a cambio un «objeto JSON» validado que contiene los indicadores socioemocionales mapeados a las dimensiones del currículo ecuatoriano (Zabala, 2025).

Para Vygotsky (2025), la IA actúa como un mediador para proporcionar una estrategia pedagógica en el apoyo integral a un estudiante, con el objetivo de cumplir los objetivos académicos en un marco de protección de derechos. Por esta razón, la Inteligencia Artificial permite la retroalimentación personalizada en tiempo real, debido al ajuste de la complejidad del contenido según el récord académico del estudiante (Nuñez, 2023). En vista de ello, el conectivismo por Siemens señala que, el aprendizaje se reestructura como un proceso de conexión de nodos de información especializados. Por consiguiente, las plataformas de IA, potenciados por modelos de lenguajes de gran escala (LLM_s) no se enfoca solo en la construcción de conocimientos en entornos interconectados, sino que, actúan como pilar de enlace entre los valores socioemocionales, académicos y estructurales. En conclusión, al implementar la IA en el sistema educativo para el seguimiento del bienestar estudiantil, se mitiga el riesgo de abandono escolar, al fortalecer la relación entre el soporte pedagógico y la realidad socioeconómica (Quiroz, 2025).

2.3.5 IA Explicable (XAI) y ética algorítmica

En primera instancia, el concepto transversal en la IA Explicable (XAI), se define como técnicas (SHAP), una herramienta que se enfoca en la interpretación de resultados de modelos complejos de *machine learning* denominado «caja negra» (Soriano-Sánchez & Jiménez-Vázquez, 2023). Estos resultados permiten a los humanos comprender de una manera clara y cuantificable cada característica que influye en el programa, lo que lo posiciona como un instrumento primordial en áreas como la medicina, la educación y las finanzas. De ahí que, es esencial el DECE para que puedan justificar sus intervenciones basándose en los factores de riesgo específicos identificados por el algoritmo. Así, la investigación reconoce la necesidad de mitigar el sesgo algorítmico para garantizar la privacidad de los datos bajo marcos de adopción responsable.

Es por este que, los beneficios técnicos al implementar este algoritmo en el ámbito educativo incluyen:

- Escalabilidad y velocidad. Analiza las bases de datos de miles de estudiantes con bajo consumo de memoria.
- Manejo de valores faltantes. Corresponde a la cifra de escuelas donde los registros de asistencia y/o reportes psicológicos son incompletos.
- Importancia de las características. Permite identificar aquellos factores con mayor peso en la predicción para un estudiante específico.

Tabla 6. Marco conceptual de IA y psicopedagogía educativa.

Término Conceptual	Definición Técnica	Relevancia en el Modelo
DPS (Dropout Probability Score)	Puntaje calculado mediante variables académicas y «conductuales» (Nuñez, 2023).	Clasificación del «nivel de riesgo».
RAG (Retrieval-Augmented Generation)	Técnica que combina generación de texto con búsqueda en bases de «conocimiento» (Zabala, 2025).	Interpretación contextual de comentarios con «relevancia pedagógica».
Sentimiento Académico	Polaridad emocional detectada en la narrativa del «estudiante» (Infante & Araya, 2020).	Identificación de «aislamiento» y «ansiedad por carga de trabajo».
Analítica de Aprendizaje	Medición y «análisis» de datos sobre los estudiantes y sus contextos (Zabala, 2025).	Optimización del «entorno educativo» y «prevención de la deserción».

Nota. Elaboración propia basada en desafíos y retos de la inteligencia artificial en la educación ecuatoriana: Una mirada desde la enseñanza y el rol del docente (Nuñez, 2023).

Al implementar herramientas de IA Explicable en la educación, así como optimiza la eficiencia administrativa, también otorga precisión al medir la ética y solidez en decisiones pedagógicas. Las variables complejas al modificarlas en indicadores comprensibles (el DPS y el «sentimiento académico»), permite un traspaso de un acompañamiento académico a una prevención de la deserción. Con ello, asegura que la tecnología actúe como un enfoque de guía para mejorar el rendimiento académico y no como un sesgo automatizado en la trayectoria del estudiante.

2.4. Bases legales que fundamentan el estudio

El desarrollo de la presente investigación se fundamenta en su ejecución, la cual integra modelos predictivos de Inteligencia Artificial en el sistema educativo ecuatoriano. No obstante, es importante sustentar según los lineamientos jurídicos jerarquizados que garantizan el derecho a la educación, la protección integral de niños y adolescentes, así como la seguridad en el tratamiento de datos sensibles. A continuación, se detalla la normativa vigente:

- **Constitución de la República del Ecuador**

De acuerdo con el marco constitucional del Estado que garantiza derechos fundamentales a través de mecanismos preventivos y regulatorios (Asamblea Nacional Constituyente, 2008):

Art. 26.- La educación es un derecho de las personas a lo largo de su vida y un deber ineludible e inexcusable del Estado. Constituye un área prioritaria de la política pública y de la inversión estatal, garantía de la igualdad e inclusión social y condición indispensable para el buen vivir. Las personas, las familias y la sociedad tienen el derecho y la responsabilidad de participar en el proceso educativo.

Art. 28.- La educación responderá al interés público y no estará al servicio de intereses individuales y corporativos. Se garantizará el acceso universal, permanencia, movilidad y egreso sin discriminación alguna y la obligatoriedad en el nivel inicial, básico y bachillerato o su equivalente.

Es derecho de toda persona y comunidad interactuar entre culturas y participar en una sociedad que aprende. El Estado promoverá el diálogo intercultural en sus múltiples dimensiones (Asamblea Nacional Constituyente, 2008).

El aprendizaje se desarrollará de forma escolarizada y no escolarizada.

Art. 44.- El Estado, la sociedad y la familia promoverán de forma prioritaria el desarrollo integral de las niñas, niños y adolescentes, y asegurará el ejercicio pleno de sus derechos; se atenderá al principio de su interés superior y sus derechos prevalecerán sobre los de las demás personas.

Las niñas, niños y adolescentes tendrán derecho a su desarrollo integral, entendido como proceso de crecimiento, maduración y despliegue de su intelecto y de sus capacidades, potencialidades y aspiraciones, en un entorno familiar, escolar, social y comunitario de afectividad y seguridad. Este entorno permitirá la satisfacción de sus necesidades sociales, afectivo-emocionales y culturales, con el apoyo de políticas intersectoriales nacionales y locales.

Art. 66, Num. 19: El derecho a la protección de datos de carácter personal, que incluye el acceso y la decisión sobre información y datos de este carácter, así como su correspondiente protección. La recolección, archivo, procesamiento, distribución o difusión de estos datos o información requerirán la autorización del titular o el mandato de la ley.

Art. 347, Núm. 5 y 6: Será responsabilidad del Estado (Asamblea Nacional Constituyente, 2008):

5. Garantizar el respeto del desarrollo psicoevolutivo de los niños, niñas y adolescentes, en todo el proceso educativo.
6. Erradicar todas las formas de violencia en el sistema educativo y velar por la integridad física, psicológica y sexual de las estudiantes y los estudiantes.

- **Ley Orgánica de Educación Intercultural y su Reglamento**

La normativa sectorial regula la gestión de la permanencia y el rol de las alertas tempranas (Asamblea Nacional del Ecuador, 2015):

Art. 2. - Principios. - La actividad educativa se desarrolla atendiendo a los siguientes principios generales, que son los fundamentos filosóficos, conceptuales y constitucionales que sustentan, definen y rigen las decisiones y actividades en el ámbito educativo:

- a. Universalidad.** - La educación es un derecho humano fundamental y es deber ineludible e inexcusable del Estado garantizar el acceso, permanencia y calidad de la educación para toda la población sin ningún tipo de discriminación. Está articulada a los instrumentos internacionales de derechos humanos;
- h. Acceso y permanencia.** - Se garantiza el derecho a la educación en cualquier etapa o ciclo de la vida de las personas, así como su acceso, permanencia, movilidad y egreso sin discriminación alguna;

Art. 6, Literal m.- Propiciar la investigación científica, tecnológica y la innovación, la creación artística, la práctica del deporte, la protección y conservación del patrimonio cultural, natural y del medio ambiente, y la diversidad cultural y lingüística;

CAPÍTULO III

DE LAS MATRÍCULAS Y EL INGRESO AL SISTEMA DE EDUCACIÓN

Art. 162.- De la matrícula posterior al inicio del año lectivo. - La Autoridad Educativa Nacional normará el procedimiento y las evaluaciones pedagógicas que sean necesarias previas a la matrícula de un estudiante en un establecimiento educativo cuando ésta se efectúe una vez iniciado el año lectivo.

CAPÍTULO VI

DE LA ASISTENCIA DE LOS ESTUDIANTES

Art. 171.- Inasistencia recurrente. Cuando la inasistencia de un estudiante fuere recurrente y estuviere debidamente justificada, la máxima autoridad de la institución educativa solicitará la aplicación de las medidas previstas en la normativa expedida por el Nivel Central de la Autoridad Educativa Nacional, a fin de garantizar la continuidad de los estudios, el apoyo psicopedagógico y las tutorías académicas correspondientes según el caso.

CAPÍTULO VII

DE LA OFERTA DE FORMACIÓN PERMANENTE PARA LOS PROFESIONALES DE LA

EDUCACIÓN

Art. 313.- Tipos de formación permanente. La oferta de formación en ejercicio para los profesionales de la educación es complementaria o remedial. La formación permanente de carácter complementario se refiere a los procesos de desarrollo profesional, capacitación, actualización, formación continua, mejoramiento pedagógico y académico para que provean a los docentes de conocimientos y habilidades distintas de las aprendidas en su formación inicial.

- **Código de la Niñez y Adolescencia (CNA)**

Art. 1: Dispone la protección integral que el Estado y la familia deben garantizar para el disfrute pleno de sus derechos en un marco de dignidad y equidad (Asamblea Nacional del Ecuador, 2014).

Art. 37: Ratifica el derecho a una educación de calidad y la obligación del sistema educativo de garantizar la permanencia del adolescente hasta el bachillerato (Asamblea Nacional del Ecuador, 2014).

- **Ley Orgánica de Protección de Datos Personales (LOPDP)**

Debido a que la investigación utiliza IA para manejar información socioemocional, se rige bajo la LOPDP (Asamblea Nacional del Ecuador, 2021):

Art. 1.- Objeto y finalidad. - El objeto y finalidad de la presente ley es garantizar el ejercicio del derecho a la protección de datos personales, que incluye el acceso y decisión sobre información y datos de este carácter, así como su correspondiente protección, Para dicho efecto regula, prevé y desarrolla principios, derechos, obligaciones y mecanismos de tutela.

CAPÍTULO IV

CATEGORÍAS ESPECIALES DE DATOS

Art. 25.-Categorías especiales de datos personales. - Se considerarán categorías especiales de datos

personales, los siguientes:

- a. Datos sensibles;
- b. Datos de niñas, niños y adolescentes;
- c. Datos de salud; y,
- d. Datos de personas con discapacidad y de sus sustitutos, relativos a la discapacidad.

Art. 26.-Tratamiento de datos sensibles. - Queda prohibido el tratamiento de datos personales sensibles salvo que concurra alguna de las siguientes circunstancias:

- a. El titular haya dado su consentimiento explícito para el tratamiento de sus datos personales, especificando claramente sus fines.
- b. El tratamiento es necesario para el cumplimiento de obligaciones y el ejercicio de derechos específicos del responsable del tratamiento o del titular en el ámbito del Derecho laboral y de la seguridad y protección social.
- c. El tratamiento es necesario para proteger intereses vitales del titular o de otra persona natural, en el supuesto de que el titular no esté capacitado, física o jurídicamente, para dar su consentimiento.
- d. El tratamiento se refiere a datos personales que el titular ha hecho manifiestamente públicos.
- e. El tratamiento se lo realiza por orden de autoridad judicial.
- f. El tratamiento es necesario con fines de archivo en interés público, fines de investigación científica o histórica o fines estadísticos, que debe ser proporcional al objetivo perseguido, respetar en lo esencial el derecho a la protección de datos y establecer medidas adecuadas y específicas para proteger los intereses y derechos fundamentales del titular.
- g. Cuando el tratamiento de los datos de salud se sujete a las disposiciones contenidas en la presente ley.

- **Políticas Ministeriales: Plan Nacional "Cuidemos de ti"**

Expedido mediante el Acuerdo Ministerial MINEDUC-MINEDUC-2024-00064-A, este plan constituye la base operativa directa de la investigación (Ministerio de Educación del Ecuador, 2024):

Objetivo General: Prevenir los casos de abandono escolar, reducir la repitencia e incrementar la reinserción mediante la articulación de programas y proyectos institucionales.

Estrategias de Retención: Establece el fortalecimiento de capacidades para implementar estrategias preventivas contra riesgos psicosociales, tales como la violencia, el trabajo infantil y la inseguridad.

Investigación basada en evidencia: El plan promueve la indagación sistemática sobre el abandono escolar para la toma de decisiones de política pública fundamentadas en datos reales.

CAPÍTULO III

MARCO METODOLÓGICO

3.1. Paradigma, Enfoque, alcance, modalidad, tipo de estudio y diseño de investigación

En la actualidad el «abandono escolar» a nivel bachillerato constituye uno de los nudos críticos más urgentes para las políticas públicas en el Ecuador, donde se juntan variables socioemocionales con el entorno académico (Zabala, 2025). En este sentido, la gestión de «permanencia escolar» se ha gestionado bajo conceptos descriptivos y reactivos que limita la intervención institucional ante riesgos de «abandono». En cambio, el aumento de herramientas en relación a la Analítica de Aprendizaje (Learning Analytics) y sistemas de Inteligencia Artificial (IA) brinda una oportunidad en la administración moderna (UNESCO, 2023). Así, el modelo analiza el entorno a evaluar para anticipar un riesgo en base a métricas y registros del DECE para proteger la formación académica del alumnado.

En base a la epistemología, el estudio se apoya en el paradigma pragmático que alinea el «criterio de utilidad» como eje central en su metodología. Bajo el concepto de Quituzaca et al., (2025), el conocimiento científico no se busca solo por «amor a la verdad», sino por la capacidad de resolver problemas en tiempo real. El enfoque descrito materializa un esquema moderno en una estructura multimodal al integrar la precisión algorítmica del Machine Learning con el Procesamiento de Lenguaje Natural (NLP) (Carrillo & Gonzáles, 2023). De esta forma, al usar la plataforma Orange Data Mining, la investigación pasa de una *caja negra* informática a un sistema de procesamiento de datos estructurado por medio de las narrativas estudiantiles.

Al fusionar este diagnóstico humano y accionable, la arquitectura del modelo predictivo pasa a un diseño mixto explicativo secuencial (DEXPL) que garantiza una triangulación de datos (Mota, 2022). Dado que, la técnica equilibra valores críticos con la percepción del bienestar del alumno, el algoritmo XGBoost procesa variables cuantitativas que categoriza los niveles de riesgo (Tomaszewski, 2021). Por lo tanto, la técnica de balanceo SMOTE con las métricas de fidelidad (F1-Score) incrementa la precisión operativa en base a la evidencia empírica. Por el contrario, el rendimiento académico no se comprende solo de un buen rendimiento académico, ya que, alberga diversos factores en la formación del estudiante. En base a lo mencionado, el sistema se alimenta de una base de

datos cualitativos invisibles que facilita una comprensión profunda ante el «ecosistema formativo».

A su vez, los reportes del Departamento de Consejería estudiantil (DECE) se codifican bajo la estructura interpretativa CASEL, lo cual facilita que la subjetividad explique cifras cuantitativas (Dewey & Faerna, 2022). Así, el enfoque de innovación tecnológica potencia el sistema administrativo por medio de un Dropout Probability Score (DPS) que se ajusta con los estándares nacionales. Según Liu et al., (2025), al integrar valores SHAP, el programa de IA transforma predicciones matemáticas en rutas y protocolos hacia el docente para conseguir una intervención clara y ética. Por ese motivo, al usar un modelo de IA en la Unidad Educativa la investigación se estructura bajo las normativas de protección de datos y privacidad de menores. Al terminar el desarrollo de esta herramienta, los docentes responderán con técnicas pedagógicas ante el abandono escolar y así tener un mecanismo de «justicia educativa».

En estos instrumentos la validación de resultados se fundamenta en una división de datos que permite al algoritmo aprender y clasificar los niveles de riesgo de «abandono» (Delgado & Riquelme, 2022). Para comenzar, se realizarán pruebas «train/test Split», que reserva el 80% de la muestra para el aprendizaje del algoritmo y el 20% para la evaluación de su eficacia en datos no vistos. Así mismo, al aplicar una validación cruzada, asegura la estabilidad del modelo y minimiza el sesgo en las predicciones. Esta metodología sirve para medir la calidad del sistema a través de métricas de evaluación estandarizadas, las cuales se conforman por la precisión «*accuracy*», la sensibilidad «*recall*» y el puntaje F1 «*F1-score*» (Hernández, 2023). Por el contrario, la precisión «*accuracy*» determina la eficacia general de las predicciones, mientras que la sensibilidad «*recall*» es primordial para asegurar la mayor cantidad posible de estudiantes en riesgo (Jin, 2023).

Tabla 7. Métricas de calidad del sistema

Métrica de evaluación	Definición operativa	Importancia en el abandono escolar
Precisión « <i>accuracy</i> »	Eficacia general de las «predicciones».	Determina la «confiabilidad global» del sistema (Matbouli & Alghamdi, 2022).
Sensibilidad « <i>recall</i> »	Identificación de «alumnos en riesgo».	Evitar «falsos negativos» para abarcar a estudiantes que necesitan apoyo (Coghlan &

		D'Alfonso, 2021).
Puntaje F1 «F1-score»	Equilibrio entre «precisión» y «sensibilidad».	Valida la solidez del modelo en su «rendimiento» (Tomaszewski, 2021).

Nota. Elaboración propia basada en Machine learning for the educational sciences (2021).

En otras palabras, estas dos métricas evitan que el sistema genere «falsos negativos» e ignore a estudiantes en riesgo que necesitan apoyo oportuno. Por último, el puntaje F1 «*F1-score*» aporta un equilibrio robusto entre ambas métricas dado que, valida el modelo en sus clasificaciones (Tomaszewski, 2021). Según Concepción et al., (2024), dichas métricas permiten cuantificar con exactitud la capacidad del modelo para identificar a los estudiantes en situación de riesgo. Una vez que el sistema capte los campos integrados en su sistema operativo, el modelo conseguirá de manera instantánea, rápida y consistente categorizar posibles riesgos. De esta manera, el modelo se vuelve confiable y equilibrado para la toma de decisiones administrativas y pedagógicas preventivas.

3.1.1. Paradigma

El estudio se enmarca en el paradigma pragmático, el cual concibe el conocimiento como válido para resolver problemas prácticos del mundo real. Bajo este concepto, el paradigma no busca pensamientos dogmáticos, sino que, se posiciona en lo que William James denominaba el «valor de flujo» de las ideas (James, 2023). Dicha teoría es válida, si su aplicación produce resultados prácticos que más allá de buscar una validez estadística, mejore la vida del estudiante en el transcurso del año lectivo. Tal orientación filosófica convierte a la propuesta en un sistema práctico de trabajo según el criterio de utilidad (Dewey & Faerna, 2022). Debido a que, establece la evolución del cerebro humano como una fuente de adaptabilidad en su entorno para sobrevivir y resolver problemas a futuro.

De esta manera, la investigación analiza la «verdad» del modelo de IA en su perfección matemática intrínseca y en su capacidad real para mitigar el abandono escolar mediante medidas anticipadas ante posibles riesgos (Delgado & Riquelme, 2022). Por ello, el estudio trasciende la observación de la realidad al transformar el conocimiento y validar una construcción mediada por la praxis a través de la acción técnica y los instrumentos de análisis. Al aplicar este paradigma, la investigación entiende que la realidad educativa sigue un proceso de cambio, por lo cual el modelo debe tener la capacidad de adaptarse a las dinámicas socioemocionales de los adolescentes en un entorno de crisis. Así, este enfoque ofrece al investigador la libertad de adaptar metodologías que prioricen la

capacidad predictiva con la precisión diagnóstica del sistema a utilizar (Jin, 2023).

- XGBoost como Instrumento de Precisión: Materializa el pragmatismo al buscar la «eficiencia» y la «minimización del error predictivo». Su capacidad de ensamble, donde varía «información débil», se corrige por secuencias para generar un «aprendiz fuerte» capaz de ofrecer predicciones robustas sobre el rendimiento académico (Coghlan & D'Alfonso, 2021).
- Instrumentos de Interpretación: El Procesamiento de Lenguaje Natural (NLP) mediante los widgets de Orange opera el análisis pragmático, que busca entender las narrativas del estudiante. La utilidad aquí es lingüística y social, ya que, transforma la subjetividad emocional en datos codificados que el educador puede utilizar para tomar decisiones inmediatas (Zhou & Ye, 2023).

La educación en la pedagogía y en la gestión mediada por la tecnología asegura que el alumno no abandone y siga aprendiendo, este indicador constituye la operacionalización del pragmatismo en el ámbito educativo y tecnológico.

Por lo tanto, el estudio trasciende la observación de la realidad para transformar el conocimiento y validar una construcción mediada por la praxis a través de la acción técnica y los instrumentos de análisis. Tal enfoque ofrece al investigador la libertad de adaptar métodos que prioricen la capacidad predictiva con la precisión diagnóstica del sistema a utilizar (Jin, 2023). Al aplicar este paradigma, la investigación entiende que la realidad educativa sigue un proceso de cambio, por lo cual el modelo debe tener la capacidad de adaptarse a las dinámicas socioemocionales de los adolescentes en un entorno de crisis.

3.1.2. Enfoque

La investigación adopta un enfoque mixto, el cual permite una integración de componentes cuantitativos y cualitativos para abordar la complejidad del «abandono escolar». Así, el enfoque cuantitativo se manifiesta en el uso de técnicas analíticas predictivas y de un aprendizaje automático para identificar patrones de riesgo asociados a este fenómeno (Delgado & Riquelme, 2022). A partir del procesamiento de datos académicos y de la interacción estudiantil, el enfoque cualitativo se orienta al análisis de narrativas y expresiones socioemocionales de los estudiantes. Por medio de técnicas de procesamiento de lenguaje natural (NLP), permite interpretar dimensiones subjetivas que

no pueden ser capturadas exclusivamente por indicadores numéricos (Golec & Hachaj, 2025).

En este contexto, el enfoque no limita una serie de métodos que busca la triangulación de datos, es decir se ejecuta a través de los resultados obtenidos de indicadores estadísticos de rendimiento académico con el análisis interpretativo de las narrativas socioemocionales (Coghlan & D'Alfonso, 2021). Al usar esta metodología, permite el cruce de información al contrastar datos numéricos detectados por el aprendizaje supervisado con la subjetividad obtenida del «procesamiento de lenguaje natural» (NLP) (Golec & Hachaj, 2025). Tras evaluar las métricas de rendimiento académico con la realidad emocional del estudiante, el sistema se adapta a las necesidades académicas que el DECE necesita para abordar nuevas metodologías de prevención. Por ello, el modelo predictivo con el proyecto de la Inteligencia Artificial mejora su visión holística al enriquecer su algoritmo por parte del sujeto como de su comportamiento estadístico.

El objetivo se centra en la transformación de cifras numéricas de «*dataset*» sobre las variables del desempeño académico y de la variable socioemocional para mejorar el rendimiento algorítmico bajo la guía de un docente (James, 2023). En esta etapa el investigador identifica correlaciones estadísticas significativas con indicadores de riesgo que validan la precisión predictiva del diseño. De esta manera, la dimensión cualitativa facilita la interpretación de datos necesaria para entender el ¿por qué? de los hallazgos obtenidos. Al identificar la variable socioemocional y la percepción de bienestar de los alumnos que cursan el Bachillerato, se representa como factores críticos que el sistema no logra detectar antes de que surja una crisis académica (Mota, 2022).

Esta técnica enriquece el «Procesamiento de Lenguaje Natural» (NLP) de datos cualitativos que valida el uso de la IA como una herramienta fiable ante problemas de «abandono» (Tomaszewski, 2021). En este estudio, la dimensión se operativiza a partir del análisis de contenido cualitativo de los reportes del «Departamento de Consejería Estudiantil» (DECE). En donde cuyo proceso, se aplicará una codificación híbrida, ya que, combina la rigurosidad de la teoría con la riqueza de la realidad observada. En primer lugar, una codificación deductiva basada en categorías predefinidas como autorregulación, entorno familiar y desmotivación (Carrillo & Gonzáles, 2023). En segundo lugar, una codificación inductiva que permite identificar niveles críticos propios del contexto de los

alumnados.

Según García et al., (2023), al utilizar el «Procesamiento de Lenguaje Natural» (NLP) como guía para la extracción y clasificación de datos, el marco interpretativo se basa en la validación manual del investigador. Así, quien examine la capacidad del algoritmo para entender el contexto de las etiquetas generadas por la IA, medirá la probabilidad de «abandono», que permita interpretar los factores críticos ajenos al modelo tradicional (Bolaño-García & Duarte-Acosta, 2024). Dicho de otro modo, al combinar estas metodologías ofrecen una comprensión holística del fenómeno educativo, entre la analítica predictiva con la riqueza descriptiva del análisis socioemocional. Por ese motivo, supera la estructura de modelos basados en datos numéricos, al aportar una capa de interpretabilidad semántica al transformar la predicción en un conocimiento humano frente al riesgo académico.

El primer aspecto de análisis se sustenta en las competencias del modelo CASEL, el cual actúa como eje de evaluación en el estado socioemocional del estudiante. Desde el punto de vista de Frye et al., (2024), este carácter deductivo permite identificar niveles de metacognición en reportes del DECE sobre la capacidad del adolescente para gestionar el estrés académico. Por esta razón, se analiza la conciencia social para detectar patrones de aislamiento, tensiones intergrupales y la falta de equipos de apoyo en instituciones educativas. Dicha categorización teórica facilita al sistema transformar los textos para clasificar comportamientos alineados a la psicología educativa internacional (Coghlan & D'Alfonso, 2021).

En consecuencia, el estudio analiza el problema desde una premisa general que nace del contexto del bachillerato ecuatoriano ante el fenómeno de manera realista. El enfoque de estas dimensiones es la vulnerabilidad socioeconómica y el entorno familiar, en el cual capturan datos sobre la inserción laboral temprana, la insuficiencia de apoyo familiar y de la tutela familiar (Ventura & Peña, 2021). Cabe subrayar en esta fase la identificación de indicadores de seguridad y de entornos macrosociales, que sirven para tratar conceptos sobre el impacto en la inseguridad del barrio y la estabilidad laboral. Al alimentar el algoritmo con categorías inductivas, captura la complejidad del entorno y aumenta la precisión técnica ante problemas reales

Tabla 8. Estrategia metodológica: enfoque mixto e IA predictiva

Componente	Enfoque cuantitativo	Enfoque cualitativo
------------	----------------------	---------------------

Herramienta clave	Algoritmos supervisados y modelos predictivos mediante el «Aprendizaje Automático» (Machine Learning)	Procesamiento de Lenguaje Natural (NLP): A través de «Text Mining de Orange» para el análisis de reportes (Jin, 2023).
Fuente de datos	Valores numéricos, «dataset», desempeño académico.	Informes «socioemocionales», reportes del DECE y opiniones del alumnado (Tomaszewski, 2021).
Objetivo específico	Categorizar «factores de riesgo» y establecer relaciones estadísticas.	Conocer el «¿por qué?» de los hallazgos para capturar las dimensiones que no se detectan (Zhou & Ye, 2023).
Proceso de análisis	Codificación de variables de desempeño en criterios de validación.	Codificación Híbrida: «deductiva» e «inductiva».

Nota. Elaboración propia MOOC student dropout prediction model based on learning behavior features and parameter optimization (2023).

Al combinar el enfoque cualitativo y cuantitativo con herramientas de inteligencia artificial para identificar valores estadísticos de riesgo, las dimensiones socioemocionales explican los factores que tiene un estudiante en su preparación estudiantil. Por ende, el sistema se fortalece al integrar esta información en su sistema y garantizar una guía de apoyo a los docentes.

En resumen, al determinar herramientas tecnológicas para la traducción de datos cualitativos en información cuantitativa, el uso de «Text Mining de Orange» ofrece beneficios en el manejo de información no estructurada (Zhou & Ye, 2023). Debido al enfoque de esta dimensión en la polaridad emocional y la intensidad de la alerta, el algoritmo analiza la actitud que el autor proyecta en los reportes. Al señalar el ámbito de la psicología educativa, el algoritmo asigna valores numéricos al «malestar estudiantil» que alimentan el «dataset» del modelo predictivo. Según King et al., (2024), este proceso garantiza que los relatos por parte del estudiante junto con la interpretación del profesional del DECE, se codifiquen en unidades de análisis cualitativas. A fin de que, el sistema final clasifique el riesgo de abandono de una forma autónoma y respaldados por la experiencia humana.

3.1.3. Alcance

El alcance del estudio se desarrolla en procesos complementarios para asegurar una visión integral que aborda el fenómeno del abandono escolar desde una perspectiva multidimensional. En la primera fase, el estudio describe las características académicas, conductuales y socioemocionales de los estudiantes de bachillerato y busca patrones con tendencias relevantes. Tal fase es esencial para el proyecto de IA, pues ofrece una «línea

base» de datos (Zabala, 2025). Así mismo, el proceso de identificar variables como la ansiedad, el rendimiento académico y la disciplina estructura las entradas «*features*» principales para el entrenamiento del algoritmo.

Por consiguiente, al usar una técnica explicativa busca trascender la observación en las relaciones de causalidad e influencia entre las variables independientes y el abandono escolar. Por esta razón, la técnica señala que los estudiantes están en riesgo y que factores socioemocionales actúan como detonantes del abandono. En este nivel de análisis es vital dotar al modelo de IA de interpretabilidad, con el fin de comprender la raíz del problema (Golec & Hachaj, 2025). A través de la construcción y entrenamiento de modelos analíticos de aprendizaje «*Learning Analytics*» que emplean algoritmos de aprendizaje automático, el estudio busca anticipar la incidencia potencial de la variable dependiente.

Al alcanzar una reacción reactiva a una intervención proactiva, el sistema de IA genera alertas tempranas basadas en patrones históricos detectados en el «*dataset*» de variables socioemocionales (Coghlan & D'Alfonso, 2021). Para culminar, la integración de estos tres niveles asegura que la propuesta tecnológica tenga una base sólida de fundamentos científicos y pedagógicos. Dichos conceptos estructuran la realidad del estudiante y explican las causas emocionales que sirven como fuente para entrenar al sistema, y así, obtener una herramienta de soporte. En base a lo dicho por Zamora et al., (2023), el alcance de la investigación al ser correlacional - predictivo, busca identificar relaciones entre variables y, partir de ellas, construir un modelo capaz de predecir la probabilidad de abandono escolar, al convertir una base de datos en un mecanismo de «justicia educativa». De esta forma, aporta estrategias de intervención personalizadas que prevengan la exclusión de los jóvenes en el sistema escolar y de una intervención preventiva por parte del DECE.

Tabla 9. Análisis del proyecto: IA para la prevención del abandono escolar

Dimensión	Fase / proceso	Descripción y componentes clave	Propósito en el modelo de IA
Descriptiva	Línea Base	Análisis de características académicas, conductuales y socioemocionales.	Establecer variables « <i>features</i> » de entrada para el entrenamiento (Jin, 2023).
Explicativa	Análisis de Causalidad	Identificación de relaciones entre variables.	Dotar al modelo de «interpretabilidad» (Tomaszewski, 2021).

Predictiva	Learning Analytics	Uso de algoritmos de « <i>Machine Learning</i> » sobre patrones históricos.	Generar «alertas tempranas» para anticipar el abandono (Ventura & Peña, 2021).
Aplicada	Intervención Proactiva	Integración de «fundamentos pedagógicos» y científicos.	Crear una «herramienta de soporte» para el DECE (Zhou & Ye, 2023).

Nota. Elaboración propia basada en MOOC student dropout prediction model based on learning behavior features and parameter optimization (2023).

La aplicación de un sistema de Learning Analytics que convierte datos multidimensionales en una herramienta de inteligencia predictiva. El Departamento de Consejería Estudiantil adquiere la información de alumnos en riesgo para que intervengan de manera ética, y así, garantizar la eficiencia del modelo.

3.1.4. Modalidad

Al tener un carácter aplicado, la investigación desde un punto teórico no se limita a analizar sobre el problema del «abandono escolar», sino que ofrece una solución práctica a esta problemática social (Golec & Hachaj, 2025). Dado que, su propósito fundamental trasciende la reflexión teórica en la resolución de una problemática social, esta modalidad implica que el conocimiento generado se materializa en un producto tecnológico tangible. Ya que, este modelo predictivo resultante no es un ejercicio abstracto, sino una herramienta diseñada para ser integrada en la gestión operativa de las instituciones educativas, transformando los datos socioemocionales en estrategias de retención escolar efectivas.

El estudio asume una modalidad de campo, lo que garantiza la autenticidad y la pertinencia de los datos procesados. Al recolectar la información directamente de las aulas de Bachillerato donde se manifiesta el fenómeno, se asegura que el «*dataset*» de entrenamiento refleje la contextualización del sistema educativo ecuatoriano (James, 2023). En este contexto, la proximidad con la fuente primaria de información ayuda al algoritmo de IA a identificar variables contextuales de modelos genéricos que suelen ignorar. De esta forma su nivel de precisión y confiabilidad es fiable en próximos estudios.

El desarrollo de esta modalidad aplicada y de campo se justifica por la necesidad de buscar el conocimiento a partir de contextos de la realidad local con la Inteligencia Artificial. Por ende, factores socioemocionales conducen al abandono escolar dentro del contexto ecuatoriano, ya que suelen diferir los estándares globales en la recolección de datos *in situ* (Maldonado & Soledispa, 2024). Esta tecnología facilita procesar la

idiosincrasia de los jóvenes de bachillerato en el aprendizaje de los comportamientos específicos y ofrecer respuestas alineadas a su entorno sociocultural. Por consiguiente, la investigación aporta un avance técnico en el área de la analítica de aprendizaje que ofrece una solución socialmente relevante y metodológicamente robusta para el fortalecimiento del sistema educativo nacional (Mota, 2022).

3.1.5. Tipo de estudio

Este estudio es de tipo no experimental, acción que garantiza una observación directa de los fenómenos en un entorno natural, sin alterar la información presentada de las variables independientes (Matbouli & Alghamdi, 2022). Por esta razón, en lugar de manejar datos que van más allá de entornos simulados, se aplica una evaluación con un análisis de los factores socioemocionales en entornos académicos. De esta manera, el contexto natural tal como ocurre en la realidad del estudiante de Bachillerato, facilita la codificación de datos en el sistema (Hernández, 2023). Así mismo, un proyecto de Inteligencia Artificial en el ámbito estudiantil es fundamental, ya que permite entrenar el modelo con datos reales y no sesgados por una intervención.

El estudio aplica un corte transversal, debido a la recolección de información que se sitúa en un punto único y determinado del tiempo. Por ende, este diseño permite realizar un mapeo situacional del estado actual del abandono escolar y sus indicadores emocionales. Al obtener una base de datos de las variables en la muestra seleccionada, se facilita la construcción de un «dataset» estático pero robusto. Es decir, sirve como base para validar la eficacia del modelo de IA en la identificación de patrones de riesgo en un escenario temporal específico antes de su implementación a largo plazo (Jin, 2023).

En cuanto a la perspectiva analítica, la investigación posee un carácter correlacional, ya que se centra en examinar y cuantificar el grado de relación existente entre las variables académicas, conductuales y socioemocionales del estudiante. Este diseño correlacional permite al investigador identificar cómo la ansiedad y el entorno familiar representan una vinculación con la variable dependiente (Golec & Hachaj, 2025).

Este diseño, al ser no experimental, transversal y correlacional, garantiza una base científica sólida para la toma de decisiones basada en datos temporales. Según Pacheco al., (2025), al no intervenir en las variables, el estudio garantiza la ética y la objetividad de los resultados, lo cual genera que el modelo de IA identifique relaciones ya existentes en el ecosistema escolar. Por esta razón, la estructura de esta investigación cumple con los

requisitos metodológicos de una tesis de maestría al proporcionar un marco de referencia confiable en predicciones precisas y contextualizadas para que el modelo logre traducirse en intervenciones pedagógicas.

Tabla 10. Resumen técnico del diseño

Característica	Propósito metodológico	Función en el modelo de IA
No experimental	Garantizar «ética» y «objetividad» al no intervenir en las variables.	Identificación de relaciones preexistentes en el «entorno escolar» (Hernández, 2023).
Transversal	Proporcionar una visión puntual del «estado del estudiante».	Base de prueba «dataset» para la validación de la precisión diagnóstica (Dewey & Faerna, 2022).
Correlacional	Comprobar la influencia de «factores socioemocionales» en la permanencia.	Proveer la «base de datos» para que el sistema categorice riesgos de forma autónoma (Zabala, 2025).

Nota. Elaboración propia basada en Using AI Chatbots to Provide Self-Help Depression Interventions for University Students: A Randomized Trial of Effectiveness. *Internet Interventions* (2022).

En conclusión, el diseño de esta investigación se estructura con el esquema del modelo de Inteligencia Artificial (IA) al identificar patrones preexistentes dentro del fenómeno estudiado que fortalecen la predicción y la funcionalidad analítica del modelo en un contexto educativo.

3.1.6. Diseño de la investigación

El diseño de la investigación es mixto de tipo Explicativo Secuencial (DEXPL), porque equilibra la profundidad subjetiva de la experiencia humana con la rigidez de los parámetros cuantitativos para explicar el proceso de los resultados obtenidos. En base a la investigación de Villaroel et al., (2025), la diferencia del modelo DEXPL con un modelo general, se encuentra en la estructura de las fases secuenciales en donde los resultados cuantitativos son descritos por datos cualitativos. Por ende, al usar dicho diseño en esta investigación funciona como puente de validación estadística para un enfoque sistémico del conocimiento humano frente al fenómeno estudiado.

La estructura que conforma el modelo se describe a continuación:

- Fase Cuantitativa: El tratamiento de datos estructurados se realiza por medio del algoritmo XGBoost en Orange Data Mining. El propósito es establecer valores críticos de riesgo que valide el modelo por medio de medidas específicas de alta precisión. En esta fase, la precisión se respalda en el abastecimiento de valores de Exactitud «Accuracy», Sensibilidad «Recall» y del Puntaje F1 «F1-score». Por

medio de estos indicadores se determina con objetividad qué índice de la muestra se ubica en una zona de riesgo «abandono escolar» (Tomaszewski, 2021).

- Fase Cualitativa: En esta fase se profundiza la interpretación de resultados a través del análisis de narrativas estudiantiles presentes en los reportes del DECE, después de ser detectadas por el patrón estadístico de riesgo. El módulo de Text Mining de Orange, captura la subjetividad de los adolescentes, el cual permite codificar textos a números que revela los factores de cualidad que el algoritmo por sí solo no podría detallar (Coghlan & D'Alfonso, 2021).



Ilustración 1. IA contra el abandono escolar

Por tanto, la metodología DEXPL posibilita que los datos numéricos guíen los resultados de las entrevistas para un análisis cualitativo que tenga coherencia en la triangulación de resultados (Hernández, 2023). En cambio, el sistema de IA integra el sustrato emocional del estudiante, de esta forma, el personal del DECE recibe una explicación cualitativa fundamentada en un sustento investigativo. En el marco metodológico se responde la naturaleza del «abandono escolar», que integra una innovación disruptiva con la realidad que vive el estudiante en el proceso educativo.

3.2. Matriz de Operacionalización de variables

Tabla 11. Operacionalización de las variables en el estudio

Variable	Dimensión	Indicadores
		Desvinculación de entorno

Abandono escolar	Riesgo de abandono	virtuales
		Alerta por bajo rendimiento
	Trayectoria escolar	Continuidad de matrícula
		Tasa de ausentismo
		Abandono escolar
Académico	Conducta	
	Rendimiento	
	Materias reprobadas	
	Asistencia	
Variable socioemocional	Bienestar emocional	Niveles de ansiedad
		Manejo de estrés
		Crisis emocionales
	Habilidades sociales	DECE
		Percepción de convivencia
Reporte de conflictos		
Socioeconómica y entorno	Actividades académicas	
	Presión económica	
	Apoyo familiar	
Tecnológico	Vulnerabilidad socioeconómica	
	Inseguridad del entorno	
	Brecha digital	
		Cobertura de red
		Acceso a dispositivos

Nota. Elaboración propia.

3.3. Población y muestra

La toma de datos poblacionales y de muestra en un proyecto de Inteligencia Artificial, se distingue de la investigación tradicional por la calidad y volumen de información que determinan la capacidad del sistema a evaluar (Matbouli & Alghamdi, 2022).

La población para estudiar en esta investigación se conformará de estudiantes de Bachillerato General Unificado (BGU) de la Unidad Educativa Nelson Torres en la Zona 9 (Distrito Metropolitano de Quito), sector que concentra la mayor vulnerabilidad estructural y «abandono educativo» post-pandemia.

Para calcular el tamaño de la muestra con el rigor estadístico necesario, se utiliza la fórmula para poblaciones finitas, cuyo alcance en subconjuntos seleccionados sea representativo del universo estudiantil (Matbouli & Alghamdi, 2022). En este sentido, el cálculo se basa en un nivel de confianza del 95% con un margen de error máximo del 5%.

Fórmula universal aplicada:

$$n = \frac{N * Z^2 * p * q}{e^2 * (N-1) + Z^2 * p * q}$$

Los parámetros que contiene dicha fórmula se definen de acuerdo con los estándares de precisión científica:

- n: Tamaño de la muestra a obtener
- N: Tamaño de población total (Alumnos de BGU matriculados en el ciclo 2024-2025)
- Z: Nivel de confianza. En un nivel de 95%, el valor crítico es 1,96.
- e: Margen de error admitido 5% que equivale a 0,05.
- p: Probabilidad de ocurrencia del fenómeno.
- q: Probabilidad de no ocurrencia.

Al reemplazar los datos constantes en la fórmula, el proceso se estructura de esta manera:

$$n = \frac{102 * (1,96)^2 * (0,5) * (0,5)}{(0,05)^2 * (N-1) + (1,96)^2 * (0,5) * (0,5)}$$

Este proceso facilita que los datos adquiridos alimenten el algoritmo XGBoost en Orange Data Mining al poseer la validez estadística necesaria para que las predicciones de «abandono escolar» tengan un peso de viabilidad.

Los criterios de selección de la muestra son:

- Criterios de inclusión: Estudiantes registrados en el nivel de Bachillerato para el periodo escolar 2024-2025 y que tenga un historial académico en los registros del Departamento de Consejería Estudiantil (DECE) (Maldonado & Soledispa, 2024).
- Criterios de exclusión: Estudiantes que forman parte de los niveles de Educación Inicial o General Básica (EGB), y cuyos datos se encuentren incompletos en más del 30% lo que frenaría la alimentación del sistema de IA (Hernández, 2023)

En conclusión, la delimitación de la muestra con los reportes que superen el 70% entregados por parte del DECE, se centra en la optimización del sistema al integrar una base de datos de alta densidad para que el algoritmo XGBoost prediga el «abandono escolar» (Maldonado & Soledispa, 2024). De esta forma, adquiere una validez técnica necesaria para resolver problemas en instituciones con mayor grado de «abandono».

3.4. Técnicas e instrumentos de recolección de información.

Dado que la investigación se enfoca en la variable socioemocional en estudiantes,

el modelo predictivo requiere de datos naturales con una diversidad numérica, categórica y ordinal (Mota, 2022). Así, al utilizar una triangulación de técnicas bajo el concepto de DEXPL, asegura que la recolección de datos alimente con precisión en las unidades funcionales independientes de Machine Learning y Text Mining en Orange Data Mining (Tomaszewski, 2021).

- Observación documental: Adquirir variables de control como el historial académico y datos socioeconómicos de la Zona 9 de Quito para describir quienes forman parte del estudio en tiempo real (Delgado & Riquelme, 2022).
- Análisis de contenido: Identificar reportes y opiniones cualitativos de psicólogos y docentes que forman parte del DECE cuya función se basa en la detección del «abandono escolar». Al analizar estos datos se categoriza en modelos CASEL y en categorías emergentes (Ventura & Peña, 2021). Estos relatos son procesados en el módulo de Text Mining de Orange para saber el «porqué» de las alertas.
- Encuesta: Aplicada a estudiantes para comprobar la información con las variables de entorno social y seguridad. Los datos estructurados en la escala Likert genera el dataset primario para entrenar el algoritmo XGBoost para gestionar un umbral numérico de riesgo (Zhou & Ye, 2023).

La implementación de esta triangulación por parte del modelo DEXPL transforma el sistema de la IA en una base multidimensional que el algoritmo detecta como alertas tempranas ante fenómenos cotidianos en instituciones con mayor «abandono escolar».

3.4.1. Formato del instrumento de recolección de datos

Los instrumentos son herramientas de soporte donde se recolectarán los indicadores definidos en la operacionalización. El instrumento principal será el «Cuestionario de Diagnóstico Socioemocional para Predicción de Abandono» (CDSPA). En este formato la estructura facilita la digitalización para convertirlo en un «*DataFrame*» (CSV/JSON) dentro de su procesamiento de datos (Delgado & Riquelme, 2022).

Tabla 12. Cuestionario de diagnóstico socioemocional

Bloque	Dimensión	No. de ítems	Tipo de escala
Datos generales	Perfil demográfico	5	Género, Edad, «estudios de los padres» (Carrillo & Gonzáles, 2023).
Académico	«Rendimiento» y «Disciplina»	4	Promedios, «materias reprobadas», asistencia (Coghlan & D'Alfonso, 2021)
Socioeconómico	«Entorno» y «Tecnología»	4	Cobertura de red (Coghlan & D'Alfonso, 2021)
Socioemocional	«Ansiedad» y «Estrés» (DECE)	10	«Likert» 1-5: Nunca a Siempre (Dewey & Faerna, 2022)
Clima escolar	«Convivencia» y «Seguridad»	5	«Likert» Percepción de bullying (Salazar y otros, 2023).

Nota. Elaboración propia basada en Digital phenotyping: an epistemic and methodological analysis.

Los datos reflejados por parte de Likert, serán tratadas como variables ordinales para la sistematización del modelo de Machine Learning. De esta manera, la robustez del modelo predictivo se influenciará en un proceso riguroso de validación en su operación dentro de áreas educativas con problemas de «abandono».

El CDSPA será valorado por juicio de expertos en el campo de la educación e inteligencia artificial, quienes indicarán cada ítem bajo los criterios de pertinencia. Para tales fines, se vale del Coeficiente de Validez de Contenido (CVC), en el cual se escogerá solo ítems que alcancen un puntaje superior a 0,80. Mientras que, al ejecutar una prueba piloto con 30 estudiantes pertenecientes al bachillerato por medio del alfa de Cronbach (α) (Ventura & Peña, 2021). Con este proceso se considerará un valor de $\alpha \geq 0,70$ para proceder con el abastecimiento de información dentro del algoritmo XGBoost en el cual las variables socioemocionales consten con una predicción de riesgo.

Tabla 13. Rango de Alfa de Cronbach

Alfa de Cronbach	Consistencia Interna
$\alpha \geq 0,9$	Excelente
$0,8 \leq \alpha < 0,9$	Buena
$0,7 \leq \alpha < 0,8$	Aceptable
$0,6 \leq \alpha < 0,7$	Cuestionable
$0,5 \leq \alpha < 0,6$	Pobre
$\alpha < 0,5$	Inaceptable

Nota. Elaboración propia basada en GPLResearch

Se considera un valor de 70 para proceder con el abanto de información.

3.5. Técnica de análisis de datos

A partir del registro recopilado de datos, la información con los instrumentos especificados se ocupará la herramienta Orange Data Mining a través de dos fases concurrentes mediante:

- Análisis de Datos Estructurados (XGBoost): El algoritmo de Gradient Boosting permite procesar variables numéricas y categóricas.

Tabla 14. Tipo de variable y función del modelo

Tipo de Variable	Función en el Modelo	Ejemplo
Numérica	Establece el «umbral predictivo».	«Promedio académico» (López y otros, 2025).
Categoría	Define el «perfil de la muestra».	Sector de la Zona 9.
Ordinal	Mide la «intensidad de la percepción».	Nivel de «estrés» (Bajo/Medio/Alto).

Nota. Elaboración propia basada en Factores socioemocionales que influyen en la deserción escolar: un estudio en contextos de vulnerabilidad (2025).

Así mismo, la técnica SMOTE (*Synthetic Minority Over-sampling Technique*) balancea el «dataset» con desertores y no desertores para evitar sesgos en la formación académica. Por ende, al utilizar valores SHAP se garantiza la explicabilidad del riesgo al identificar qué variables sociodemográficas tienen mayor índice en la predicción (Matbouli & Alghamdi, 2022).

- Procesamiento de Lenguaje Natural (NLP): El uso del Text Mining de Orange, analiza los datos no estructurados de reportes del DECE. Así mismo, el programa asigna niveles semánticos que indiquen «apatía afectiva-institucional» bajo el concepto de CASEL (Zabala, 2025). Dichas narraciones subjetivas transforman estos datos en indicadores de riesgo que el sistema optimizara.
- Integración Multimodal: A través de los resultados de ambos módulos se consolida en un Dropout Probability Score (DPS) final, el cual clasifica el riesgo en niveles (Verde, Amarillo, Naranja, Rojo) según los estándares del MINEDUC. De esta manera, facilita el desarrollo de rutas y protocolos para una intervención ante futuros problemas institucionales (Coghlan & D'Alfonso, 2021).

En este sistema se procesará los datos por medio de la arquitectura multimodal en la plataforma Orange Data Mining. Al usar el algoritmo XGBoost analiza datos estructurados con técnicas SMOTE y SHAP, y minería de texto con NLP para cuantificar la subjetividad de reportes (Hernández, 2023). Por lo tanto, al combinar los análisis resultará en un Dropout Probability Score (DPS) que clasifica el riesgo de abandono escolar en cuatro niveles según las características del MINEDUC.

De esta forma, al analizar la naturaleza del fenómeno, el estudio se adhiere en base a la Ley Orgánica de Protección de Datos Personales (LOPDP) de Ecuador y al principio del interés superior del menor (Asamblea Nacional del Ecuador, 2021). Así, la confidencialidad mediante procesos de seudonimización, donde los datos de cada

estudiante son remplazados por códigos alfanuméricos antes de ser ingresados en la plataforma Orange, se verán seguros dentro del sistema. En esta fase, al tener el acceso a los reportes del DECE bajo protocolos de seguridad restringidos, la información presente se manejará de manera profesional para evitar ser expuesto ante terceros (Ministerio de Educación del Ecuador, 2024).

Al señalar el sesgo algorítmico, el modelo se revisará tras un periodo para detectar disparidades involuntarias asociadas a las condiciones socioeconómicas (Delgado & Riquelme, 2022). Según Basantes et al., (2024), la interpretabilidad constituye un eje ético al funcionar la visión de Orange con valores SHAP, esta unión asegura al modelo como una herramienta guía en el personal del DECE y a los docentes. De esta manera, al combinar esta información con los datos obtenidos, el sistema explica las causas de riesgo que tiene un estudiante en su preparación académica y fortalece la toma de decisiones pedagógicas.

CAPÍTULO IV

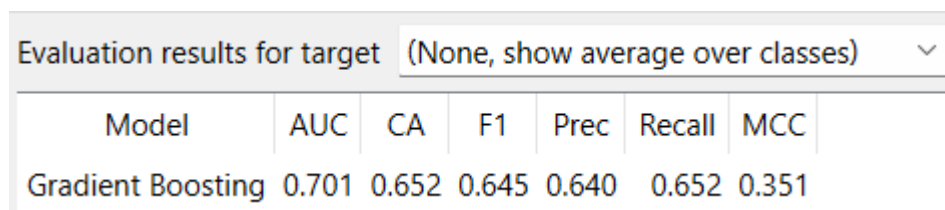
ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

4.1. Análisis de resultados.

En la presente investigación, se realizó una encuesta en Google forms en base a las dimensiones a investigar, en el que participaron sesenta y nueve estudiantes de la Unidad Educativa Nelson Torres en la Zona 9. Esta sección permitió examinar las experiencias y percepciones de los participantes respecto a su formación académica.

De acuerdo con los datos obtenidos se procesaron mediante técnicas de Aprendizaje Automático «Machine Learning» (Eager & Brunton, 2023). Por lo cual, la estructura de este análisis se clasificó en fases de preparación técnica y en la interpretación estadística descriptiva de cada ítem a analizar que garantizó la validez científica del estudio.

A fin de garantizar el rigor científico del estudio, se procedió con la validación y confiabilidad del instrumento en el cual, la medida de fiabilidad de la encuesta se verificó a través del coeficiente Alfa de Cronbach. Por lo tanto, se obtienen valores que el algoritmo mide de manera estable las variables socioemocionales, familiares y económicas a través de reactivos. En cambio, para determinar la confiabilidad del procesamiento de datos, se basó en la evaluación del algoritmo XGBoost junto con la métrica de exactitud «accuracy» y con el F1 score (Barra & Pilicita, 2025). De esta manera, se confirma que el modelo predictivo tiene la capacidad de detectar los factores de riesgo que tienen los estudiantes en el transcurso de su formación académica.



Model	AUC	CA	F1	Prec	Recall	MCC
Gradient Boosting	0.701	0.652	0.645	0.640	0.652	0.351

Ilustración 2. Procesamiento de datos y evaluación del modelo predictivo

La figura señala las métricas de desempeño del algoritmo XGBoost donde el CA «accuracy» acierta con un 65,2%, el F1-Score acierta con un 64,5% y el recall con un 65,2% lo que corrobora la capacidad de predicción que el sistema tiene para detectar los factores de riesgo sobre el análisis de datos presentada.

En primer lugar, la fase de procesamiento de información obtenida, se realizó en una etapa crítica de limpieza de datos para asegurar la calidad del conocimiento y la convergencia del algoritmo XGBoost (Golec & Hachaj, 2025). En este proceso se eliminó la columna «Marca Temporal», ya que, el momento exacto de la respuesta fue un dato irrelevante para el abandono escolar y generó una confusión en la lectura del modelo predictivo. El ítem clave para la detección del abandono escolar se centra en la columna “*He pensado en dejar el colegio en algún momento del presente año lectivo*”. En Orange, esta pregunta nos ayudó con el objetivo central del estudio al cambiarlo por una variable «Target» que detectó el problema a resolver en relación entre otras variables «*features*» (Jin, 2023).

En segundo lugar, el sistema Orange Data Mining para detectar las variables a evaluar el Código del Estudiante con la Edad fueron clasificadas como variables tipo «*meta*», lo que permitió la transparencia del proceso en los futuros casos de riesgo. Las variables que son clasificadas como «*meta*» en el sistema no participan en el cálculo de las funciones de pérdida del algoritmo (Eager & Brunton, 2023). Al vincular cada registro permitió identificar de manera precisa los diferentes riesgos que tuvieron los estudiantes en el ámbito educativo. De esta manera, la intervención del DECE se adoptó a los diversos casos que conforma cada estudiante, sin dejar de lado el entrenamiento del algoritmo dentro del modelo. Por otro lado, las inasistencias recolectadas en rangos de texto de 1-2 día, 3-5 días, fueron codificadas a valores numéricos discretos 2, 4, 6 lo que permitió el cálculo de valores cuantitativos.

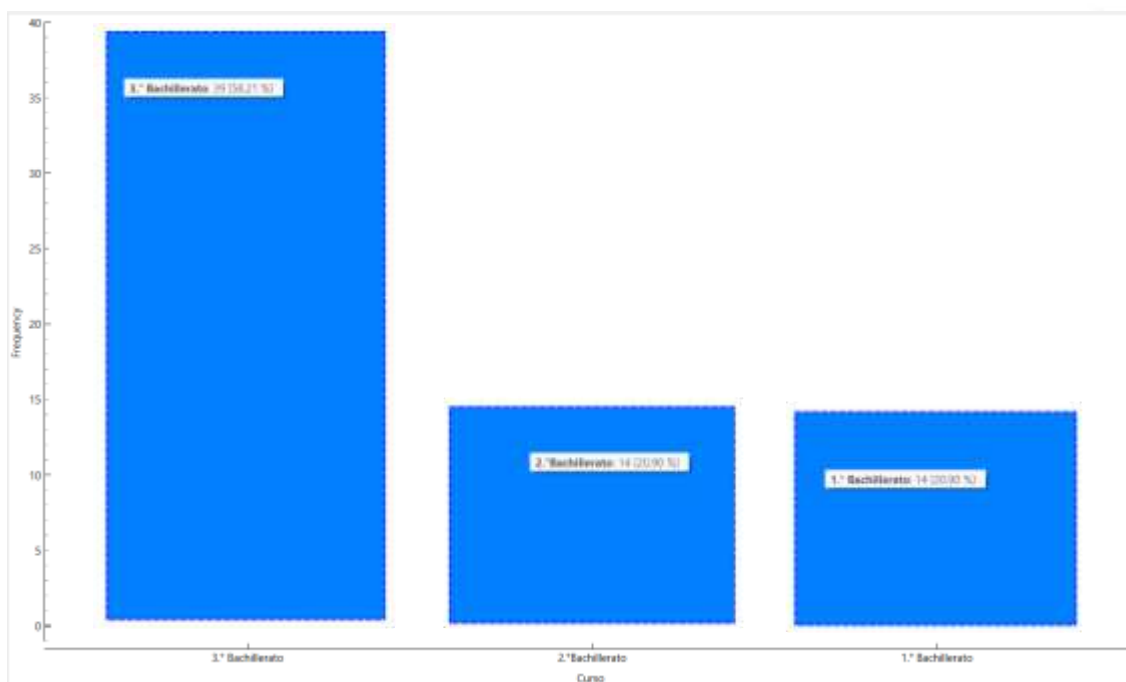
Por consiguiente, para que el sistema Orange Data Mining funcione de manera eficiente, se simplificó los encabezados a etiquetas técnicas para facilitar la lectura en los gráficos de las matrices de importancia. Así, al renombrar los ítems de la encuesta realizada por palabras clave como Carga Académica y Ansiedad - Estrés, facilitó el análisis dentro del programa Orange Data Mining al adquirir puntos clave para que algoritmo XGBoost procesara la información ingresada y de esta manera detecte los principales factores de riesgo en las variables a evaluar (Suberviola, 2021).

En síntesis, el desarrollo de este análisis aseguró que las conclusiones no sean observaciones aisladas, sino parte de un conocimiento que integra el desarrollo de intervenciones aptas para el riesgo del abandono escolar. Así, al finalizar la preparación técnica se procedió a la presentación de los resultados mediante la tabulación de datos y su respectiva representación gráfica.

Tabla 15. Distribución de estudiantes según el curso

Curso	Frecuencia (n)	Porcentaje (%)
1.º Bachillerato	14	20,90%
2.º Bachillerato	14	20,90%
3.º Bachillerato	39	58,21%
Total	67	100,00%

Nota. Elaboración propia.



En la ilustración con la tabla precedente, se visualizó un alto volumen en el tercer año de bachillerato con un valor de 58,21%. Como se describió la fase de preparación técnica, la variable fue categorizada como «Meta» para un mejor entendimiento en la lectura de la información dentro del sistema (Eager & Brunton, 2023). Esta simetría es importante para el modelo predictivo, ya que, el algoritmo XGBoost identificó los factores de riesgo específicos en el transcurso del año lectivo. Por ello, al tener una alta concentración en el último año

lectivo, es importante destacar que los indicadores de abandono escolar incrementan por la carga académica y la transición profesional.

Tabla 16. Diversidad de género de los participantes

Género	Frecuencia (n)	Porcentaje (%)
Femenino	38	55,88%
Masculino	29	42,65%
Prefiero no decirlo	1	1,47%
Total	68	100,00%

Nota. Elaboración propia.

Nota. Elaboración propia.

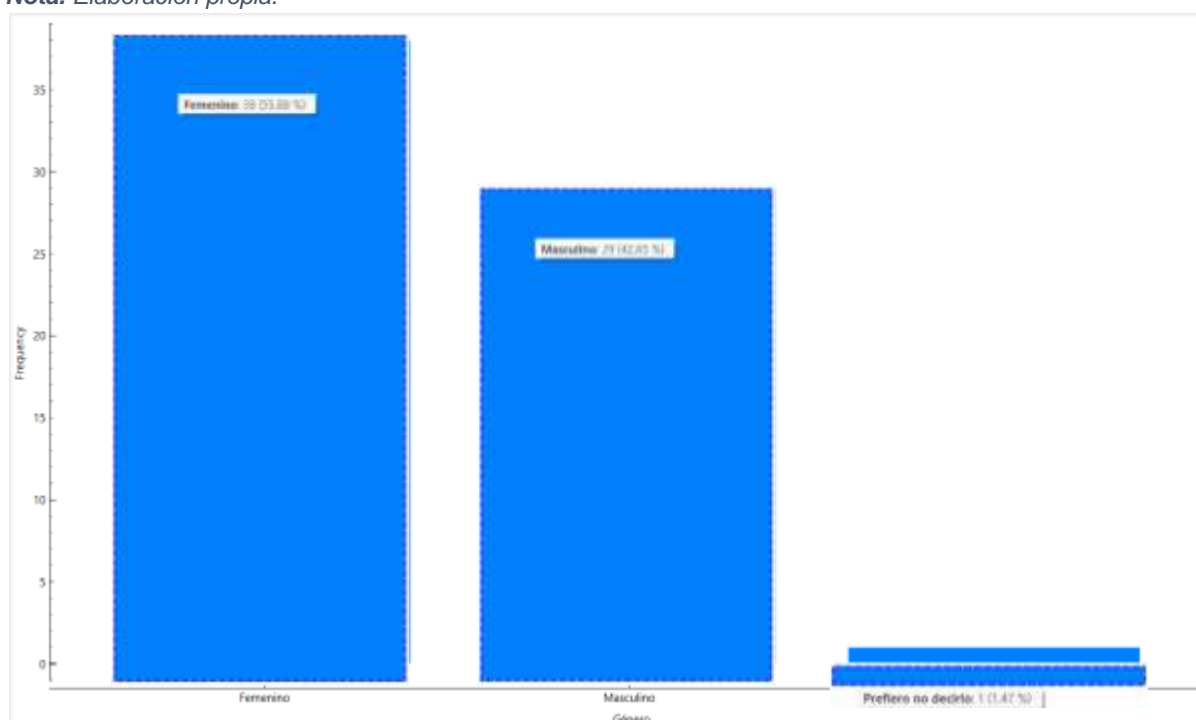


Ilustración 4. Diversidad de género de los participantes

En la siguiente tabla con la ilustración presentada, se identificó una mayor participación por parte del género femenino con una representación porcentual del 55, 88, seguida del género masculino con 42,65 y con una privacidad del género con un valor de 1,47. Dado que, existe una distribución de género, se evidenció una predominancia femenina, lo que resulta relevante en el análisis del estudio. Sin embargo, al tener un índice de 1,47% de personas que opta por la privacidad de su género, esto fortalece la inclusión en el análisis dentro del sistema que identifica la diversidad de identidades. Esta distribución permitió al algoritmo entender de forma integral y presentar variaciones según el género en las variables críticas como la Ansiedad Estrés y la Carga Académica (Suberviola, 2021).

Tabla 17. Estadística descriptiva de comprensión de contenidos

Opciones	Frecuencia (n)	Porcentaje (%)
1	3	4,35%
2	3	4,35%
3	11	15,94%
4	27	39,13%
5	25	36,23%
Total	69	100,00%

Nota. Elaboración propia.

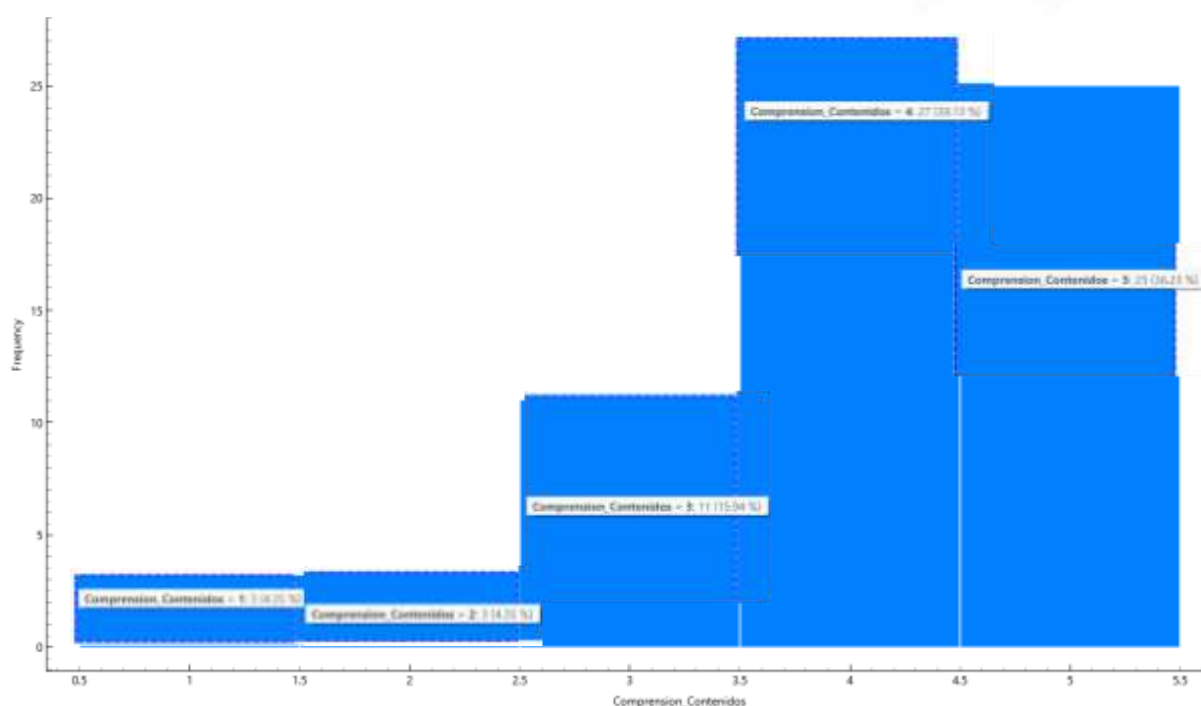


Ilustración 5. Estadística descriptiva de comprensión de contenidos

En la presente ilustración junto con la tabla, se visualizó una comprensión de contenidos de 27 participantes con un valor porcentual de 39,13%, lo que representó una percepción positiva de los estudiantes respecto a su entendimiento académico. Al tener un resultado positivo en este ítem, reflejó en los estudiantes una comprensión positiva en los contenidos impartidos por el docente, lo que generó en el estudiantado una incorporación en la formación educativa (Aguilera, 2022). De esta manera, el sistema Orange Data Mining al detectar una dispersión de 4,35% distribuidos en la opción 1 y 2, este dato representó una cifra en el conocimiento que no fue adecuada. Por lo que, al utilizar este valor como un indicador de estabilidad dentro de la formación estudiantil para entrenar el modelo predictivo.

Tabla 18. Análisis de la variable cumplimiento de tareas

Opciones	Frecuencia (n)	Porcentaje (%)
1	2	2,90%
2	10	14,49%
3	11	15,94%
4	18	26,09%
5	28	40,58%
Total	69	100,00%

Nota. Elaboración propia.

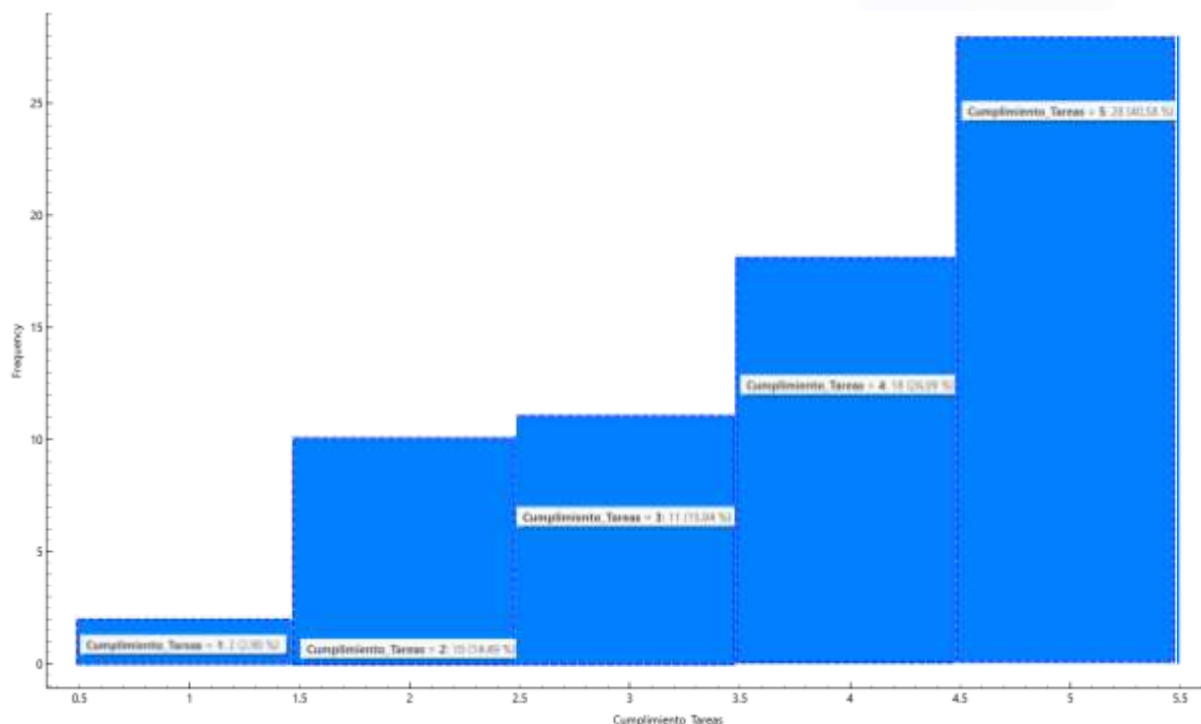


Ilustración 6. Análisis de la variable cumplimiento de tareas

Al analizar este ítem, se muestra una tendencia de alto nivel en la opción 5, con una participación de 28 estudiantes que equivale al 40.58%. Dicho porcentaje, señaló que una gran parte de los alumnos encuestados percibieron un cumplimiento de tareas, lo que generó una motivación al cumplir con sus obligaciones y generando la continuidad en el proceso educativo. Por otro lado, el 2,90% de los estudiantes, el sistema los detectó de una manera directa como un grupo crítico, ya que, el cumplimiento de tareas se reflejó en su rendimiento académico y generó una desmotivación en el estudiante. En esta parte el sistema tomó como alerta al cumplimiento de tareas como una señal de riesgo ante posibles casos de abandono debido a que existe una interacción con factores críticos que aumentan su probabilidad (Zhou & Ye, 2023).

Tabla 19. Análisis de la variable rendimiento y esfuerzo

Opciones	Frecuencia (n)	Porcentaje (%)
1	2	2,90%
2	4	5,80%
3	5	7,25%
4	26	37,68%
5	32	46,38%
Total	69	100,00%

Nota. Elaboración propia.

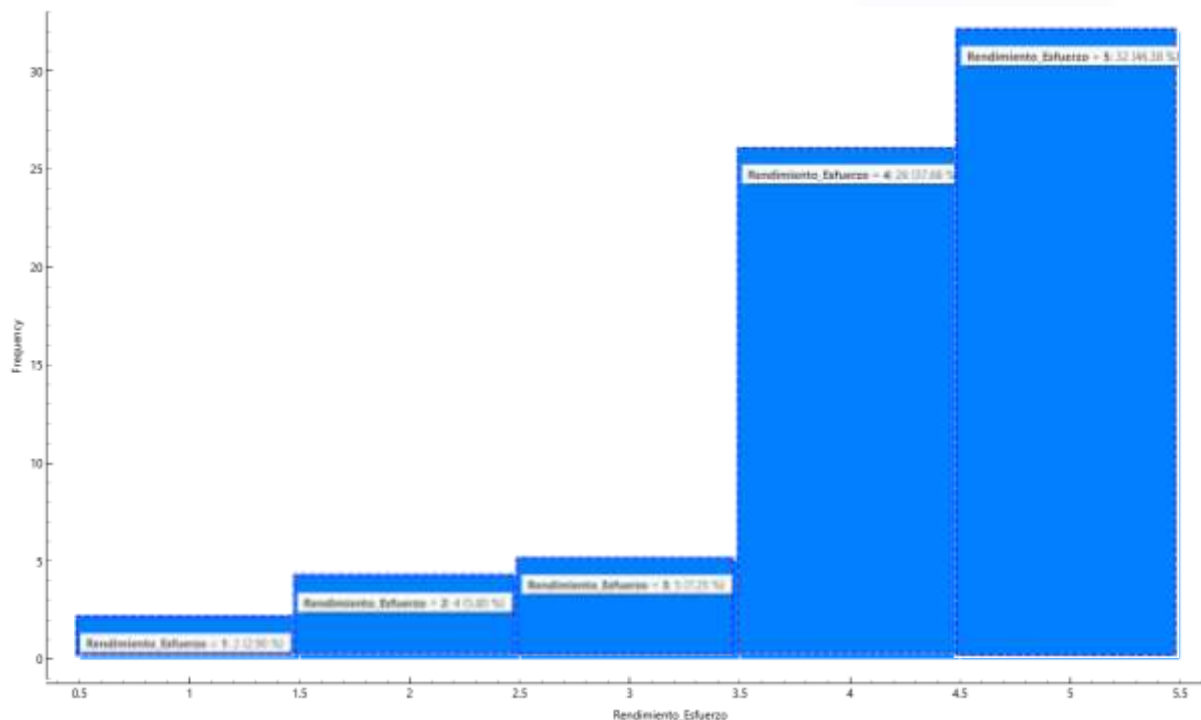


Ilustración 7. Análisis de la variable rendimiento y esfuerzo

En la presente ilustración se determinó un porcentaje de 46,38% que equivale a 32 estudiantes los cuales percibieron una autopercepción de alto desempeño. Al tener un resultado positivo de más de la mitad de la población, el rendimiento y esfuerzo se asocia con niveles adecuados de motivación y confianza en el proceso académico. Por otro lado, el algoritmo al tener un porcentaje bajo de 2,90% de los alumnos, reduce la probabilidad de clasificar la variable como un factor de riesgo de abandono. De este modo, al tener una mayor participación de estudiantes en esta categoría, el sistema disminuye la probabilidad de riesgo a grupos vulnerables de abandono escolar (Suárez & Castro, 2022). En otros términos, no significa que el algoritmo carezca de importancia debido a su porcentaje presentado, ya que, su impacto de riesgo es limitado.

Tabla 20. Distribución de la carga académica

Opciones	Frecuencia (n)	Porcentaje (%)
1	12	17,39%
2	10	14,49%
3	20	28,99%
4	17	24,64%
5	10	14,49%
Total	69	100,00%

Nota. Elaboración propia.

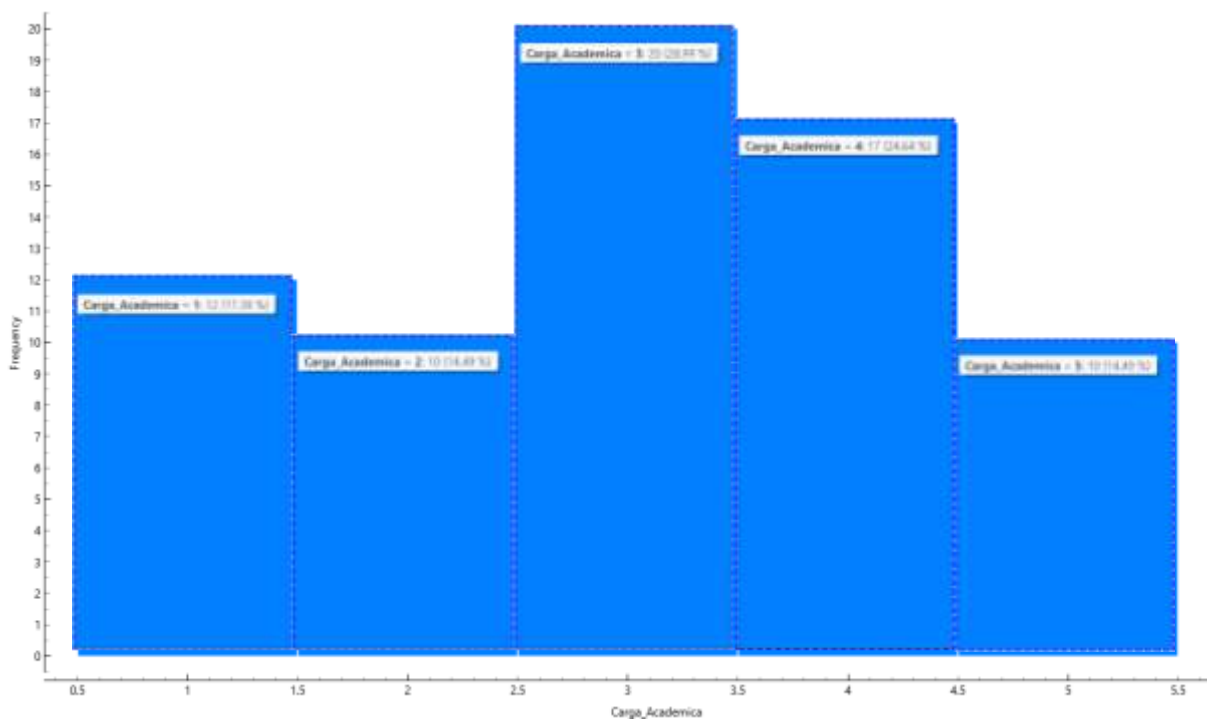


Ilustración 8. Distribución de la carga académica

En este índice, la muestra se diferenció en un punto más alto en el nivel medio con un porcentaje de 28,99% de los estudiantes encuestados. De este modo, el algoritmo toma como relevante la variable por la diversidad de resultados, lo que lo posiciona como un potencial estresor que desarrolla alertas de riesgo socioemocionales en el modelo predictivo. Dado que el sistema está programado para detectar variables críticas, el algoritmo identificó patrones más precisos en relación con el abandono escolar. En conclusión, al tener una mayor concentración en el nivel intermedio otorga un valor analítico significativo al modelo predictivo para detectar situaciones de riesgo (Infantas & Heredia, 2023).

Tabla 21. Análisis de la variable apoyo docente

Opciones	Frecuencia (n)	Porcentaje (%)
1	6	8,70%
2	1	1,45%
3	2	2,90%
4	15	21,74%
5	45	65,22%
Total	69	100,00%

Nota. Elaboración propia.

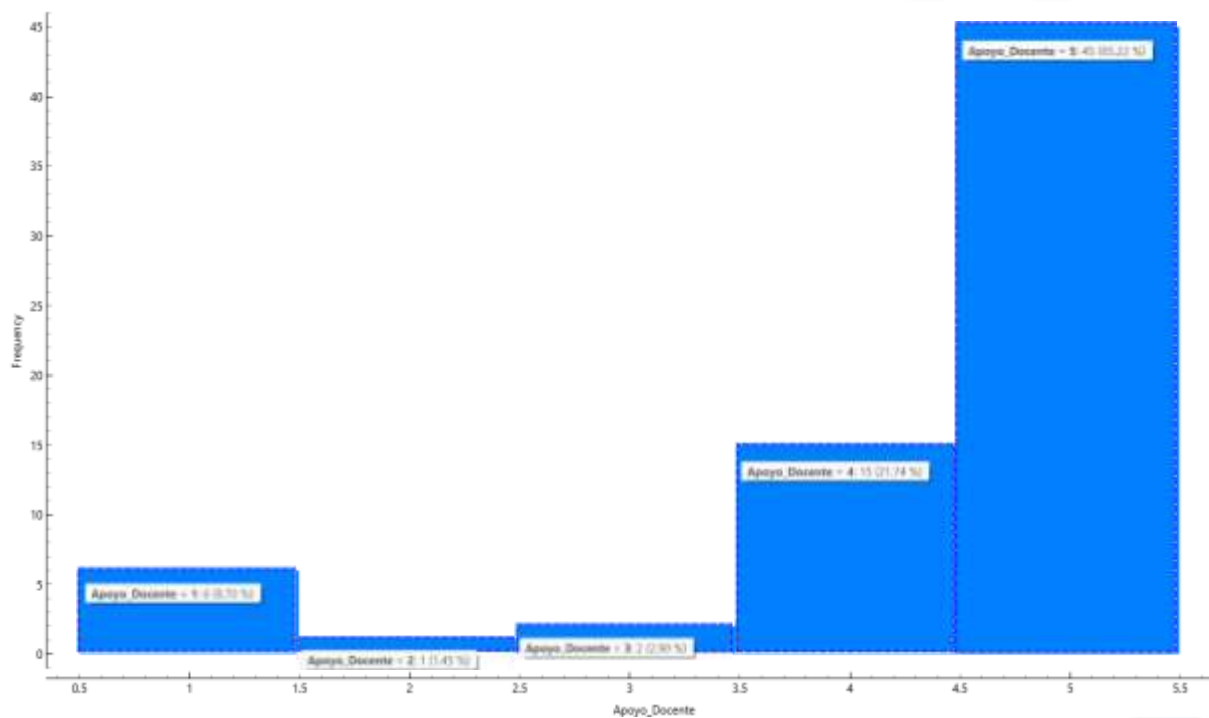


Ilustración 9. Análisis de la variable apoyo docente

En este apartado, el apoyo docente reflejó un valor de 65,22% por parte de los estudiantes encuestados, el cual operó como el eje central que impulsa el éxito académico y el equilibrio emocional. Esta variable se proyecta como un factor protector en la formación académica frente a otras variables de riesgo que el algoritmo detectó después de codificar la minería de datos suministrada. Dado que, este porcentaje es positivo en los datos entregados, el sistema lo coloca como un factor protector en relación con el abandono escolar. En base a lo propuesto por Mendevíl et al., (2025), al tener una relevancia positiva en el estudio, refleja que los alumnos identifican al docente como un acompañamiento integral que favorece el fortalecimiento y la estabilidad académica.

Tabla 22. Análisis de la variable agotamiento emocional

Opciones	Frecuencia (n)	Porcentaje (%)
1	21	30,43%
2	11	15,94%
3	15	21,74%
4	12	17,39%
5	10	14,49%
Total	69	100,00%

Nota. Elaboración propia.

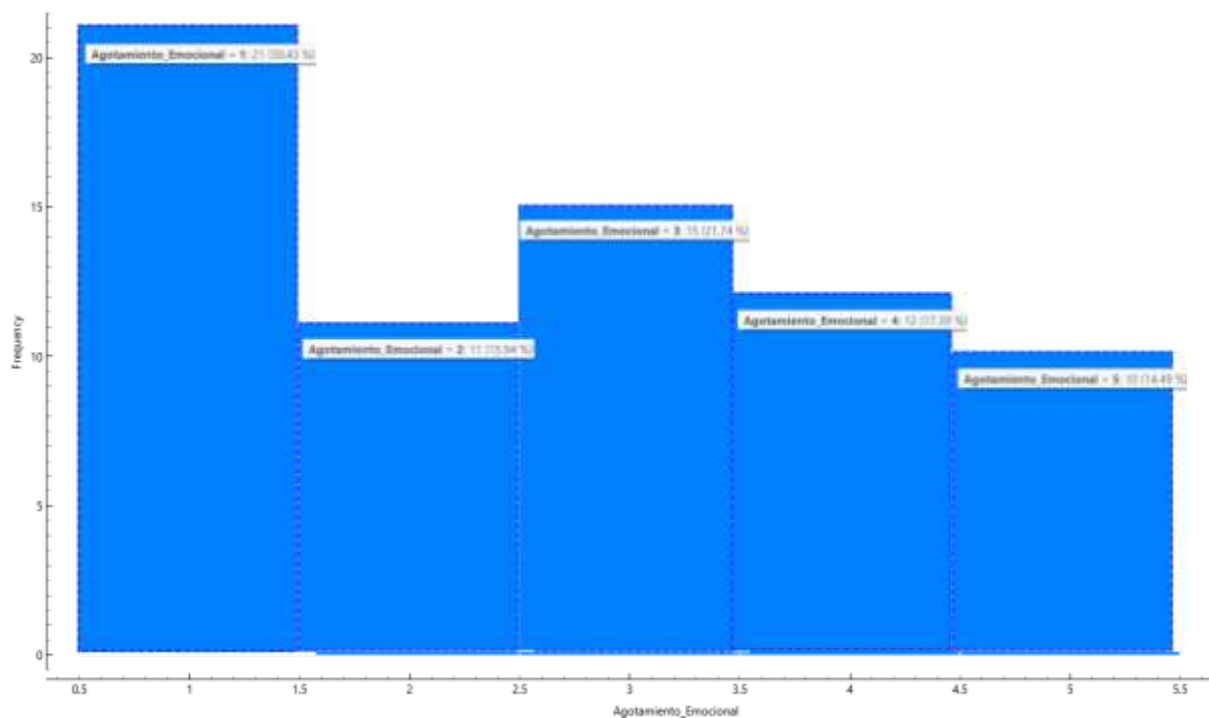


Ilustración 10. Análisis de la variable agotamiento emocional

A pesar de tener un índice de 30,43% en el agotamiento emocional de los estudiantes, un 31,88% que equivale a un tercio de los alumnos que se encuentran en el punto 4 y 5 señalaron una fatiga mental preocupante en la muestra. Esta variable es detectada como un punto crítico para el sistema predictivo al tener un desgaste emocional que afecta en el desarrollo de aprendizaje por parte del alumno. Así, al presentar una acumulación de carga académica y factores emocionales, el desarrollo de un modelo de acompañamiento integral toma relevancia mitigar estos cuadros críticos para estabilizar las emociones del estudiante. Al funcionar como un catalizador que acelera el riesgo de abandono escolar, el software de Orange Data Mining realiza el proceso de detección que categoriza posibles riesgos basado en el aprendizaje autónomo «*Machine Learning*» (Eager & Brunton, 2023).

Tabla 23. Niveles de ansiedad y estrés percibido

Opciones	Frecuencia (n)	Porcentaje (%)
1	27	39,71%
2	10	14,71%
3	19	27,94%
4	6	8,82%
5	6	8,82%
Total	69	100,00%

Nota. Elaboración propia.

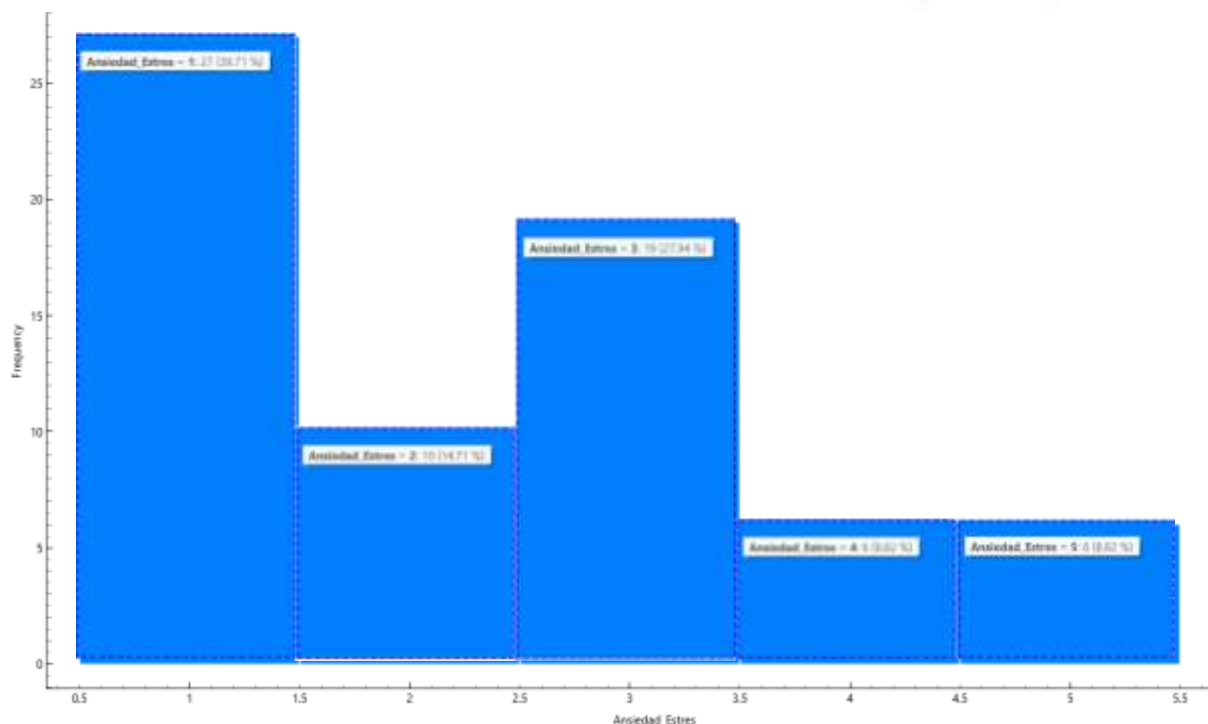


Ilustración 11. Niveles de ansiedad y estrés percibido

Esta ilustración señaló que la población encuestada determina un nivel de ansiedad y estrés correspondiente al 39,71%, lo que significa que una parte de la muestra experimenta malestar en la formación académica. Este porcentaje implica que los factores psicoemocionales inciden en el desempeño académico y la permanencia escolar, que constituye un elemento de atención para el análisis del estudio. Sin embargo, el sistema Orange Data Mining profundiza la clasificación de los niveles emocionales al analizar grandes volúmenes de información (Barra & Pilicita, 2025). Por lo tanto, en las opciones 4 y 5 el sistema detectó valores de 8,82% presentes en los estudiantes asociados a una inestabilidad emocional que el algoritmo categoriza como un factor para el abandono escolar. Desde un enfoque analítico, los

resultados obtenidos establecen la relación entre los estados emocionales y el riesgo de abandono escolar que representan un punto crítico que afecta la continuidad académica.

Tabla 24. Desmotivación académica

Opciones	Frecuencia (n)	Porcentaje (%)
1	31	46,27%
2	7	10,45%
3	12	17,91%
4	5	7,46%
5	12	17,91%
Total	69	100,00%

Nota. Elaboración propia.

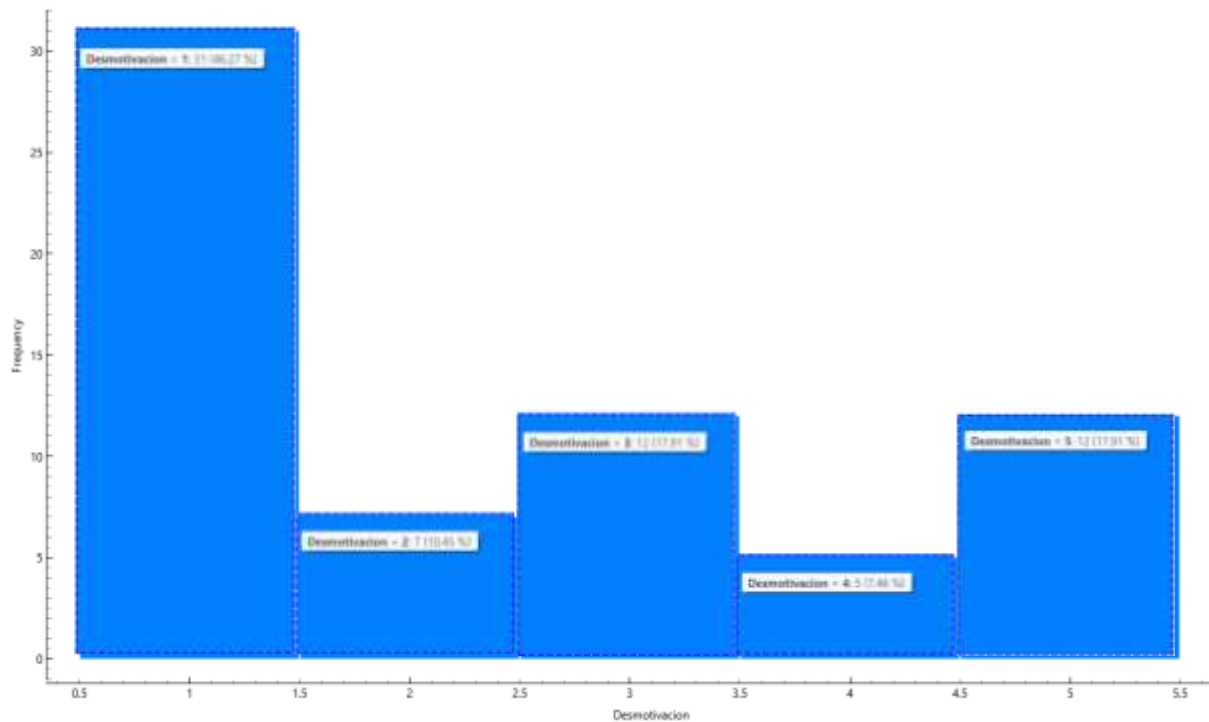


Ilustración 12. Desmotivación académica

En la siguiente ilustración, se comprobó que el 46,27% de la mitad de los estudiantes encuestados no presentan indicadores de desmotivación, lo que significa que parte de la muestra mantiene niveles de interés y compromiso en la educación. El algoritmo al detectar estos valores en su sistema, categorizó esta variable como un factor protector con relación al bajo rendimiento y al abandono escolar (Delgado & Riquelme, 2022). Desde esta perspectiva, la motivación se caracterizó en el ámbito educativo como un componente clave en los logros académicos. Por otro lado, valores de 17,91% en las opciones 3 y 5 los estudiantes representan un nivel máximo de desinterés que requiere medidas de intervención adecuada.

Así, el modelo predictivo, identifica patrones de riesgo que permitió clasificar este grupo como un nivel de alto riesgo ante posibles casos de abandono. Al colocar este nivel a la variable, actúa como un factor predictor en estudiantes con desinterés máximo que el sistema detectó para disminuir el índice de abandono escolar. A pesar de que, exista un porcentaje alto con motivación estudiantil, establecer el desarrollo de estrategias focalizadas en el rendimiento académico generará una disminución del riesgo escolar.

Tabla 25. Capacidad para afrontar retos académicos

Opciones	Frecuencia (n)	Porcentaje (%)
1	3	4,41%
2	6	8,82%
3	10	14,71%
4	15	22,06%
5	34	50,00%
Total	69	100,00%

Nota. Elaboración propia.

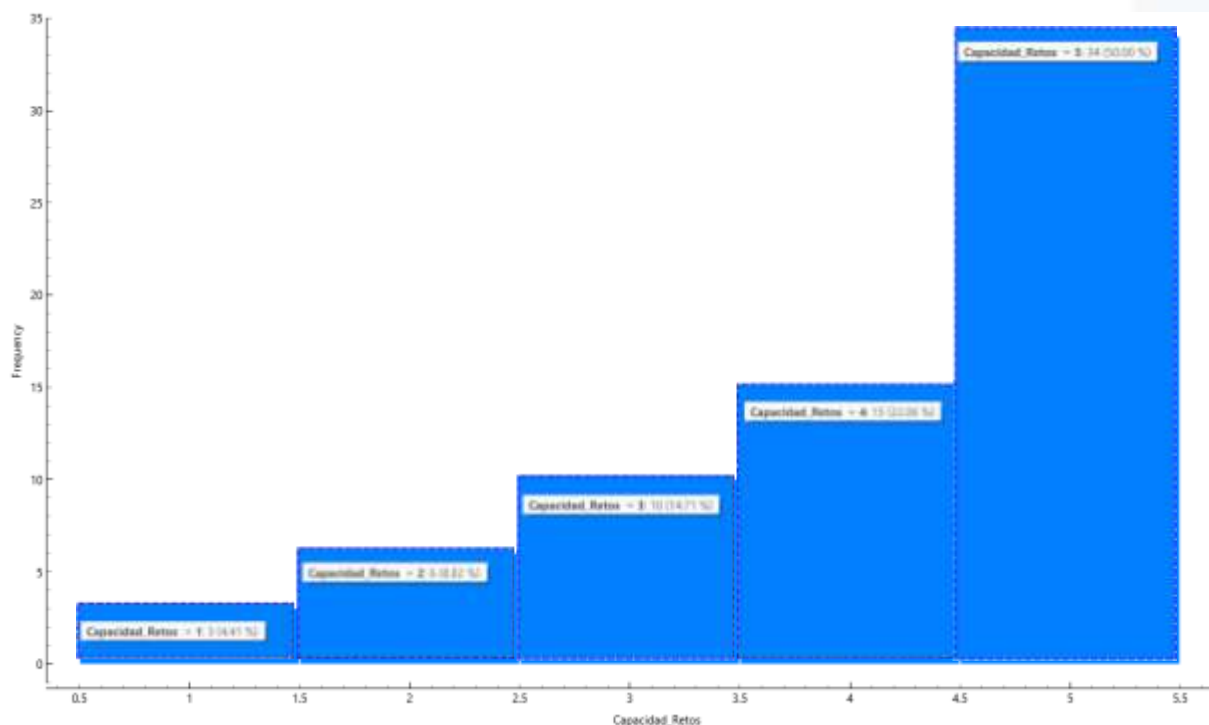


Ilustración 13. Capacidad para afrontar retos académicos

En los hallazgos obtenidos de esta variable, se evidenció una tendencia positiva del 50% de estudiantes en el nivel máximo de la escala, lo que apunta hacia una autoeficacia académica. Esta proporción de la población estudiantil que confió en sus capacidades para enfrentar los desafíos del proceso formativo, el sistema los destaca con una menor tendencia a posibles

riesgos de abandono (Suárez & Castro, 2022). Por ello, el modelo predictivo interpreta este índice como un mecanismo de defensa dentro del análisis que el sistema clasifica. Desde este punto, cuando se manifiestan índices prominentes en la desvinculación académica, la autoeficacia aporta un peso significativo en la reducción de la probabilidad de abandono escolar.

Tabla 26. Sentido de pertenencia institucional

Opciones	Frecuencia (n)	Porcentaje (%)
1	5	7,25%
2	5	7,25%
3	6	8,70%
4	14	20,29%
5	39	56,52%
Total	69	100,00%

Nota. Elaboración propia.

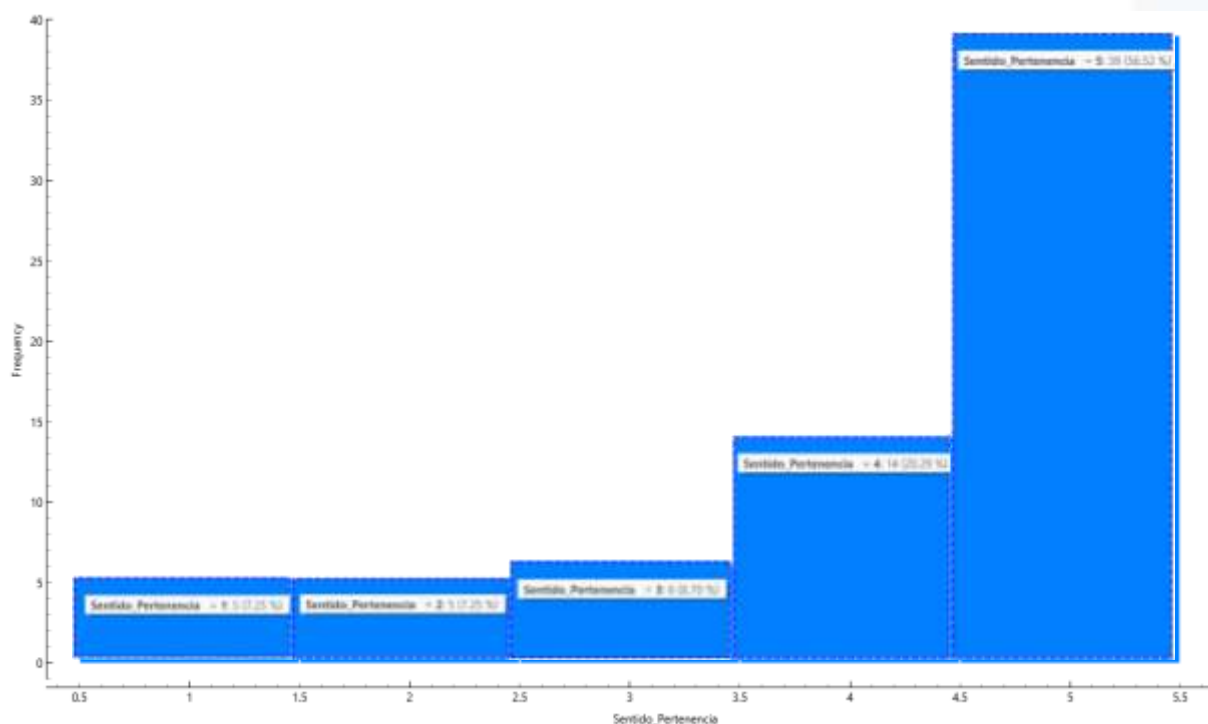


Ilustración 14. Sentido de pertenencia institucional

Tras los hallazgos obtenidos en este ítem, más de la mitad de la población destaca un nivel de 56,52% en la identificación con los valores, la integración en equipo y el compromiso hacia propósitos académicos de excelencia. De esta manera, estos resultados colocan al sentido de

la pertenencia como un indicador clave en la desvinculación académica dentro de la formación estudiantil. En base a la investigación realizado por Álvarez et al., (2025), esta variable contribuye al entrenamiento del sistema en la reducción del abandono escolar por tener una alta participación de estudiantes. Al referirse como un amortiguador, su función dentro del sistema de indicadores toma la dirección con relación a estudiantes con mayor sentido de pertenencia y a la presencia de estresores externos. Por ende, dicha variable opera como protector psicosocial en el fortalecimiento del vínculo institucional y estudiantil al incrementar la persistencia escolar frente a nuevos cambios de enseñanza. Por lo tanto, estos resultados fortalecen las políticas educativas orientadas a la participación estudiantil y la cohesión comunitaria para lograr un impacto directo en el índice del abandono escolar.

Tabla 27. Apoyo emocional por parte del docente

Opciones	Frecuencia (n)	Porcentaje (%)
1	5	7,25%
2	3	4,35%
3	5	7,25%
4	14	20,29%
5	42	60,87%
Total	69	100,00%

Nota. Elaboración propia.

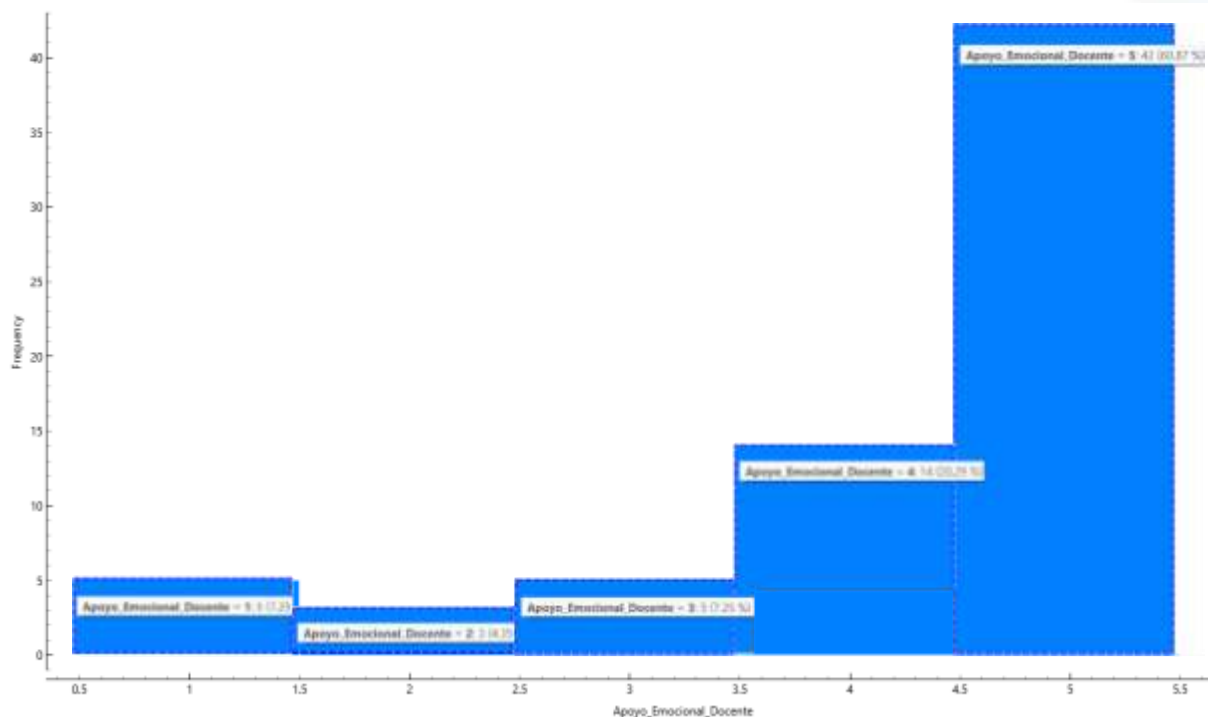


Ilustración 15. Apoyo emocional por parte del docente

En cuanto a la variable apoyo emocional por parte del docente, la muestra poblacional presentó un grado de 60,87% lo que significa que más de la mitad de la muestra encuestada percibió al educador como un compañero de respaldo ante dificultades personales y académicas. Esto señaló que, a mayor participación por parte del docente en presencia de problemas internos y externos de los estudiantes, menor será el porcentaje del riesgo que el sistema detectó. Según Cayambe et al., (2024), en el transcurso del entrenamiento y la validación del modelo predictivo, los estudiantes que tengan un porcentaje alto de apoyo emocional son menos propensos a desertar la formación académica. Esta variable al ser detectada por el algoritmo como protector, protege a estudiantes de factores adversos que presentan en la institución educativa. Por lo tanto, este porcentaje garantiza una percepción positiva del rol docente al configurar el sistema educativo como un componente estratégico en modelos preventivos al abandono escolar.

Tabla 28. *Relación con amigos en el colegio*

Opciones	Frecuencia (n)	Porcentaje (%)
1	8	11,59%
2	7	10,14%
3	7	10,14%
4	17	24,64%
5	30	43,48%
Total	69	100,00%

Nota. Elaboración propia.

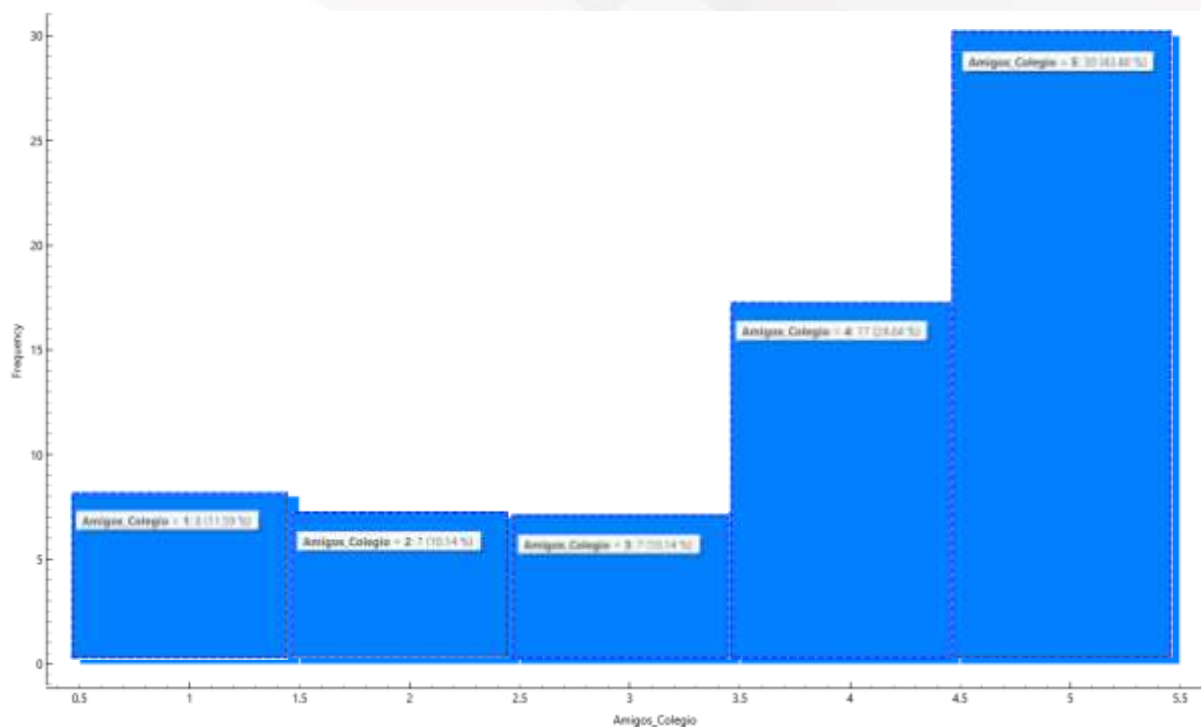


Ilustración 16. Relación con amigos en el colegio

En esta evidencia de las personas encuestadas, el 43,48% de la población destacó tener una excelente relación con sus compañeros de estudio, lo que refleja un ambiente sano en grupos que mantienen vínculos interpersonales positivos. Según Vélez et al., (2020), este porcentaje propone el desarrollo de estrategias favorables para la convivencia y apoyo entre pares que potencien el clima educativo en el sentido de la integración social. En cambio, el 21,73% de la muestra se encuentra en la opción 1 y 2, que refleja una baja interacción entre compañeros. El sistema lo cataloga como una parte vulnerable en relación a variables fuertes como la pertenencia, la motivación y el compromiso educativo. Así, el modelo advierte que los estudiantes que se destacan por trabajar de una manera individual son más vulnerables a factores de estrés y apoyo emocional al vincularse en trabajos grupales.

Tabla 29. Análisis de ausentismo

Opciones	Frecuencia (n)	Porcentaje (%)
0	28	40,58%
2	18	26,09%
4	12	17,39%
6	11	15,94%
Total	69	100,00%

Nota. Elaboración propia.

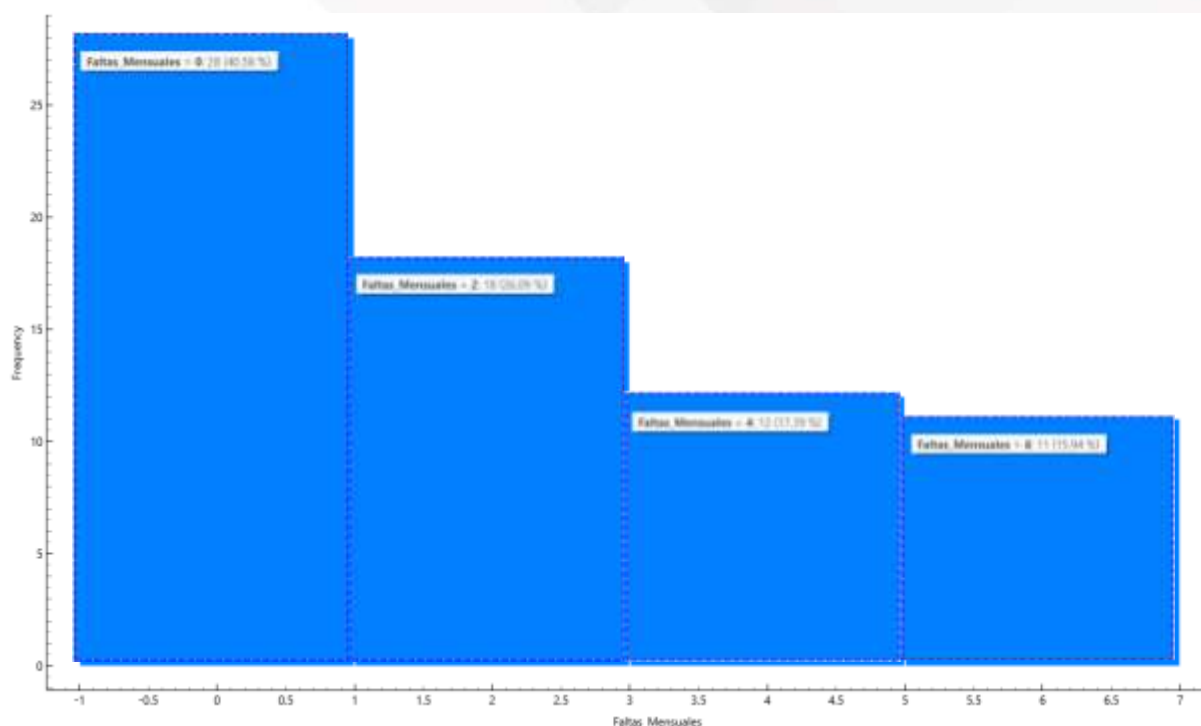


Ilustración 17. Análisis de ausentismo

Desde un enfoque descriptivo, el 40,58% de la población encuestado señaló que su asistencia en la institución educativa está orientada en el proceso de enseñanza y aprendizaje en su periodo académico. Por el contrario, el 15,94% de los estudiantes presentan niveles críticos de más de 6 días de inasistencia en la institución, lo cual el algoritmo emite alarmas sobre esa población ante posibles riesgos de abandono. Por otro lado, el sistema Orange Data Mining, marca a esta variable con valores de ganancia «Gain», ya que, este indicador mide la capacidad de las variables entre estudiantes con o sin intención de abandono. Por ello, esta variable al tener los valores más altos de «Gain» dentro del sistema, cumple con el propósito de explicar y predecir el fenómeno que se está estudiando. En base a lo investigado por Molina (2025), si el número de inasistencias incrementa, se correlaciona al de abandono escolar que debilitó la continuidad educativa de los alumnos. Así, esta variable se transforma en un punto clave para clasificar y detectar posibles riesgos que el sistema emitirá para que exista una intervención preventiva enfocadas en estudiantes con ausentismo de más de 6 días.

Tabla 30. Importancia percibida del bachillerato

Opciones	Frecuencia (n)	Porcentaje (%)
1	3	4,35%
2	0	14,49%
3	1	1,45%

4	4	5,80%
5	61	88,41%
Total	69	100,00%

Nota. Elaboración propia.

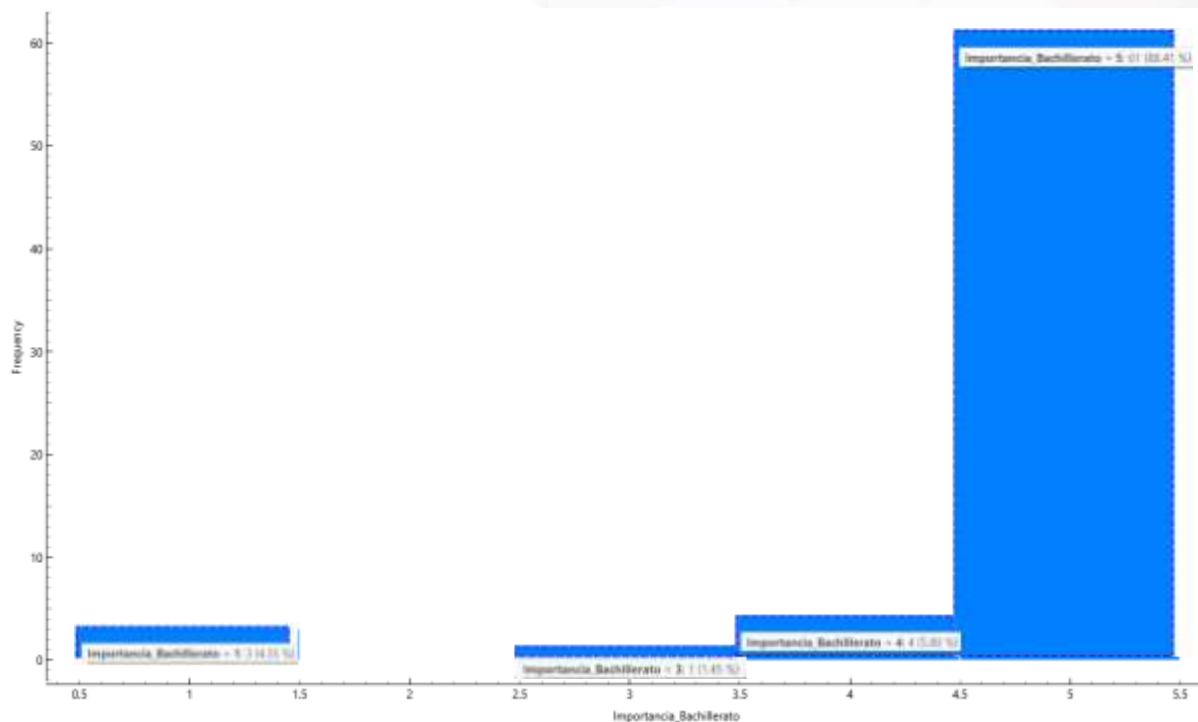


Ilustración 18. Importancia percibida del bachillerato

Al medir la valoración de los estudiantes para obtener un título a futuro, gran parte de la población encuestada destaca un compromiso en la culminación de sus estudios con una incidencia del 88,41%. Dicho porcentaje posiciona a la variable dentro del sistema como un indicador de motivación que oriente a los estudiantes a cumplir sus metas, a la continuidad formativa y a su visión a largo plazo (Gracia & Sánchez, 2020). De esta manera, el modelo predictivo integra esta variable como ancla al tener una participación positiva en la culminación de sus estudios y reducir el sesgo institucional en el abandono.

Tabla 31. Situación familiar percibida

Opciones	Frecuencia (n)	Porcentaje (%)
1	29	42,65%
2	7	10,29%
3	8	11,76%
4	10	14,71%
5	14	20,59%
Total	69	100,00%

Nota. Elaboración propia.

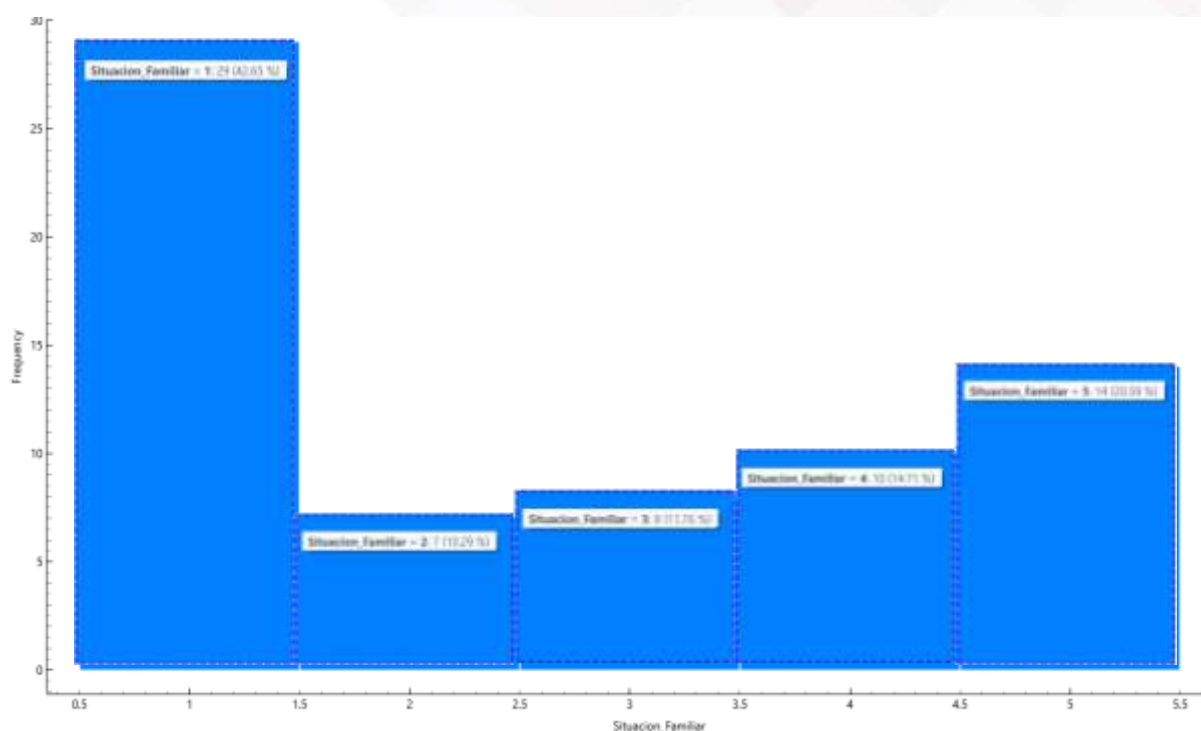


Ilustración 19. Situación familiar percibida

En cuanto a la percepción y apoyo familiar dentro del hogar, el 42,65% de los estudiantes encuestados destacan tener un entorno favorable en el transcurso de su proceso educativo. Desde la perspectiva de Concepción et al., (2024), este ambiente es adecuado para cuidar a los estudiantes frente a factores internos y externos que al crear una red de seguridad se optimiza el aprendizaje para la sostenibilidad del compromiso estudiantil. Por otro lado, existe un índice del 20,59% de personas vulnerables al padecer de problemas familiares que incide en el proceso y en su compromiso educativo dentro de la institución. El algoritmo a través de la minería de datos interpreta esta variable como un factor de riesgo, al alterar la estabilidad emocional junto con su compromiso en la culminación de sus estudios. Al tener una diversidad de dimensiones que afectan a otros indicadores negativos, el abandono escolar se incrementara valores críticos que el sistema detectara de inmediato.

Tabla 32. Motivos laborales

Opciones	Frecuencia (n)	Porcentaje (%)
Sí	43	62,32%
No	26	37,68%
Total	69	100,00%

Nota. Elaboración propia.

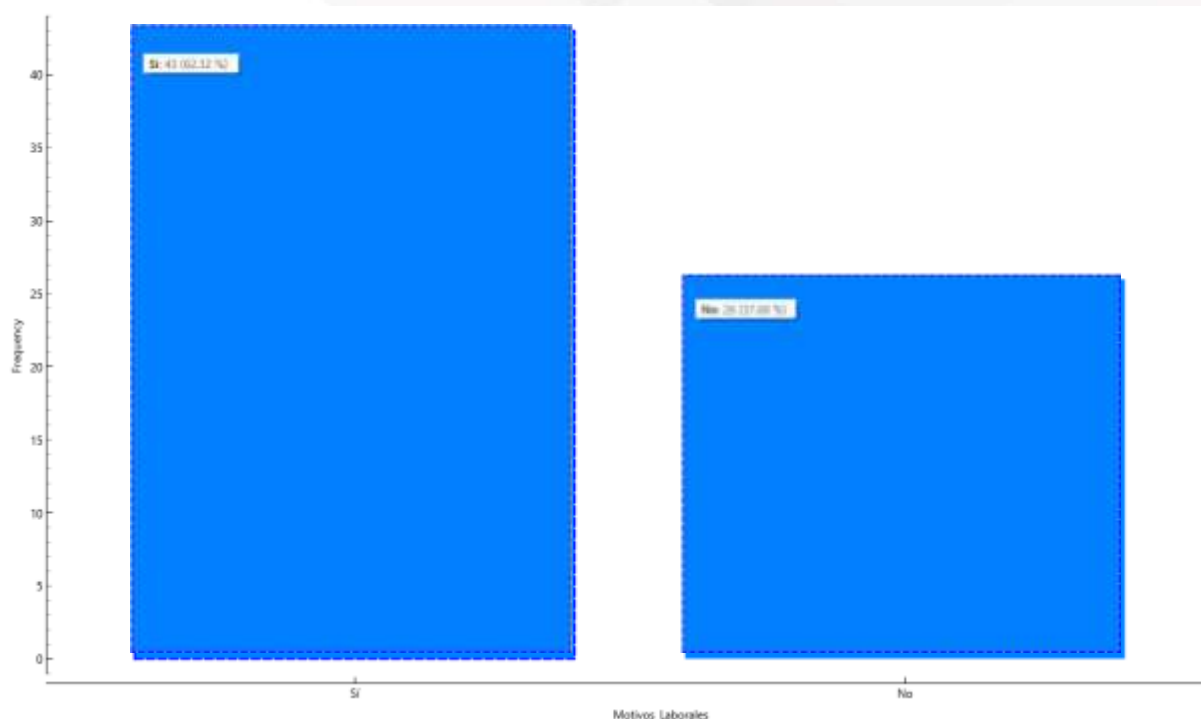


Ilustración 20. Motivos laborales

En cuanto al índice de motivos laborales, más del 62,32% de la población indicó que dicho factor impide el seguimiento de sus estudios académicos, debido al tiempo laboral, al desempeño que ejercen en cada clase y a la permanencia educativa (Gairín & Olmos, 2022). Por otro lado, el 37,68% de los encuestados representa una menor proporción de la población al señalar que la situación laboral no es un factor crítico para continuar en el proceso educativo. Por ello, al tratar una variable binaria nominal, su análisis se destaca en la comparación paralela entre ambas categorías. Este efecto permitió al modelo predictivo detectar un mayor volumen de respuestas en la categoría SI dentro de las personas vulnerables, lo cual presenta una mayor frecuencia en el bajo rendimiento académico. En conclusión, los motivos laborales se configuran en el sistema como una variable contextual que se considera con niveles altos de riesgo académico ante posibles causas de abandono.

Tabla 33. Motivos económicos

Opciones	Frecuencia (n)	Porcentaje (%)
Sí	17	24,64%
No	52	75,36%
Total	69	100,00%

Nota. Elaboración propia.

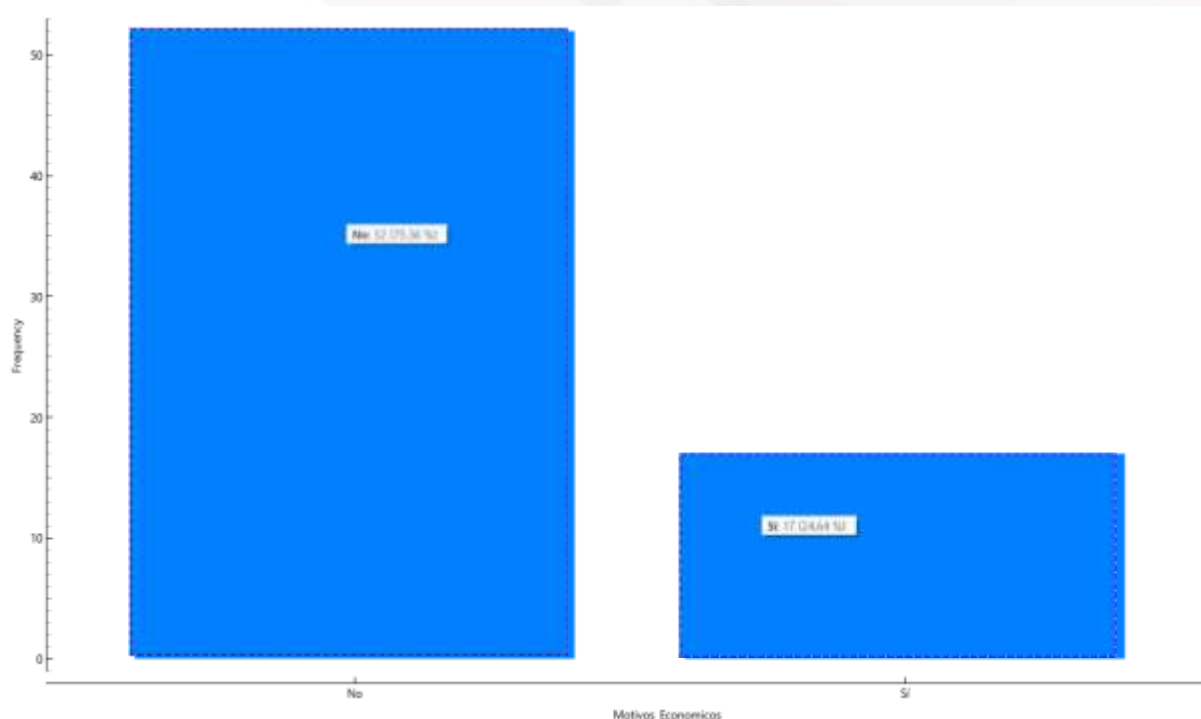


Ilustración 21. Motivos económicos

Al señalar una categoría socioeconómica, el 24,64% de la población señaló que los factores económicos repercuten en el transcurso de su año lectivo, debido a la falta de recursos necesarios para culminar sus estudios. Por otro lado, cerca de un cuarto de la población no presenta una restricción económica para desenvolverse en la institución educativa. Desde este punto, los resultados evidencian que, aunque la mayoría de las personas no presentan dificultades económicas en el sistema educativo, existe una población significativa que reconoce dificultades relacionadas al presupuesto estudiantil (Mercado & De la Cruz, 2024). En este apartado, a pesar de que exista una división poblacional significativa, el algoritmo lo cataloga como un factor mediador en relación con el desempeño académico que incide directo al abandono escolar.

Tabla 34. Motivo de salud

Genero	Frecuencia (n)	Porcentaje (%)
Sí	43	62,32%
No	26	37,68%
Total	69	100,00%

Nota. Elaboración propia.

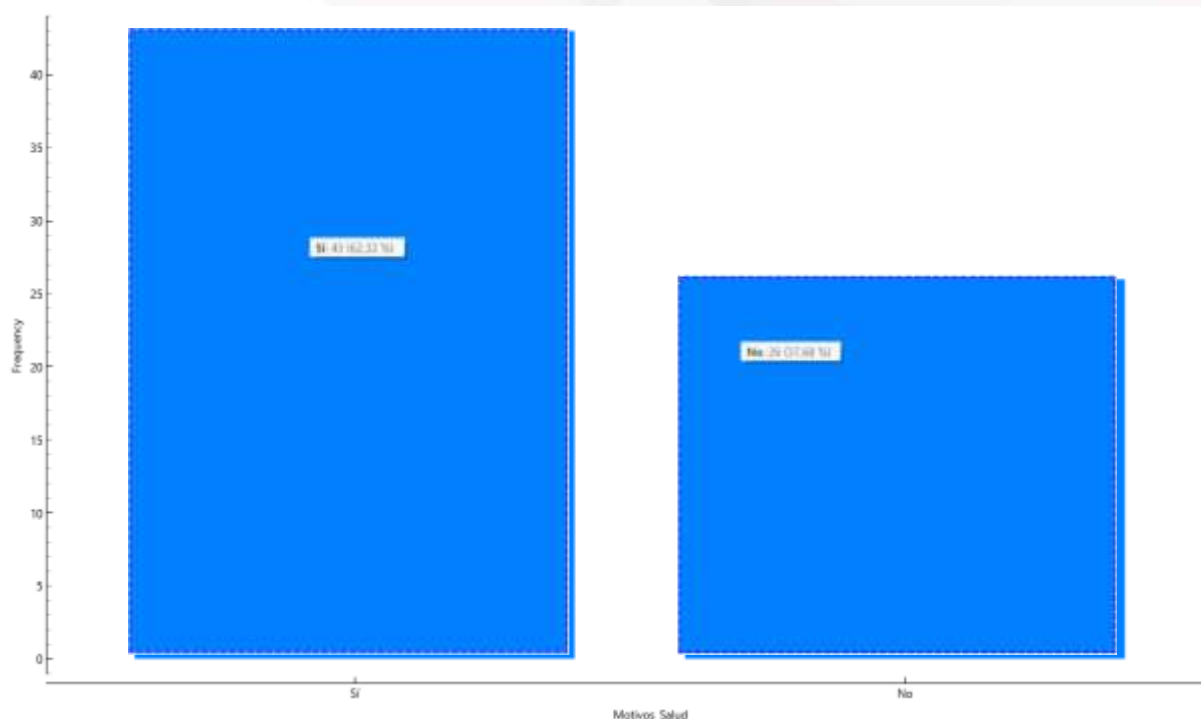


Ilustración 22. Motivo de salud

En la categoría de motivos de salud, la población se destacó con un porcentaje de 62,23% al indicar que su situación influye en el desempeño académico. En cambio, el 37,68% señaló que no tiene una afectación directa en su rendimiento académico. De esta manera, los resultados analizados evidencian que los factores de salud resaltan de una forma significativa dentro de la población estudiada. Por lo tanto, al reflejar más de la mitad de los estudiantes en el concepto de la salud, la existencia de condiciones psicosomáticas influye en el desempeño académico. A diferencia de otros rangos evaluados, los motivos de salud presentan una alta demanda en el ámbito educacional, lo que lo posiciona como un factor crítico en el análisis de riesgo (Gairín & Olmos, 2022). En conclusión, esta dimensión se configura en una variable contextual de manera directa con el abandono escolar deteriorando el desempeño académico.

Tabla 35. Responsabilidad familiar

Género	Frecuencia (n)	Porcentaje (%)
Sí	23	33,82%
No	45	66,18%
Total	68	100,00%

Nota. Elaboración propia.

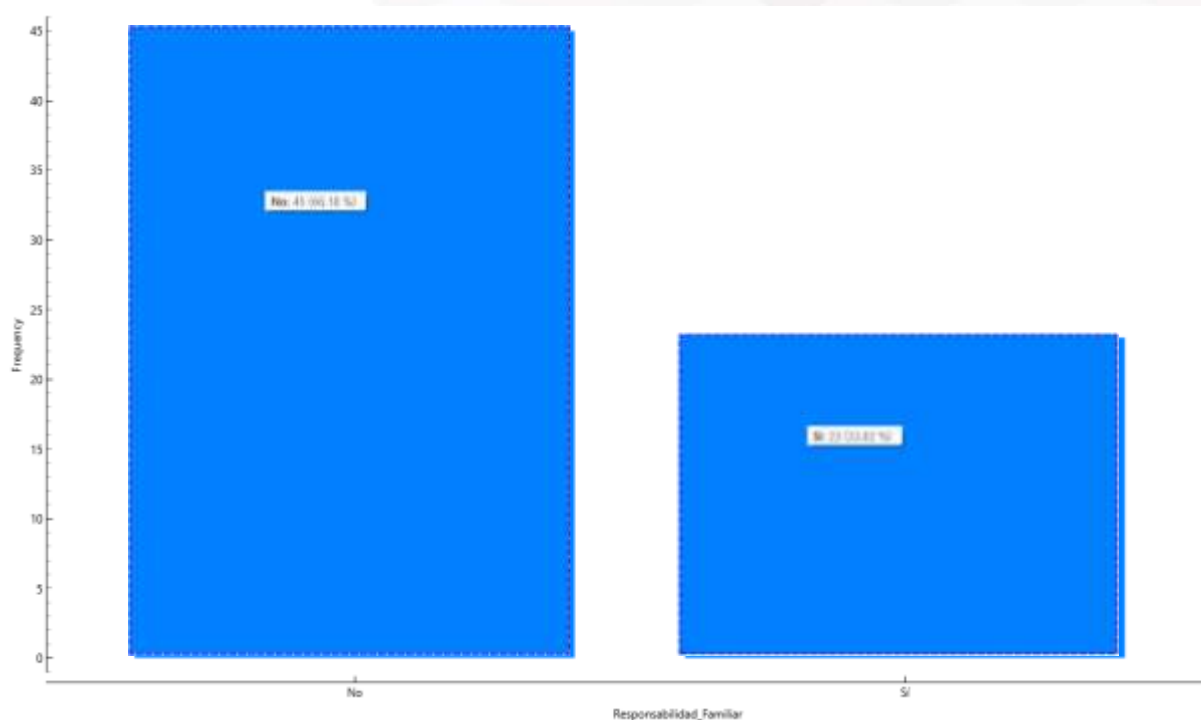


Ilustración 23. Responsabilidad familiar

En la última categoría está la responsabilidad familiar en el cual reflejó una población de 33,82% al tener un compromiso familiar que influye en la vida académica de cada estudiante. Por el contrario, el 66,18% de los estudiantes no enfrentan dificultades de responsabilidad lo que les permitió continuar con sus estudios sin tener que dividir su tiempo y energía en tareas de cuidado y sustento económico. Desde este punto de vista, los estudiantes que no muestran un mayor porcentaje de carga familiar, se diferencian del grupo que sí asume obligaciones domésticas y responsabilidad económica. Dado que, estas tareas influyen en el proceso de aprendizaje y enseñanza el tiempo de estudio se ve limitado al generar un nivel de estrés y una carga cognitiva en las actividades académicas (Sánchez & Dávila, 2022). Este índice incrementa cuando se combina con factores laborales y de salud que producen escenarios de mayor riesgo de exclusión educativa. En conclusión, la mayor parte de la población al no contar con una responsabilidad familiar, el grupo minoritario requiere de una atención de los factores asociados en el rendimiento académico.

4.2. Discusión de los Resultados/Triangulación (por cada ítem)

Distribución de estudiantes según el curso

Al tener un grado de 58,21% de la población de estudiantes en el tercer año lectivo, investigaciones resaltan en este periodo lectivo como uno de los casos más recurrentes de riesgo de abandono escolar debido a la carga académica junto a sus decisiones profesionales. Según el análisis realizado por Mello et al., (2025), detecta un mayor abandono académico en América Latina, donde la presión por graduarse se presenta como una carga multifactorial en la tolerancia hacia la carga académica. Por ello, al ingresar la información en el programa el algoritmo XGBoost lo clasifica en las variables «meta», en donde, su valor explicativo cobra un peso significativo en las intervenciones futuras del DECE. Así, el modelo predictivo detecta al grado educativo con un punto estratégico para que las acciones por parte del DECE sean efectivas en grupos con mayor probabilidad de vulnerabilidad.

Diversidad de género de los participantes

En cuanto al género de los estudiantes encuestados, la participación femenina resalta con participación de 55,88% por parte del grupo de análisis a estudiar. En un estudio realizado por Loaiza et al., (2023), señala que el abandono escolar es causado por la presión económica al buscar una inserción laboral temprana por parte del género masculino, mientras que el género femenino se vincula a factores de estrés debido a las demandas presentes en la institución. Por lo tanto, la diferencia de género no es un factor determinante en el abandono escolar, ya que, al existir un enfoque diferencial resulta el eje central para comprender los problemas de cada estudiante en el transcurso de su formación académica. En otras palabras, el género interviene en el estudio como una categoría de análisis para comprender la naturaleza de las causas multifactoriales del desenganche escolar. De esta manera, el algoritmo relaciona las interacciones de la diversidad de género con otros factores que le permitió clasificar los perfiles con mayor vulnerabilidad ajustada a su realidad multidimensional.

Comprensión de contenidos

Los participantes presentaron un 75,36% en la comprensión de contenidos, el cual evidencia un índice positivo en los estudiantes al comprender los contenidos impartidos por el educador dentro del proceso educativo. Este porcentaje dentro del modelo predictivo, lo posiciona como un factor protector en la claridad del aprendizaje cuya función impulsa la retención estudiantil para tener un proceso estable en su formación. En el estudio realizado por Tinto (1987, citado en Mendivil et al., (2025)) mantiene su postura en la integración académica como un factor establece al asegurar la continuidad del estudiante en el sistema educativo

eliminando el abandono. De esta manera, se logra comprender cuando el estudiante se adapta al estudio para entender los contenidos que el docente imparte consiguiendo acabar con el abandono escolar. Así, el programa Orange Data Mining consigue estabilizar el sistema educativo en las categorías con mayor presencia positiva para disminuir falsas alertas y optimizar la detección de riesgos reales dentro del modelo.

Cumplimiento de tareas

En el análisis realizado se evidenció que el 40,58% de los estudiantes presentan un grado de responsabilidad en el cumplimiento de tareas al generar un compromiso académico que disminuye la probabilidad del abandono escolar frente a variables externas. El resultado registrado, subraya que una parte de la población mantiene prácticas de estudios efectivas al cumplir con las tareas asignadas del docente, lo cual estabiliza el rendimiento en la preparación académica. Bajo una perspectiva multifacética acerca de la permanencia escolar, el desapego progresivo centrado en la carga académica se marca como una de las principales causas de la desconexión institucional (Suberviola, 2021). De esta manera, al tener un incumplimiento de tareas, ausentismo masivo y una baja participación su resultado se ve reflejado en el bajo rendimiento del estudiante al no tener una gestión en su compromiso conductual. Por otro lado, el modelo predictivo identifica al 2,9% de la muestra como el centro de un grupo vulnerable en el cual la institución debe diseñar intervenciones enfocadas a las realidades que presentan los estudiantes.

Rendimiento y esfuerzo

En este ítem más de la mitad de los estudiantes (46,38%) manifestaron un desempeño positivo en sus capacidades académicas al validar los procesos de enseñanza dentro de la institución y reducir el abandono escolar. Estos resultados apuntan al desarrollo de estrategias de autorregulación del aprendizaje que incrementan el sentido de pertenencia con el fin de potenciar su conocimiento para resolver problemas a futuro. En consecuencia, autores como Álvarez et al., (2025), afirman que la autoeficacia y la pertenencia son componentes importantes en el progreso educativo de cada estudiante al enfrentar la carga académica. De esta manera, disminuye la probabilidad de un abandono en el sistema educativo cuando el alumno asume los bajos resultados como desafíos de superación. Así mismo, el programa clasifica al ítem como un factor protector al fortalecer la permanencia escolar mediante estrategias de planificación y evaluación del conocimiento con el propósito de estabilizar el sistema educativo.

Carga académica

En cuanto a la carga académica existió una propagación en los niveles de presión más altos del 39,13% de los estudiantes, el cual indicó que una proporción de la muestra percibió una demanda académica exhaustiva. Las investigaciones enfocadas en la salud mental de los estudiantes respaldaron esta interpretación cuando el estrés académico supera las herramientas internas que un estudiante tiene para gestionar situaciones estresantes (Infantas & Heredia , 2023). En este sentido, al adquirir una presión constante en los alumnos, su resultado en clases se deriva en un agotamiento emocional que disminuye el rendimiento y provoca un efecto negativo en la continuidad escolar. De esta forma, el algoritmo del programa Orange Data Mining establece a esta variable como un factor crítico que debilitó el progreso estudiantil al interactuar con otros índices de riesgo. Al tener un incremento de estudiantes, el algoritmo del modelo predictivo activa alertas tempranas en la variable, siempre y cuando se adopten estrategias de acompañamiento socioemocional.

Apoyo del docente

En el rol del docente como guía de apoyo, se presentó una participación positiva de 65,22% por parte de estudiantes que consideran al educador como un elemento clave en la experiencia educativa. Estos datos transforman al docente en uno de los principales factores protectores de la educación frente a las necesidades específicas de cada alumno para garantizar una inclusión educativa. Al tener una comunicación efectiva del docente hacia el estudiante, el vínculo pedagógico favorece la inclusión escolar con el fin de crear un entorno de confianza emocional y de protección en el entorno. En la investigación propuesta por Tinto (1987, citado en Mendivil et al., (2025)), menciona que la calidad académica se rige a las interacciones del docente al estudiante en la toma de decisiones sobre su formación académica para evitar que el alumno se sienta excluido de la comunidad estudiantil. Por esta razón, el estudiante al recibir un apoyo continuo en el transcurso de su formación, aumenta el sentido de pertenencia en la institución lo cual le permitió al algoritmo operar esta variable como un factor protector.

Agotamiento emocional

Bajo el concepto del agotamiento emocional el 31,88% de los estudiantes poseen niveles críticos de fatiga mental que surgen a partir de la carga académica lo que dificulta la recuperación psicológica de esa población. Dado que este porcentaje tiene una relevancia significativa en relación a la presión escolar, su impacto es mayor debido a la dificultad de

estabilizar al estudiante dentro del sistema educativo. Según Tito et al., (2024), el bienestar estudiantil con la salud mental al ser un ciclo bidireccional presenta una consistencia en relación a la fatiga cognitiva, el cual impide que el estudiante continúe con su preparación académica y profesional. En otros términos, el estrés crónico y el desgaste ocupacional perjudican los procesos cognitivos dirigidos a la capacidad de enfoque de cada estudiante, lo que ocasiona que el aprendizaje se limite por estos factores. Al presentar un cansancio continuo el alumno se vio involucrado en un abandono silencioso el cual consiste en presentarse en clases, pero su nivel de participación y su capacidad de aprendizaje disminuye.

Niveles de ansiedad y estrés

Los estudiantes que padecen de niveles de ansiedad y estrés el algoritmo los identifica como un grupo minoritario que representa el 17,4%, cuyo impacto sigue siendo relevante dentro del análisis. Así, la existencia de la ansiedad en casi una quinta parte de la población requiere de la presencia de estrategias pedagógicas que interfieran de manera significativa el desempeño académico y cuidando su bienestar general. En base a lo propuesto por Thomas et al., (2024), la ansiedad escolar actúa como un factor inhibidor en los procesos cognitivos, que al impactar en el rendimiento académico del estudiante deteriora de forma progresiva la autoconfianza del estudiante. Por este motivo, la ansiedad al continuar en un margen crítico, el estudiante percibió su entorno educativo como una fuente de amenaza, en el cual su estabilidad emocional se deteriora optando como estrategia de alivio el abandono escolar.

Desmotivación académica

En la desmotivación académica la presencia del 17,91% de los estudiantes determina niveles críticos en la población, debido a la desconexión que marca el proceso educativo con el bienestar del alumno. Sin embargo, al no representar un grupo mayoritario de la muestra, este porcentaje es significativo para el algoritmo dado que se reporta el desinterés académico con el proceso formativo lo cual deriva al estudiante a optar por el abandono. Según el estudio de Mejía et al., (2026), evidenció que la ausencia de un proyecto de vida el estudiante no tiene una orientación sobre lo que quiere ser a futuro, lo que conlleva a debilitar su visión y aumenta el riesgo de abandono. En otras palabras, cuando el alumno no tiene un propósito a largo plazo y, al combinar factores de estrés, desmotivación y ansiedad, aumenta la probabilidad del abandono. Por ello, el modelo predictivo identificó la relación entre el desinterés y la intención de abandonar los estudios como una fuente de información que diferencia a los estudiantes con mayor y menor riesgo.

Capacidad para afrontar retos académicos

En la población encuestada sobre la capacidad de afrontar retos académicos se evidenció que más del 50% de los participantes poseen la capacidad de enfrentar las situaciones y problemas académicos y personales. Dado que, su porcentaje es alto frente a los recursos internos relacionados a la resiliencia estudiantil, funciona como blindaje frente a los factores externos que generó un gasto mental y emocional. En relación con la investigación realizada por (Suárez & Castro , 2022), los factores protectores a destacar en la resiliencia se convierten en un punto clave frente al contexto de vulnerabilidad. De esta manera, dicho ítem se configura como un factor de resiliencia dentro del grupo de indicadores evaluados, ya que cumple con la función de reducir el índice de abandono. Así, los estudiantes que tienen mayor percepción ante la presencia de obstáculos, no se les dificulta sobrellevar su conocimiento para enfrentar los desafíos y promover la motivación hacia sus compañeros con el fin de erradicar la incidencia del abandono.

Sentido de pertenencia institucional

En la población encuestada, el 56,52% de los estudiantes presentaron un índice alto de compromiso con la institución al sentir un vínculo afectivo que fortalece la permanencia escolar. Debido a un resultado positivo, los estudiantes al tener un ambiente sano lleno de valores con su grupo de estudio, se proyecta una estructura sólida en donde la educación es una herramienta útil y no un obstáculo en la enseñanza. De esta manera, el sentido de pertenencia el algoritmo lo codificó como un factor de barrera ante posibles riesgos de abandono. En un estudio realizado por Álvarez et al., (2025), al incorporar la inclusión social y académica en la formación del estudiante, impulsa la continuidad escolar al generar una serie de características, valores y metas en una comunidad sólida para el individuo. Por ende, al tener un vínculo institucional los niveles de abandono escolar disminuyeron debido a que este ítem funciona como protector ante el abandono al fortalecer la parte emocional y social de la institución.

Apoyo emocional por parte del docente

La población señala un respaldo emocional por parte del docente con un 60,87%, en el cual el educador al ser un transmisor de información se convierte en un referente afectivo en base al proceso de aprendizaje. Al tener este porcentaje, el rol del Departamento de Consejería Estudiantil (DECE) se refuerza a través de la interacción del estudiante con el profesor lo que reduce el riesgo del abandono escolar y generó una respuesta positiva en la permanencia. Por

ello, los docentes al disponer de un conjunto de conocimientos y habilidades les permitió identificar y entender las emociones que alberga cada estudiante para favorecer un ambiente sano y protector en la institución. Desde el punto de vista de Cayambe et al., (2024), el acompañamiento afectivo por parte del docente reduce el riesgo a la desvinculación institucional, ya que, en el sentido de pertenencia, el compromiso institucional se fortalece por las intervenciones que se realizan. En otras palabras, cuando el estudiante se ve respaldado por un adulto conocedor en las necesidades emocionales dentro de la escuela, disminuye la opción de aislamiento.

Relación con amigos en el colegio

Los estudiantes que presentan una excelente relación con sus amigos se visualizan en los resultados obtenidos con un peso de 43,48% en donde se evidenció que existe una red de apoyo entre pares. De esta forma, al desarrollar estrategias que vinculan la integración social con el bienestar institucional, transforma el fenómeno de estudio en un ecosistema de protección que impide el desinterés al compromiso académico y al abandono. En base a la investigación realizada por Vélez et al., (2020), al integrar estrategias sociales en la institución educativa, los estudiantes generaron amistades positivas que fortalecen factores académicos en el sentido de pertenencia, la motivación y la participación académica en el ámbito escolar. Por ello, si una institución carece de métodos o herramientas sociales, los alumnos tienden al aislamiento social lo que incide a una reacción en cada uno al afectar las emociones y bajar su rendimiento académico.

Faltas mensuales

En el indicador de faltas, los estudiantes que presentan ausentismo en la institución alcanzan una cifra de 15,94% que lo posiciona como un factor de alerta dentro del análisis, ya que, forma parte del conjunto de predictores en la desvinculación escolar. Esta tasa, indica que un grupo de la población encuestada se sitúa en niveles críticos por el aumento de faltas, ya que, existe una relación entre la continuidad académica y el abandono escolar. En base a lo investigado por Molina (2025), si un estudiante acumula el 20% de faltas injustificadas el sistema educativo lo deriva como pérdida de año, lo que debilita la percepción del alumno. Es decir, el alumno pierde esa visión de bienestar en la institución y comienza a apreciar el ambiente como un lugar de amenaza, lo que intensifica el deseo de abandonar su formación académica. De esta manera, al manejar variables cuantificables dentro del sistema Orange

Data Mining el algoritmo calcula las probabilidades de abandono y anticipa los posibles escenarios de estudiantes que alcancen su límite en base a las normas establecidas.

Importancia percibida del bachillerato

El análisis revela que un 73,91% de los encuestados considera importante la culminación de sus estudios, ya que potencia el bienestar psicológico y social para generar una visión clara en su proyecto de vida. Este índice es favorable en el ámbito educativo al aumentar las posibilidades de ingresar a la educación superior y a la adopción de un trabajo bien remunerado. Cabe señalar que, al tener un objetivo claro para culminar los estudios, el estudiante aumentó la motivación personal para resolver problemas emocionales y familiares que debilita la idea del abandono escolar (Gracia & Sánchez, 2020). Por lo tanto, si el estudiante mantiene su orientación en la culminación de sus estudios, aumenta su resistencia para enfrentar problemas externos e internos que dificultan su progreso educativo. Así, el modelo predictivo cataloga a esta variable como un protector, al reforzar la continuidad formativa del estudiante y evitar futuros riesgos de abandono.

Situación familiar percibida

La situación familiar dentro de los estudiantes se representa con un 20,59% al padecer problemas críticos dentro de su hogar, lo que impacta directamente en su rendimiento impidiendo la culminación de sus estudios. Al presentar este índice, las instituciones optan por fortalecer su entorno para los estudiantes que enfrentan un ambiente inestable en su hogar, ya que, este factor perjudica en su estado mental que lo desvincula de su preparación profesional. Desde la perspectiva de Concepción et al., (2024), las condiciones familiares debilitaron la adaptación biopsicosocial del estudiante, lo que provoca en el individuo un pensamiento negativo sobre sus metas y objetivos académicos para que se enfoque en actividades externas. En conclusión, esta variable forma parte del conjunto protector que el sistema detecta en el análisis ante el abandono escolar, debido a su porcentaje positivo adquirido por los estudiantes encuestados.

Motivos laborales

El 62,32% de las personas encuestadas, por motivos laborales tienden a dejar sus estudios por una presión económica, lo cual el modelo predictivo dentro de su sistema lo posiciona como un factor de riesgo. Dado que, el porcentaje tiene una tendencia de más de la mitad de la población, produce una fractura en el sistema educativo debido al aumento de abandono que

generó esta variable. En otras palabras, los estudiantes que tienen una tensión laboral, son más susceptibles al abandono, ya que se enfocan en la estabilidad económica de su familia dejando a un lado su formación profesional (Gairín & Olmos, 2022). Por otro lado, el ámbito laboral no siempre suele ser el principal indicador del abandono académico, pero una vez que genera ingresos, su presencia en las instalaciones educativas comienza a disminuir lo que provoca un aumento en el abandono escolar. Por ende, una vez que el sistema detecta alarma en estudiantes de riesgo, el DECE debe gestionar estrategias pedagógicas de acompañamiento para que los alumnos logren culminar sus estudios y aumenten su visión a largo plazo.

Motivos económicos

El 24,64% de los estudiantes representaron una cuarta parte de la muestra con respecto a las limitaciones en recursos económicos lo que afecta el desempeño escolar con respecto a la adquisición de materiales para mejorar el aprendizaje. Este porcentaje evidencia la presencia de una vulnerabilidad que trasciende en el ámbito pedagógico. Por ello, es evidente la desigualdad educativa en América Latina y el Caribe, ya que, se destaca los costos indirectos sobre el transporte, la conectividad y la alimentación de las cuales llegan a convertirse en barreras sin importar la motivación o el interés que presentan los estudiantes por salir adelante (Mercado & De la Cruz, 2024). Por otro lado, al tratar de cubrir los gastos generó un círculo cada vez más amplio de pobreza y el deterioro de la vinculación educativa es más rápida por causa de las limitaciones del capital de cada familia. Es por eso que, las estrategias académicas que emplean cada institución se basan en becas, apoyo financiero, entre otros que generaron condiciones adecuadas para evitar el abandono escolar.

Motivo de salud

En motivos de salud más de la mitad de los estudiantes representaron un 62,32%, al tener problemas físicos y mentales siendo variables indirectas en el desempeño académico, asistencia regular y estabilidad emocional. La estabilidad emocional (ansiedad, estrés, depresión, etc) es un factor estructural importante en la dinámica educativa con respecto a la concentración y motivación. En base a la investigación realizada por (Gairín & Olmos, 2022) indica que el malestar emocional colectivo en los adolescentes ha tenido un incremento radical ya que las modalidades híbridas tienen una continuidad con respecto a la asistencia. Este hecho ha provocado la integración de protocolos de salud integral en los planes del DECE creando así estrategias preventivas como el acompañamiento psicológico y personalizado. Sin embargo, dejar aún lado la importancia del bienestar físico da como

resultado una mejora en el rendimiento estudiantil a nivel cognitivo, emocional (reducción de cortisol y aumento de endorfinas) y social.

Responsabilidad familiar

Al obtener un 33,82% en este ítem, se evidenció que los estudiantes asumieron responsabilidades domésticas en las cuales se encuentran las actividades de cuidado (atender a niños pequeños, adultos mayores o personas con discapacidad). Esto implica que un tercio de la muestra enfrenta cargas adicionales en la distribución de tiempo disponible, siendo así un factor influyente en el rendimiento académico. Por lo cual, las actividades formativas como la lectura, la investigación y el estudio se encuentran afectadas al provocar una dificultad en la concentración y cansancio. Desde un punto de vista psicológico, la tensión emocional y mental, las mujeres adolescentes son más susceptibles a la presión familiar por ejercer tareas domésticas (Sánchez & Dávila, 2022). Al tener un “rol tradicional”, los factores vulnerables como lo económico crea la búsqueda de empleo inadecuado, mal remunerado e inseguro en su ambiente laboral, lo que provoca una barrera en la permanencia escolar.

CAPÍTULO V

5.1. Conclusiones

En función de los análisis realizados, se consiguió implementar el programa Orange Data Mining a través del algoritmo XGBoost, que se cataloga como una herramienta de alta calidad en la detección temprana de posibles riesgos de abandono. Por el contrario, la diferencia que existe entre los modelos descriptivos tradicionales, se destaca en identificar las relaciones entre dos o más variables en relación a la carga académica, las faltas mensuales y la motivación educativa. Este factor otorga la precisión predictiva del modelo para facilitar las estrategias con las intervenciones preventivas por parte del DECE.

Al señalar el abandono escolar como el punto de estudio, en los resultados obtenidos se evidencio que dicho fenómeno actúa con un proceso multicausal, ya que, al tener índices de 62,32% en los ítems motivos laborales y condiciones de salud se catalogan como los principales agentes externos ante el abandono escolar. De esta forma, se alcanzó a identificar los factores de riesgo que influyen en la continuidad formativa del estudiante al limitar su visión a largo plazo. Cabe mencionar que, el algoritmo XGBoost logró identificar al ítem de faltas mensuales como un factor de riesgo debido a la existencia de ausentismo de más de 6 días por parte de los estudiantes.

Al tener un rango de 58,21% de la muestra en el último nivel educativo, el riesgo de abandono escolar es crítico por la presencia de una sobre carga académica de 58,21% y de una desmotivación institucional de 31,88%. Sin embargo, el apoyo por parte del docente al tener una cifra de 65,22% y con un 60,87% en el apoyo emocional, representan un factor protector para determinar y desarrollar técnicas preventivas dentro del sistema educativo. Así, al clasificar variables en el rango meta, permite que el sistema tenga una transparencia para evitar que la IA sea una caja negra para el personal del DECE. Al aplicar esta técnica, la triangulación confirma que, al tener una interacción del docente con el alumno los factores de riesgos como la situación familiar y la ansiedad pierden validez en el sentido de pertenencia escolar.

Como resultado de esta investigación, se concluye que la integración de variables socioemocionales en herramientas de IA incrementa significativamente la precisión cuya viabilidad queda demostrada al identificar que un 58,21% de la muestra en el último nivel educativo se encuentra en una situación de vulnerabilidad.

Finalmente, la propuesta trasciende la detección al integrar estrategias de transición, clasificación de variables durante el proceso formativo, esta visión proactiva permitirá crear una “alarma inteligente” para las escuelas, en lugar de saturar o esperar a que un estudiante tenga el riesgo de abandono escolar. Este modelo desarrollado se fundamenta en la implementación de Orange Data Mining, esta elección técnica responde a la necesidad de contar con herramientas de alta precisión para una detección temprana de abandono escolar, superando así las limitaciones de los modelos estadísticos convencionales.

Las instituciones educativas al incluir estrategias de transición durante el proceso formativo, consiguen mitigar el impacto de riesgo en el sistema actual. Por esta razón, el sistema proporciona al personal educativo las posibilidades que tiene un grupo crítico de abandonar la educación con el fin de dotar al DECE información de los estudiantes para aplicar estrategias preventivas.

5.2. Recomendaciones

En la Unidad Educativa se recomienda integrar el modelo Orange Data Mining como una herramienta que funciona en la gestión institucional y potencia los recursos didácticos de vanguardia del DECE. Por ello, las inasistencias y el rendimiento son variables indispensables en el procesamiento de datos, debido a la necesidad de gestionar las alarmas ante la probabilidad de riesgo que el sistema detecte. Así mismo, estudiantes que presentaron un agotamiento emocional, se aconseja implementar cursos de inteligencia emocional y estrategias dirigidas al estrés académico para garantizar una armonía en el proceso de estudio. De esta forma, estos factores afectan al estudiante en su rendimiento académico, ya que, al usar técnicas que garanticen una reducción en la sobrecarga académica el resultado será positiva en la persistencia escolar.

Al reportar un índice de 62,32% en el ítem de motivos laborales, la institución debe estructurar mecanismos de flexibilidad pedagógica que integren tutorías asincrónicas y un ajuste en el tiempo de entrega en dirección a estudiantes de riesgo. Sin embargo, se sugiere fortalecer la inteligencia emocional del docente, ya que, este método otorga al educador herramientas técnicas que identifican de inmediato información silenciosa a través de tableros predictivos. Por otro lado, para disminuir la carga académica que presentan los estudiantes, se recomienda revisar la información curricular en los últimos niveles de educación superior, debido al porcentaje de 39,13% que presentó esta variable. En otros términos, al ser un factor de riesgo, se aconseja integrar una planificación transversal de evaluación que reduce el estrés del estudiante de manera significativa dentro del sistema.

Al usar este modelo predictivo se sugiere colocar las variables en tipo meta, para que el algoritmo los detecta como indicadores estratégicos de seguimiento, lo que garantiza una identificación temprana de posibles riesgos de abandono. De esta manera, las intervenciones realizadas por parte del DECE, se fortalecen en la toma de decisión y en la elaboración de estrategias sostenibles en el transcurso del año lectivo del estudiantado. Así, se transforma el entorno educativo en un ambiente sano y libres de amenaza que dificultan el proceso formativo y potencian la permanencia con éxito.

En síntesis, se propone diseñar una mesa técnica de seguimiento predictivo que este conformado por el personal del DECE y el grupo pedagógico, con el fin de traducir la información presentada en los tableros de Orange Data Mining. Por lo tanto, al elaborar planes de acción personalizados, la institución debe actuar de manera instantánea ante las

alertas emitidas por el sistema ajustando mecanismo de acompañamiento sobre los factores de riesgo que presenta cada estudiante. De esta manera, al transformar los datos en estrategias pedagógicas, el personal institucional fortalece la persistencia escolar mediante el uso de modelos vanguardistas que posiciona a la tecnología como eje hacia una educación inclusiva.

Referencias

- Aguilar, D., Ramírez, M., Gómez, H., Moscoso, O., & Rojas, J. (2025). Pobreza: Motor de la Deserción Escolar en la Zona Urbana de Pichincha, Ecuador. *GADE: Revista Científica*, 1(5), 876-891. <https://doi.org/https://doi.org/10.63549/rg.v5i1.473>
- Aguilera, J. (2022). *El desarrollo integral en el ámbito educativo y deportivo*. Editorial Latinoamericana de Educación Física.
- Álvarez, P., López, D., González, S., & Armas, A. (2025). Incidencia de las expectativas de autoeficacia y sentido de pertenencia en la intención de abandono académico del alumnado de bachillerato. *Incidencia de las expectativas de autoeficacia y sentido de pertenencia en la intención de abandono académico del alumnado de bachillerato*, 27(2), 132–148. <https://doi.org/https://doi.org/10.12795/revistafuentes.2025.25999>
- Asamblea Nacional Constituyente. (25 de 01 de 2008). *Constitución de la República del Ecuador*. https://www.defensa.gob.ec/wp-content/uploads/downloads/2021/02/Constitucion-de-la-Rpublica-del-Ecuador_act_ene-2021.pdf
- Asamblea Nacional del Ecuador. (14 de 07 de 2014). *Código de la Niñez y Adolescencia*. https://www.igualdad.gob.ec/wp-content/uploads/downloads/2017/11/codigo_ninezyadolescencia.pdf
- Asamblea Nacional del Ecuador. (25 de 08 de 2015). *Ley Organica de Educacion Intercultural*. https://educacion.gob.ec/wp-content/uploads/downloads/2017/02/Ley_Organica_de_Educacion_Intercultural_LOEI_codificado.pdf
- Asamblea Nacional del Ecuador. (26 de Mayo de 2021). *Ley Orgánica de Protección de Datos Personales*. Registro Oficial Suplemento 459 de 26-may.-2021. https://www.finanzaspopulares.gob.ec/wp-content/uploads/2021/07/ley_organica_de_proteccion_de_datos_personales.pdf
- Barra, E., & Pilicita, A. (2025). Sentiment Analysis With Transformers Applied to Education: Systematic Review. *International Journal of Interactive Multimedia and Artificial Intelligence*, 2(9), 72–83. <https://doi.org/https://doi.org/10.9781/ijimai.2025.02.008>
- Basantes, M., Miranda, A., Lara, E., Zamora, H., & Corozo, M. (2024). Desafíos y retos de la inteligencia artificial en la educación ecuatoriana: Una mirada desde la enseñanza y el rol del docente. *Revista Científica Internacional*, 11(1), 1551-1566. <https://doi.org/https://doi.org/10.69639/arandu.v12i1.694>
- Bolaño-García, M., & Duarte-Acosta, N. (2024). Una revisión sistemática del uso de la inteligencia artificial en la educación. *Revista Colombiana de Cirugía*, 39(1), 51-63. <https://doi.org/https://doi.org/10.30944/20117582.2365>
- Carrillo, R., & Gonzáles, W. (2023). *Estrategias pedagógicas implementadas por los docentes del grado noveno de la institución técnico nacional de comercio durante el aislamiento obligatorio por el Covid 19*. Universidad Francisco de Paula Santander. <https://doi.org/https://repositorio.ufps.edu.co/handle/ufps/8946>

- Castillo, E., & Santilla-Lima, J. (2023). Desafíos de la educación media: deserción escolar y sus Implicaciones. *Prometeo Conocimiento Científico*, 3(1), 41-51.
<https://doi.org/https://doi.org/10.55204/pcc.v3i1.e41>
- Cayambe , J., Sánchez , M., Núñez , I., Silva , I., & Sánchez , E. (2024). El docente como gestor de emociones en ambientes de aprendizajes. *Revista Social Fronteriza,,* 4(2), 1-25.
[https://doi.org/https://doi.org/10.59814/resofro.2024.4\(2\)234](https://doi.org/https://doi.org/10.59814/resofro.2024.4(2)234)
- Coghlan, S., & D'Alfonso, S. (2021). Digital phenotyping: an epistemic and methodological analysis. *Philosophy & Technology*, 34(4), 1905-1928.
<https://doi.org/https://doi.org/10.1007/s13347-021-00492-1>
- Comisión Económica Para América Latina - CEPAL. (13 de 08 de 2020). *La educación en tiempos de la pandemia de COVID-19*. Informe Especial de la CEPAL.
<https://www.cepal.org/es/publicaciones/45904-la-educacion-tiempos-la-pandemia-covid-19>
- Concepción, R., Nuñez, H., & Lopez, E. (2024). Reflexiones teóricas del fenómeno de la deserción académica en la educación superior. *Revista de Investigación y Evaluación Educativa*, 2(11), 88-109. <https://doi.org/https://doi.org/10.47554/revie.vol11.num2.2024.pp88-109>
- Delgado, M., & Riquelme, G. (2022). Evaluaciones de las competencias socioemocionales en el contexto educativo: una revisión. *Revista ConCiencia EPG*, 7(1), 43-74.
<https://doi.org/https://doi.org/10.32654/revistaconcienciaepg>
- Dewey, J., & Faerna, Á. (2022). *Lógica: la teoría de la investigación (1938)*.
- Eager, B., & Brunton, R. (2023). Prompting Higher Education Towards AI-Augmented Teaching and Learning Practice. *Journal of University Teaching and Learning Practice*, 20(5).
<https://doi.org/http://doi.org/10.53761/1.20.5.02>
- Frye, K., Boss, D., Anthony, C., Du, H., & Xing, W. (2024). Content analysis of the CASEL framework using K–12 state sel standards. *School Psychology Review*, 53(3), 208-222.
<https://doi.org/https://doi.org/10.1080/2372966X.2022.2030193>
- Gairín, J., & Olmos, P. (2022). *Disminuir el abandono escolar y mejorar la persistencia: causas, consecuencias y estrategias de intervención* (Vol. 177). Narcea Ediciones.
- García, A., Martínez, L., & Torres, P. (2022). Factores de la deserción escolar en el nivel medio: Un enfoque desde la resiliencia. *Revista de Investigación Pedagógica*, 3(14), 88-105.
- García, H., Hernández, L., Gámez, R., & Gaona, G. (2023). Perspectivas del uso de chatbots en la educación superior: Caso de estudio de la universidad tecnológica de Gutiérrez Zamora. *Revista Multidisciplinaria de Ciencia, Innovación y Desarrollo*, 32-37.
- Gobierno de la República del Ecuador. (1 de julio de 2024). *Informe exhaustivo nacional beijing+30*.
https://www.unwomen.org/sites/default/files/2024-09/b30_report_ecuador_es.pdf
- Golec, J., & Hachaj, T. (2025). Ten natural language processing tasks with generative artificial intelligence. *Applied Sciences*, 15(16), 9057-9072.
<https://doi.org/https://doi.org/10.3390/app15169057>

- Gonzales, V., Flores, M., Diaz, F., & Cruz, J. (2023). Effect of servant leadership on happiness at work of university teachers: The mediating role of emotional salary. *Problems and Perspectives in Management*, 21(2), 449-458. [https://doi.org/http://dx.doi.org/10.21511/ppm.21\(2\).2023.42](https://doi.org/http://dx.doi.org/10.21511/ppm.21(2).2023.42)
- Gracia, M., & Sánchez, A. (2020). La heterogeneidad del abandono educativo en las transiciones posobligatorias. Itinerarios y subjetividad de la experiencia escolar. *Revista de Sociología*, 105(2), 235-257. <https://doi.org/https://doi.org/10.5565/rev/papers.2775>
- Hernández, L. (2023). Inteligencia artificial generativa en la educación: retos éticos, potencial pedagógico y perspectivas futuras. *Innovarium International Journal*, 1(1), 1-12. <http://revinde.org/index.php/innovarium/article/view/1>
- Hilbert, S., Coors, S., Kraus, E., Bischl, B., Lindl, A., Frei, M., & Stachl, C. (2021). Machine learning for the educational sciences. *Review of Education*, 9(3). <https://doi.org/https://doi.org/10.1002/rev3.3310>
- INEC. (2025). *gob.ec*. https://www.ecuadorencifras.gob.ec/documentos/web-inec/EMPLEO/2025/Diciembre_2025/202512_Metodologia_ENEMDU.pdf
- Infantas, S., & Heredia, F. (2023). Estrés académico y el rendimiento académico de los estudiantes de un instituto de educación. *Revista Veritas ET SCIENTIA - UPT*, 12(1), 50-62. <https://doi.org/https://doi.org/10.47796/ves.v12i01.774>
- Infante, M., & Araya, S. (2020). *Enfoque interdisciplinario en la educación escolar*. Revista Interdisciplinaria de Educación.
- James, W. (2023). *Varietades de la experiencia religiosa: un estudio de la naturaleza humana*. Trotta.
- Jin, C. (2023). MOOC student dropout prediction model based on learning behavior features and parameter optimization. *Interactive Learning Environments*, 31(2), 714-732. <https://doi.org/https://doi.org/10.1080/10494820.2020.1802300>
- King, R., Wang, Y., Fu, L., & Leung, S. (2024). Identifying the top predictors of student well-being across cultures using machine learning and conventional statistics. *Scientific Reports*, 1(14), 8376-8386. <https://doi.org/https://doi.org/10.1038/s41598-024-55461-3>
- Liu, H., Peng, X., Song, C., Xu, C., & Zhang, M. (2022). Using AI Chatbots to Provide Self-Help Depression Interventions for University Students: A Randomized Trial of Effectiveness. *Internet Interventions*. <https://doi.org/https://doi.org/10.1016/j.invent.2022.100495>
- Liu, Z., Zhou, X., & Liu, Y. (2025). Predicción de la deserción estudiantil mediante aprendizaje conjunto con análisis de IA explicable basado en SHAP. *Journal of Social Systems and Policy Analysis*, 2(3), 111-132. <https://doi.org/https://doi.org/10.62762/JSSPA.2025.321501>
- Loaiza, D., García, I., Romero, J., Ronquillo, P., & Diaz, M. (2023). Identificación de los factores de la deserción académica en el sistema. *Ciencia Latina Revista Científica Multidisciplinar*, 7(2), 11121-11136. https://doi.org/https://doi.org/10.37811/cl_rcm.v7i2.6190

- López, P., Sánchez, D., & Lucio, M. (2025). Factores socioemocionales que influyen en la deserción escolar: un estudio en contextos de vulnerabilidad. *Revista Científico-Académica Multidisciplinaria*, 4(9), 1-19. <https://doi.org/https://doi.org/10.23857/pc.v10i4.9365>
- Maldonado, R., & Soledispa, S. (2024). Factores de deserción escolar de los estudiantes de la unidad educativa Leónidas Plaza km. 20. *Revista Científica Multidisciplinaria SAPIENTIAE*, 7(14), 175-185. <https://doi.org/https://doi.org/10.56124/sapientiae.v7i14.0012>
- Matbouli, Y., & Alghamdi, S. (2022). Modelos de regresión de aprendizaje automático estadístico para la predicción salarial con actividades y ocupaciones de toda la economía. *Información*, 13(10). <https://doi.org/https://doi.org/10.3390/info13100495>
- Mejía, E., Ferriol, F., & Reyes, G. (2026). Influencia de la motivación estudiantil en la deserción escolar en el nivel medio superior. *Ciencia Latina Revista Científica Multidisciplinaria*, 9(6), 10453-10471. https://doi.org/https://doi.org/10.37811/cl_rcm.v9i6.22
- Mello, J., Escobar, R., Segura, M., Giménez, S., Hernández, A., Pérez, P., & De la Iglesia, M. (2025). Predictores claves del abandono escolar en Paraguay: un análisis de datos masivos. *Revista De Educación*, 409, 243–272. <https://doi.org/https://doi.org/10.4438/1988-592X-RE-2025-409-686>
- Mendevíl, B., Arellano, A., Ríos, N., & Lizardi, M. (2025). Predicting student dropout from day one: XGBoost-Based early warning system using pre-enrollment data. *Applied Sciences*, 15(16). <https://doi.org/https://doi.org/10.3390/app15169202>
- Mercado, X., & De la Cruz, A. (2024). Abandono escolar un riesgo latente en la vida de los jóvenes. *Revista Con-Sciencias Sociales*, 16(31), 25–48. <https://doi.org/https://doi.org/10.35319/conciencias.202431155>
- Ministerio de Educación del Ecuador. (2023). *Guías informativas de riesgos psicosociales*. <https://educacion.gob.ec/wp-content/uploads/downloads/2024/04/Guias-informativas-de-riesgos-psicosociales.pdf>
- Ministerio de Educación del Ecuador. (2024). *Estadística educativa* (Vol. 5). Dirección Nacional de Comunicación Social. Estadística educativa. <https://educacion.gob.ec/wp-content/uploads/downloads/2024/10/Estadistica-Educativa-Vol-5.pdf>
- Ministerio de Educación del Ecuador. (2024). *Modelo de Gestión de los Departamentos de Consejería Estudiantil (DECE)*. <https://educacion.gob.ec/wp-content/uploads/downloads/2023/06/modelo-DECE.pdf>
- Ministerio de Educación del Ecuador. (Septiembre de 2024). *Plan Nacional "Cuidamos de ti"*. <https://educacion.gob.ec/wp-content/uploads/downloads/2024/10/Plan-nacional-cuidamos-de-ti.pdf>
- Molina, J. (2025). Continuidad pedagógica entre primaria y bachillerato: análisis de estrategias de enseñanza para reducir la deserción escolar. *Atlas Research Journal*, 3(1), 4-18. <https://doi.org/https://doi.org/10.65305/arj.v3n1.2025.25>

- Mota, M. (2022). La cientificidad de metodologías cuantitativa, cualitativa y emergentes. *Revista Digital de Investigación en Docencia Universitaria*, 16(1).
<https://doi.org/https://doi.org/10.19083/ridu.2022.1555>
- Nuñez, G. (2023). Deserción estudiantil en institutos superiores tecnológicos de Ecuador: Una revisión de la literatura. *Revista Latinoamericana Ogmios*, 3(8), 25-32.
[https://doi.org/Deserción estudiantil en Institutos Superiores Tecnológicos de Ecuador: Una revisión de la literatura](https://doi.org/Deserción%20estudiantil%20en%20Institutos%20Superiores%20Tecnol%C3%B3gicos%20de%20Ecuador:%20Una%20revisi%C3%B3n%20de%20la%20literatura)
- Pachay-López, M., & Rodríguez-Gámez, M. (2021). La deserción escolar: Una perspectiva compleja en tiempos de pandemia. *Revista Científico-Académica Multidisciplinaria*, 6(1).
<https://doi.org/https://doi.org/10.23857/pc.v6i1.2129>
- Pacheco, A., Boude, O., Chiappe, A., & Fontán, L. (2025). Análisis de aprendizaje impulsado por IA para el desarrollo metacognitivo y socioemocional: una revisión sistemática. *Frontiers education*, 10(1672901). <https://doi.org/doi:10.3389/feduc.2025.1672901>
- Parrales, J., & Cedeño, G. (2025). Análisis y motivación del reclutamiento de menores de edad por bandas delictivas en el Ecuador. *Revista Lex*, 8(28), 220–233.
<https://doi.org/https://doi.org/10.33996/revistalex.v8i28.280>
- Quiroz, M. (2025). Inteligencia artificial generativa (IA Gen) en la transformación digital de la educación superior una revisión sistemática de literatura. *Ciencia Latina Revista Científica Multidisciplinar*, 9(2), 6339-6378.
https://doi.org/https://doi.org/10.37811/cl_rcm.v9i2.17370
- Quituzaca, J., Armijos, V., Quiroz, S., Carchi, M., Calderón, A., & Calderón, L. (2025). Uso de modelos de inteligencia artificial para la predicción del abandono escolar en instituciones educativas. *Revista Latinoamericana De Calidad Educativa*, 2(4), 318-324.
<https://doi.org/https://doi.org/10.70625/rfce/346>
- Salazar, I., Mantilla, C., Ramírez, J., Zambrano, J., & Suárez, M. (2023). Validación de un método de selección de participantes dentro de un diseño de investigación mixta secuencial para estudios de caso de cadena de suministro sostenible. *Cuadernos de Administración*, 36, 1-20.
<https://doi.org/10.11144/Javeriana.cao36.vpsmw>
- Sánchez, E., & Dávila, O. (2022). Apoyo emocional de la familia y éxito escolar en los estudiantes de educación básica. *Revista Estudios Psicológicos*, 2(1), 7-29.
<https://doi.org/https://doi.org/10.35622/j.rep.2022.01.001>
- Soriano-Sánchez, J., & Jiménez-Vázquez, D. (2023). Benefits of emotional intelligence in school adolescents: A Systematic Review. *Revista de Psicología y Educación*, 2(18), 83-95.
<https://doi.org/https://doi.org/10.23923/rpye2023.02.237>
- Suárez, X., & Castro, N. (2022). Competencias socioemocionales y resiliencia de estudiantes de escuelas vulnerables y su relación con el rendimiento académico. *Revista De Psicología*, 40(2), 879–904. <https://doi.org/https://doi.org/10.18800/psico.202202.009>
- Suberviola, I. (2021). Análisis de los factores predictivos del abandono escolar temprano. *Vivat Academia* (154), 25-52. <https://doi.org/https://doi.org/10.15178/va.2021.154.e1373>

- Thomas, T., Kamath, N., & Joseph, G. (2024). Eficacia de la meditación en el aumento de la autoconfianza, y de la reducción del estrés y la ansiedad asociados a los exámenes escolares en estudiantes de la India. *Social Medicine*, 17(3), 166-175.
<https://doi.org/https://www.medicinasocial.info/index.php/medicinasocial/article/view/1853>
- Tito, G., Linares, I., & Bermúdez, N. (2024). Educación emocional y bienestar escolar: estrategias innovadoras para afrontar la crisis de salud mental estudiantil. *Revista Multidisciplinar Ciencia Y Descubrimiento*, 2(1). <https://doi.org/https://doi.org/10.63816/snweeh47>
- Tomalá, F., Tomalá, W., Gonzabay, J., & Freire, E. (2025). *La falta de coercibilidad respecto al uso de niños y adolescentes en temas de delincuencia organizada en Ecuador* (Vol. 6). Ciencia Y Educación. <https://doi.org/https://doi.org/10.5281/zenodo.15615336>
- Tomaszewski, R. (2021). A study of citations to stem databases: acm digital library, engineering village, ieeexplore, and mathscinet. *Scientometrics*, 126, 1797–1811.
<https://doi.org/https://doi.org/10.1007/s11192-020-03795-w>
- UNESCO. (7 de 09 de 2023). *UNESCO - Guidance for generative AI in education and research*.
<https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>
- UNICEF. (04 de 05 de 2024).
<https://www.unicef.org/chile/historias/aprendizaje-digital-los-desaf%C3%ADos-para-el-sistema-educativo>
- Vélez-Miranda, M., San Andrés-Laz, E., & Pazmiño-Campuzano, M. (2020). Inclusión y su importancia en las instituciones educativas desde los mecanismos de integración del alumnado. *Revista Arbitrada Interdisciplinaria Koinonía*, 5(9), 5-27.
<https://doi.org/https://doi.org/10.35381/r.k.v5i9.554>
- Ventura, J., & Peña, B. (2021). El mundo no debería girar alrededor del alfa de Cronbach \geq 70. *Adicciones*, 33(4), 369-372. <https://doi.org/https://doi.org/10.20882/adicciones.1576>
- Villarroel, S., Castillo, L., Granda, D., & Lema, L. (2025). Desarrollo de un sistema educativo inteligente en Python para la personalización del aprendizaje mediante técnicas de inteligencia artificial como Machine Learning y procesamiento de lenguaje natural. *Revista Científico-Académica Multidisciplinaria*, 10(11), 1560-1573.
<https://doi.org/10.23857/pc.v10i11.10701>
- Zabala, J. (2025). NUMELA: Un modelo predictivo para prevenir el abandono escolar mediante Inteligencia Artificial y Neuromarketing. *Revista De Investigación E Innovación Educativa*, 3(1), 38-46. <https://doi.org/https://doi.org/10.59721/rinve.v3i1.32>
- Zamora, M., Moreira, E., Espinoza, K., Espinoza, K., & Daza, M. (2023). Análisis de los factores socioemocionales incidentes en la tasa de deserción estudiantil de nivelación-ESPAM: Analysis of the socioemotional factors incident in the student dropout rate of nivelación-ESPAM. *Maestro Y Sociedad*, 103–112.
<https://maestrosociedad.uo.edu.cu/index.php/MyS/article/view/6029>

Zerkouk, M., Mihoubi, M., Chikhaoui, B., & Wang, S. (2024). A machine learning based model for student's dropout prediction in online training. *Education and Information Technologies*, 29(12), 15793-15812. <https://doi.org/https://doi.org/10.1007/s10639-024-12500-w>

Zhou, J., & Ye, J. (2023). Sentiment analysis in education research: a review of journal publications. *Interactive learning environments*, 31(3), 1252-1264. <https://doi.org/https://doi.org/10.1080/10494820.2020.1826985>

Anexo

Anexo 1: Formato de Validación de Instrumentos de Recolección de Datos

Título de la investigación: Modelo Predictivo de IA Enfocado en la Detección Temprana del Abandono Escolar, Vinculando la Variable Socioemocional en estudiantes de Bachillerato en la Educación Ecuatoriana.

Objetivo General de la investigación: Desarrollar un modelo de Inteligencia Artificial para predecir de manera temprana el riesgo de abandono escolar en bachillerato, integrando variables socioemocionales, con el fin de fundamentar estrategias de intervención mediante Gemini y Orange.

Título del instrumento a validar: Encuesta

Encuesta socioemocional y académica para la detección temprana del riesgo de deserción escolar

Esta encuesta forma parte de un proyecto de investigación académica. Sus respuestas son anónimas y se utilizarán únicamente con fines educativos y de investigación para identificar factores asociados al riesgo de deserción escolar. No existen respuestas correctas o incorrectas.

Investigadora: Vanesa Sagñay

Año: 2025

CONSENTIMIENTO INFORMADO

La presente encuesta tiene fines académicos. La información recopilada será anónima y confidencial, utilizada únicamente para el desarrollo de un proyecto de investigación de maestría. La participación es voluntaria y puede retirarse en cualquier momento sin consecuencia alguna.

✓ **Casilla:**

Acepto participar voluntariamente en esta investigación.

Instrucciones sugeridas:

Lea cuidadosamente cada afirmación y seleccione la opción que mejor represente su situación actual. No existen respuestas correctas o incorrectas. Responda con sinceridad.

SECCIÓN 1. DATOS GENERALES

CÓDIGO DEL ESTUDIANTE

LA INICIAL DE SU NOMBRE Y UN NUMERO - EJEMPLO: V08

Edad: _____

Curso/Año de Bachillerato:

- 1. °BGU
- 2. °BGU
- 3. °BGU

Género

- Femenino
- Masculino
- Prefiero no decirlo
- Otro

SECCIÓN 2. DIMENSIÓN ACADÉMICA

Escala de respuesta:

1 = Totalmente en desacuerdo

2 = En desacuerdo

3 = Ni de acuerdo ni en desacuerdo

4 = De acuerdo

5 = Totalmente de acuerdo

Comprendo los contenidos que se imparten en la mayoría de mis clases.

- 1
- 2
- 3
- 4
- 5

Puedo cumplir con las tareas y trabajos académicos a tiempo.

- 1
- 2
- 3
- 4
- 5

Mi rendimiento académico refleja mi esfuerzo

- 1
- 2
- 3
- 4

5

Siento que la carga académica es difícil de manejar.

1

2

3

4

5

Recibo apoyo de mis docentes cuando tengo dificultades académicas.

1

2

3

4

5

SECCIÓN 3. ASISTENCIA Y VINCULACIÓN ESCOLAR

En el último mes, ¿cuántos días faltaste a clases?

1-2 días

3-5 días

Más de 5 días

Ninguno

He pensado en dejar el colegio en algún momento del presente año lectivo.

Sí

No

A veces

SECCIÓN 4. DIMENSIÓN SOCIOEMOCIONAL

Escala de respuesta:

1 = Totalmente en desacuerdo

2 = En desacuerdo

3 = Ni de acuerdo ni en desacuerdo

4 = De acuerdo

5 = Totalmente de acuerdo

Me siento emocionalmente agotado(a) por mis estudios.

1

2

3

4

5

Siento ansiedad o estrés de forma frecuente.

- 1
- 2
- 3
- 4
- 5

Me siento desmotivado(a) para asistir a clases.

- 1
- 2
- 3
- 4
- 5

Me siento capaz de afrontar los retos académicos.

- 1
- 2
- 3
- 4
- 5

Me siento parte de mi institución educativa.

- 1
- 2
- 3
- 4
- 5

Siento apoyo emocional por parte de mis docentes.

- 1
- 2
- 3
- 4
- 5

Tengo amigos o compañeros con quienes puedo contar en el colegio.

- 1
- 2
- 3
- 4
- 5

SECCIÓN 5. PROYECTO DE VIDA Y MOTIVACIÓN

Considero importante terminar el bachillerato para mi futuro.

- 1
- 2
- 3
- 4
- 5

La situación personal o familiar afecta mi motivación para continuar estudiando.

- 1
- 2
- 3
- 4
- 5

SECCIÓN 6. FACTORES DE RIESGO (SÍ / NO)

He faltado a clases por motivos laborales.

SÍ

NO

He faltado a clases por motivos económicos.

SÍ

NO

He faltado a clases por motivos de salud.

SÍ

NO

Tengo responsabilidades familiares que dificultan mi asistencia regular.

SÍ

NO

Nombre del Evaluador: Lic. Vanesa Sagñay

ID: 1752702678

Profesión: Licenciada en Pedagogía



Handwritten signature of Juan Maposita, with the identification number 171730444-6 written below it.



Handwritten signature of Abigail Basantes.

Firma del Experto:

Máster en Neurociencia Juan Maposita

Firma del Experto:

Psicopedagoga Lic. Abigail Basantes