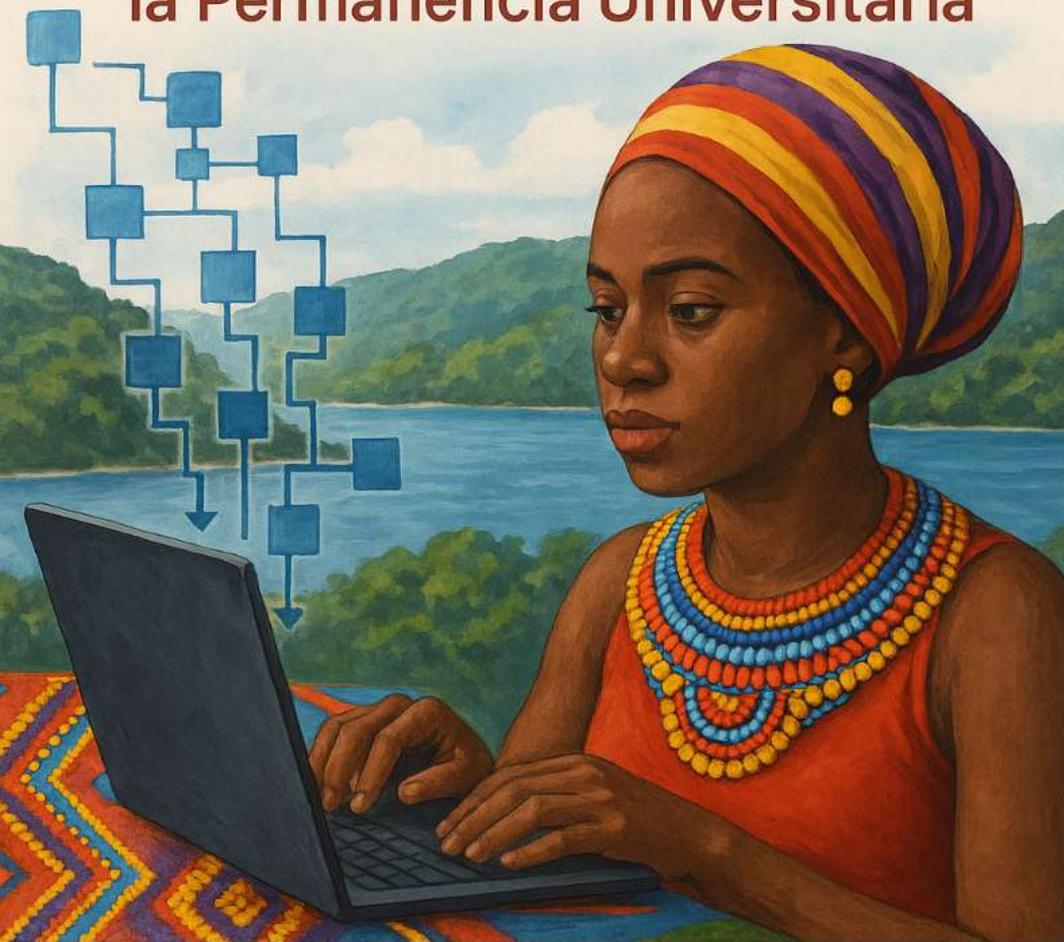


# Algoritmos contra la Deserción

Ciencia de Datos para la Permanencia Universitaria



Manuel Alberto González González





**Título: Algoritmos contra la Deserción: Ciencia de Datos para la Permanencia Universitaria**

**Autor: Manuel Alberto González González**

**Edición: Leonardo Valencia Echeverry**

**Diagramación: Leonardo Valencia Echeverry**

© MANUEL ALBERTO GONZÁLEZ GONZÁLEZ

© LIBROS PARA PENSAR

Primera Edición 2025  
ISBN: 978-628-01-8692-4

No se permite la reproducción total o parcial de este libro, ni su incorporación a un sistema informático, ni su transmisión en cualquier medio, sea este electrónico, mecánico, por fotocopia u otro método, sin el permiso previo y por escrito del autor.

Hecho en Colombia  
Printed in Colombia

Queda hecho el Depósito Legal





**ALGORITMOS CONTRA LA  
DESERCIÓN:  
CIENCIA DE DATOS PARA LA  
PERMANENCIA UNIVERSITARIA**



# Tabla de contenido

<b>INTRODUCCIÓN</b> .....	<b>11</b>
<b>JUSTIFICACIÓN DEL LIBRO</b> .....	<b>15</b>
IMPORTANCIA PARA LOS ESTUDIANTES DE INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD DEL PACÍFICO .....	16
OBJETIVOS DEL LIBRO Y PERFIL DE LECTORES.....	18
<b>PLANTEAMIENTO DEL PROBLEMA</b> .....	<b>21</b>
¿POR QUÉ UN ENFOQUE BASADO EN CIENCIA DE DATOS ES PERTINENTE HOY?.....	22
<b>METODOLOGÍA</b> .....	<b>25</b>
<b>CAPÍTULO 1. DESERCIÓN UNIVERSITARIA: UNA URGENCIA SOCIAL</b> .....	<b>29</b>
LA DESERCIÓN ESCOLAR EN COLOMBIA Y AMÉRICA LATINA .....	31
IMPACTOS SOCIALES, ECONÓMICOS E INSTITUCIONALES.....	35
LIMITACIONES DE LOS ENFOQUES TRADICIONALES .....	38
LA URGENCIA DE ACTUAR.....	41
<b>CAPÍTULO 2. DATOS PARA NO DESERTAR: EL POTENCIAL DE LA ANALÍTICA PREDICTIVA</b> .....	<b>45</b>
FUNDAMENTOS DE LA MINERÍA DE DATOS Y EL APRENDIZAJE AUTOMÁTICO.....	47
CASOS INTERNACIONALES DE USO DE ALGORITMOS EN EDUCACIÓN .....	51
ÉTICA Y RIESGOS DEL USO DE DATOS EN EDUCACIÓN .....	55
<b>CAPÍTULO 3. EL CASO DE LA UNIVERSIDAD DEL PACÍFICO</b> .....	<b>59</b>
CONTEXTO INSTITUCIONAL Y TERRITORIAL.....	59
RECOLECCIÓN Y DEPURACIÓN DE DATOS ESTUDIANTILES.....	61
DIAGNÓSTICO INICIAL DE LA DESERCIÓN.....	64

<b>CAPÍTULO 4. DISEÑO DEL MODELO PREDICTIVO.....</b>	<b>67</b>
ELECCIÓN METODOLÓGICA: KDD Y APRENDIZAJE SUPERVISADO .....	67
CONSTRUCCIÓN DEL DATASET: VARIABLES, PREPROCESAMIENTO Y BALANCEO.....	70
ALGORITMOS UTILIZADOS: RANDOM FOREST, SVM, ÁRBOLES DE DECISIÓN, ETC. ....	73
<b>CAPÍTULO 5. EVALUACIÓN DE MODELOS Y RESULTADOS.....</b>	<b>77</b>
COMPARACIÓN DE MÉTRICAS: PRECISIÓN, EXACTITUD, F1, ROC .....	77
SELECCIÓN DEL MEJOR MODELO .....	79
IMPLICACIONES PRÁCTICAS PARA LA GESTIÓN UNIVERSITARIA..	82
<b>CAPÍTULO 6. ARQUITECTURA DEL SISTEMA DE ALERTAS TEMPRANAS.....</b>	<b>85</b>
DISEÑO DE LA ARQUITECTURA CON MODELO C4 .....	85
INTEROPERABILIDAD CON ACADEMUSOFT Y ESCENARIOS DE CALIDAD .....	87
VISUALIZACIÓN DE RESULTADOS Y RECOMENDACIONES INSTITUCIONALES.....	90
<b>CAPÍTULO 7. RECOMENDACIONES, DESAFÍOS Y APRENDIZAJES.....</b>	<b>93</b>
QUÉ DEBE MEJORAR LA UNIVERSIDAD DEL PACÍFICO .....	93
DATOS QUE AÚN NO SE RECOLECTAN .....	95
<b>CONCLUSIÓN .....</b>	<b>101</b>
EL ALGORITMO COMO HERRAMIENTA, NO COMO ORÁCULO.....	101
LA PERMANENCIA COMO UN COMPROMISO INSTITUCIONAL Y SOCIAL .....	102
PRÓXIMOS PASOS: POLÍTICA PÚBLICA BASADA EN EVIDENCIA.	104
<b>BIBLIOGRAFÍA .....</b>	<b>109</b>

**SOBRE EL AUTOR.....115**



## Introducción

Cada año, miles de jóvenes en Colombia y América Latina inician con ilusión sus estudios universitarios. Lo hacen movidos por la promesa de un futuro mejor, de movilidad social y de realización personal y colectiva. Sin embargo, para muchos, ese camino se interrumpe antes de llegar a la meta: la deserción universitaria sigue siendo uno de los grandes fracasos silenciosos de nuestros sistemas educativos. No es solo un indicador de gestión institucional: es un síntoma profundo de las desigualdades estructurales que aún atraviesan nuestras sociedades.

En Colombia, cerca del 50 % de los estudiantes que ingresan a la educación superior no culminan sus estudios. Esta cifra, aunque alarmante, no alcanza a expresar la magnitud del problema en términos humanos, sociales y económicos.

Cada joven que abandona representa una trayectoria interrumpida, una familia frustrada y una comunidad que pierde una oportunidad de transformación. La deserción no ocurre en el vacío: tiene raíces en la pobreza, la exclusión territorial, la baja calidad educativa, la informalidad laboral, la violencia, y la falta de redes institucionales de apoyo.

Este libro nace de una convicción y de una urgencia. La convicción de que la ciencia de

datos, aplicada con sentido ético y político, puede contribuir significativamente a revertir este panorama. Y la urgencia de actuar ante una problemática que sigue dejando atrás, año tras año, a miles de jóvenes, especialmente en los territorios históricamente marginados.

La Universidad del Pacífico, ubicada en Buenaventura —una ciudad con profundas tensiones sociales y con una riqueza cultural invaluable— se convierte aquí en el escenario de una experiencia pionera: la construcción de un modelo predictivo para prevenir la deserción, a partir de datos reales, contextualizados y con enfoque humanista.

Por eso aquí no nos limitamos a presentar una herramienta tecnológica. Es una invitación a repensar la forma como entendemos la permanencia estudiantil, a cuestionar la fragmentación de las respuestas institucionales, y a proponer rutas de intervención integrales, basadas en evidencia, sensibilidad territorial y ética del cuidado.

A través de sus capítulos, se abordan desde los fundamentos de la minería de datos hasta los dilemas éticos de la inteligencia artificial en la educación; desde el diseño técnico del sistema de alertas hasta las recomendaciones para la política pública.

El presente texto, también interpela a las universidades públicas en su papel de garantes del derecho a la educación. Les recuerda que acompañar al estudiante no es una tarea secundaria, sino una función central. Que prever el abandono no es solo cuestión de eficiencia, sino de justicia. Que analizar los datos no es suficiente, si no se traduce en acción transformadora.

Porque detrás de cada predicción, hay una vida en juego. Y detrás de cada intervención oportuna, hay una oportunidad de futuro que no puede desperdiciarse.

Dirigido a universidades, gestores educativos, analistas de datos, investigadores sociales, responsables de política pública y ciudadanía interesada, este libro combina rigor técnico con una apuesta pedagógica y política. Muestra cómo los algoritmos, cuando se usan con propósito educativo y compromiso social, pueden convertirse en aliados para que ningún estudiante se quede atrás.

“Algoritmos contra la deserción” es, en definitiva, una propuesta para construir universidades más inteligentes, sí, pero sobre todo más humanas. Más capaces de ver a tiempo, de cuidar con datos, y de sostener sueños que no pueden seguirse quebrando por causas prevenibles.



## Justificación del libro

El libro *Algoritmos contra la Deserción: Ciencia de Datos para la Permanencia Universitaria* representa una contribución significativa a los retos contemporáneos que enfrentan las universidades públicas, en especial aquellas ubicadas en contextos de vulnerabilidad territorial como el litoral Pacífico colombiano.

Su propuesta no es solamente técnica, sino profundamente pedagógica, ética y territorial. A través de un enfoque interdisciplinario, este texto se propone transformar la forma en que las instituciones educativas comprenden, abordan y previenen la deserción estudiantil, integrando herramientas de ciencia de datos con una mirada humanista y situada.

Esta obra nace de una experiencia concreta en la Universidad del Pacífico y está escrita desde su propia realidad institucional. Esa característica le otorga una fuerza especial como referente para otras universidades públicas que enfrentan condiciones similares.

En lugar de ofrecer soluciones genéricas o descontextualizadas, este libro parte del diagnóstico profundo de las causas estructurales del abandono estudiantil y propone un modelo predictivo contextualizado, que puede ser replicado o adaptado en instituciones con limitaciones técnicas y económicas, pero con una alta vocación transformadora.

Su justificación se encuentra también en el vacío metodológico y estratégico que aún existe en muchas instituciones de educación superior frente al análisis proactivo de datos y la implementación de sistemas de alerta temprana.

Este libro propone una alternativa viable, ética y operativa, que puede ser integrada en los procesos de planeación académica y de bienestar estudiantil, fortaleciendo así la toma de decisiones basada en evidencia.

### **Importancia para los estudiantes de Ingeniería de Sistemas de la Universidad del Pacífico**

Para los estudiantes del programa de Ingeniería de Sistemas de la Universidad del

Pacífico, este libro tiene un valor formativo excepcional. En primer lugar, porque articula de manera clara y rigurosa los fundamentos de la minería de datos, el aprendizaje automático y la arquitectura de sistemas predictivos, desde un caso de estudio real en su propia universidad. No se trata de teoría abstracta, sino de conocimiento aplicado al territorio, construido con datos institucionales y desarrollado por un docente-investigador vinculado directamente a su contexto académico.

En segundo lugar, la obra ofrece a los futuros ingenieros una visión crítica y socialmente comprometida del papel que pueden cumplir las tecnologías emergentes en la transformación educativa. Les muestra que los algoritmos, cuando se diseñan con ética y con un profundo conocimiento de la realidad local, pueden convertirse en herramientas de justicia educativa y no solo en soluciones técnicas desprovistas de sensibilidad social. Así, se forma una conciencia profesional que va más allá de lo técnico: una ingeniería con propósito humano.

Además, este libro puede ser utilizado como material de apoyo en asignaturas como

Inteligencia Artificial, Minería de Datos, Bases de Datos, Sistemas de Información, Analítica de Datos o Proyectos de Grado. Su lenguaje, su estructura metodológica y su fundamentación empírica lo convierten en un recurso didáctico de alto valor, capaz de vincular la formación académica con la investigación aplicada y la intervención social.

Por último, esta publicación les brinda a los estudiantes un referente de cómo se puede producir conocimiento desde las regiones, con calidad, con pertinencia y con impacto. Les recuerda que, desde el Pacífico colombiano, también se puede innovar, investigar y proponer soluciones a los grandes desafíos del país y del mundo.

### **Objetivos del libro y perfil de lectores**

Este libro nace de una experiencia concreta: el diseño, desarrollo e implementación de un modelo predictivo de deserción en la Universidad del Pacífico, ubicada en Buenaventura, una de las regiones más excluidas del país.

A partir de esta experiencia, se propone un recorrido analítico y reflexivo que articula

elementos técnicos, institucionales y éticos para comprender y enfrentar el fenómeno de la deserción desde una perspectiva interdisciplinaria y territorial.

El objetivo principal de la obra es compartir conocimientos, metodologías y aprendizajes que puedan ser útiles para otras universidades públicas, responsables de políticas educativas, analistas de datos, investigadores sociales y organizaciones interesadas en promover la permanencia en educación superior desde enfoques basados en evidencia y sensibilidad social.

Los lectores encontrarán aquí un abordaje integral que combina fundamentos conceptuales, revisión de casos internacionales, aplicación de modelos de *machine learning*, arquitectura de sistemas de alertas, y propuestas de política pública.

Pero más allá de lo técnico, el libro invita a una reflexión profunda sobre la responsabilidad colectiva en garantizar el derecho a aprender, y sobre el potencial de la ciencia de datos como aliada de la equidad educativa en América Latina.



## Planteamiento del problema

La deserción en la educación superior es uno de los desafíos más persistentes y complejos de los sistemas educativos en América Latina. En Colombia, aproximadamente la mitad de los estudiantes que ingresan a programas universitarios no culminan sus estudios, una situación que genera consecuencias negativas para el desarrollo individual, institucional y nacional (MEN, 2023).

Más allá de las cifras, cada caso de deserción representa una historia de frustración, una inversión truncada y una pérdida de oportunidades de movilidad social.

El fenómeno de la deserción se presenta con mayor intensidad en contextos de vulnerabilidad socioeconómica, territorial y étnica, donde los estudiantes enfrentan obstáculos estructurales como la pobreza, la baja calidad educativa en niveles previos, la falta de redes de apoyo y las brechas digitales.

En regiones como el Pacífico colombiano, estas barreras se entrelazan con el abandono estatal, la violencia y la desigualdad histórica, haciendo aún más urgente el diseño de estrategias efectivas y sostenidas para promover la permanencia.

## ¿Por qué un enfoque basado en ciencia de datos es pertinente hoy?

Durante décadas, los intentos por reducir la deserción universitaria se han enfocado en políticas reactivas, centradas en becas, tutorías o programas de bienestar que, si bien han tenido impactos positivos, no siempre logran anticipar el abandono ni actuar a tiempo.

Frente a esta limitación, la ciencia de datos emerge como una alternativa transformadora. Su capacidad para procesar grandes volúmenes de información, identificar patrones invisibles y generar predicciones confiables permite desarrollar modelos de alerta temprana que orientan decisiones más oportunas y focalizadas (Baker & Inventado, 2014).

El uso de técnicas como el aprendizaje automático y la minería de datos educativas ha demostrado ser eficaz para detectar trayectorias de riesgo y predecir comportamientos estudiantiles, no con el ánimo de estigmatizar, sino de intervenir preventivamente.

Cuando se utilizan con criterios éticos y pedagógicos, estos modelos pueden apoyar a las universidades en la implementación de políticas más inteligentes, sensibles y personalizadas (Siemens & Long, 2011).

En un contexto global donde los recursos son limitados y la presión por mejorar los indicadores de calidad es creciente, incorporar la analítica predictiva en la gestión académica representa una oportunidad estratégica. No se trata solo de innovar técnicamente, sino de transformar la cultura institucional hacia una lógica de prevención, equidad y justicia educativa.



## Metodología

La construcción de un modelo predictivo para la prevención de la deserción estudiantil en la Universidad del Pacífico se apoyó en un enfoque metodológico riguroso, estructurado y contextualizado.

Este capítulo detalla los principios, etapas y criterios que guiaron el diseño, desarrollo y validación del sistema de alerta temprana, articulando conocimientos de la minería de datos, el aprendizaje automático y la analítica educativa, bajo una perspectiva ética, pedagógica y territorial.

Enfoque general: KDD y minería de datos educativa El proceso metodológico se estructuró con base en el paradigma KDD (Knowledge Discovery in Databases), una metodología consolidada para la extracción de conocimiento a partir de grandes volúmenes de datos (Fayyad et al., 1996).

Esta aproximación fue complementada con técnicas de minería de datos educativa (EDM) y aprendizaje supervisado, considerando la naturaleza binaria del fenómeno a predecir (deserta/no deserta) y la disponibilidad de un conjunto de datos etiquetado.

El KDD incluye cinco etapas: (1) selección de los datos relevantes, (2) preprocesamiento y

limpieza, (3) transformación de variables, (4) minería de datos con algoritmos predictivos y (5) evaluación e interpretación de resultados. Esta secuencia fue implementada de forma iterativa, incorporando validaciones internas en cada paso y garantizando la trazabilidad del proceso.

Recolección y depuración de datos El conjunto de datos fue obtenido a partir de las bases institucionales de la Universidad del Pacífico, incluyendo información académica, financiera, sociodemográfica y administrativa de estudiantes de pregrado entre 2015 y 2020.

En total, se consolidaron más de 13.000 registros, que fueron sometidos a un proceso de limpieza que implicó la eliminación de duplicados, el tratamiento de valores nulos y la codificación de variables categóricas.

Se seleccionaron 19 variables relevantes, agrupadas en cuatro dimensiones: rendimiento académico, situación económica, perfil sociodemográfico y trayectoria administrativa. Estas variables fueron transformadas mediante normalización, discretización y codificación para su procesamiento algorítmico.

Selección y entrenamiento de modelos Se utilizaron diversos algoritmos de clasificación binaria, incluyendo árboles de decisión,

regresión logística, máquinas de vectores de soporte (SVM), Naive Bayes y Random Forest.

Cada modelo fue entrenado sobre un conjunto de datos particionado en 70 % entrenamiento y 30 % prueba, aplicando validación cruzada de 10 pliegues para garantizar la robustez estadística.

La técnica SMOTE (Chawla et al., 2002) fue utilizada para balancear las clases, dado el desequilibrio entre estudiantes desertores y no desertores. Esta técnica generó ejemplos sintéticos de la clase minoritaria, mejorando la sensibilidad del modelo sin sacrificar precisión global.

Evaluación de modelos y selección final Los modelos fueron comparados con base en métricas como precisión, sensibilidad, especificidad, puntuación F1 y AUC-ROC.

El modelo Random Forest fue seleccionado como el más robusto, al alcanzar una precisión del 86 %, sensibilidad del 84 % y un AUC de 0.91. Su capacidad para manejar relaciones no lineales, variables faltantes y datos ruidosos lo hicieron ideal para el contexto institucional.

Criterios éticos y participativos Desde el inicio del proceso, se adoptaron principios éticos en la recolección y uso de datos personales: anonimato, consentimiento

institucional, protección de datos y uso educativo no punitivo. Además, se promovieron espacios de socialización con actores institucionales para validar las variables empleadas, garantizar la interpretabilidad de los resultados y asegurar la apropiación del sistema por parte de los usuarios clave.

En síntesis, la metodología utilizada combina rigurosidad técnica, pertinencia institucional y sensibilidad territorial. No se trata de una receta replicable mecánicamente, sino de una guía flexible que puede ser adaptada a otros contextos universitarios que busquen prevenir la deserción mediante el uso responsable de la ciencia de datos.

## Capítulo 1. Deserción universitaria: una urgencia social

En Colombia, cada inicio de semestre universitario inaugura también una carrera silenciosa contra el abandono. Aunque miles de jóvenes acceden a la educación superior con expectativas legítimas de desarrollo personal y movilidad social, una parte considerable de ellos no llegará a cruzar la meta de la graduación.

Este fenómeno, conocido como deserción estudiantil, ha dejado de ser un mero indicador institucional para convertirse en un espejo de las inequidades estructurales que atraviesan el sistema educativo y la sociedad en su conjunto.

La deserción no ocurre en el vacío. Está atravesada por condiciones socioeconómicas adversas, déficit de capital cultural, trayectorias escolares precarias, y profundas brechas territoriales que afectan especialmente a poblaciones rurales, indígenas y afrodescendientes.

Según datos del Sistema para la Prevención de la Deserción en Instituciones de Educación Superior (SPADIES), alrededor del 47 % de los estudiantes colombianos abandona sus estudios antes de concluir el primer ciclo académico de su programa (MEN, 2023).

Esta cifra, por sí sola alarmante, se vuelve aún más crítica al observar su distribución geográfica: en regiones como el Pacífico, la cifra supera ampliamente el promedio nacional.

La deserción debe entenderse como el desenlace de un proceso en el que fallan múltiples dispositivos de apoyo. En muchos casos, el abandono no es una decisión voluntaria, sino una reacción forzada ante condiciones materiales de subsistencia.

Factores como la necesidad de trabajar, la inseguridad alimentaria, el bajo rendimiento académico sin acompañamiento adecuado, y la falta de redes institucionales de apoyo generan un entorno propenso al retiro (Tovar-García & García Muñoz, 2020).

Para jóvenes que provienen de contextos históricamente marginados, la universidad representa no solo un reto académico, sino una ruptura cultural, económica y emocional con sus entornos de origen.

Este panorama es especialmente visible en la Universidad del Pacífico, ubicada en Buenaventura, uno de los municipios con mayores índices de pobreza multidimensional del país (DANE, 2023). La universidad atiende mayoritariamente a estudiantes afrocolombianos, muchos de ellos provenientes de comunidades rurales ribereñas con

condiciones de acceso limitadas. Para estos estudiantes, ingresar a la universidad no es simplemente un paso educativo, sino un acto de resistencia y afirmación social. Sin embargo, la infraestructura institucional, las políticas de bienestar y el financiamiento estudiantil no siempre están a la altura de ese esfuerzo colectivo.

Como han señalado estudios recientes sobre permanencia estudiantil, la deserción en contextos como el del litoral Pacífico no puede abordarse desde una mirada homogeneizadora. Se requiere un enfoque territorializado, interseccional y con políticas afirmativas que reconozcan las particularidades socioculturales de los estudiantes (UNESCO IESALC, 2022).

En ese sentido, la Universidad del Pacífico no solo refleja las tensiones del sistema de educación superior colombiano, sino que condensa, en un solo territorio, los desafíos más urgentes para avanzar hacia una educación realmente equitativa.

## **La deserción escolar en Colombia y América Latina**

La deserción en la educación superior no es exclusiva de Colombia, pero sí alcanza niveles particularmente alarmantes. Se trata de un fenómeno complejo, multicausal y persistente,

que limita no solo el desarrollo individual de miles de jóvenes, sino también el potencial productivo y social de las naciones.

En Colombia, el Sistema para la Prevención de la Deserción en Instituciones de Educación Superior (SPADIES) reporta que aproximadamente el 47 % de los estudiantes universitarios abandonan sus programas antes de completar el ciclo académico, especialmente durante los primeros tres semestres (MEN, 2023).

Esta cifra refleja una pérdida profunda de capital humano y de inversión pública, y se traduce en frustración, exclusión laboral y desigualdad de largo plazo.

Las causas de este abandono son múltiples y se entrelazan con los determinantes sociales más persistentes del país: pobreza, desempleo juvenil, informalidad laboral, violencia estructural, falta de apoyo familiar, y un sistema educativo que muchas veces reproduce las desigualdades en lugar de mitigarlas (Tovar-García & García Muñoz, 2020).

Aunque existen esfuerzos institucionales para frenar esta tendencia —como los créditos condonables, las becas de sostenimiento, y algunos programas de tutoría—, la mayoría de las estrategias tienden a ser reactivas y no logran anticipar los momentos críticos en los

que los estudiantes empiezan a desconectarse del proceso formativo.

Este fenómeno se agrava significativamente en las regiones más apartadas del país. En el litoral Pacífico, por ejemplo, las tasas de deserción superan con creces el promedio nacional.

En el caso específico de la Universidad del Pacífico, según la investigación base de este libro, la tasa de deserción en el segundo semestre de 2020 alcanzó el 15,32 %, con una expresión especialmente aguda entre mujeres (85,4 % de los desertores) y estudiantes de estrato 1 (68,7 %).

Estos datos no solo revelan una brecha territorial, sino también una intersección entre género, pobreza y etnicidad, que exige respuestas más profundas y estructurales.

A nivel regional, América Latina comparte esta problemática, aunque con intensidades y respuestas diferenciadas. Un informe de la CEPAL (2020) estima que solo entre el 50 % y el 60 % de los estudiantes que ingresan a la educación superior en la región logran culminar sus estudios.

Las causas se repiten: condiciones económicas precarias, falta de orientación vocacional, debilidad en las competencias

básicas, escaso acompañamiento institucional, y currículos poco contextualizados. Esta realidad configura lo que algunos autores han denominado “trayectorias frágiles” en la educación superior (UNESCO IESALC, 2022).

Sin embargo, hay ejemplos destacables. En países como Uruguay y Chile, las políticas de acompañamiento personalizado, el fortalecimiento de los sistemas de alerta temprana y la gratuidad o subsidios parciales han demostrado impactos positivos en la retención estudiantil (Brunner & Ferrer, 2018).

Estos modelos combinan financiamiento con innovación pedagógica y apoyo psicosocial, lo cual permite una lectura más integral de las necesidades estudiantiles.

Por el contrario, en países como Bolivia, Paraguay o Perú, donde los sistemas universitarios son altamente segmentados, y el acceso está mediado por profundas desigualdades sociales, las tasas de abandono siguen siendo dramáticas. Las universidades públicas suelen estar desbordadas y las privadas carecen, en muchos casos, de mecanismos sólidos de apoyo estudiantil.

En síntesis, tanto en Colombia como en América Latina, la deserción universitaria es uno de los síntomas más visibles de una deuda histórica con la equidad educativa. Es una alerta

que nos obliga a repensar el sistema más allá de la cobertura, incorporando el principio de permanencia como parte esencial del derecho a la educación.

### **Impactos sociales, económicos e institucionales**

Detrás de cada caso de deserción universitaria hay una historia interrumpida, una biografía truncada por las condiciones estructurales del entorno. No se trata únicamente de un estudiante que se retira de las aulas, sino de una vida que se desvía de su horizonte de realización. Para miles de jóvenes en Colombia y América Latina, la universidad representa no solo una promesa de movilidad social, sino una ruptura simbólica con generaciones de exclusión. Cuando esa promesa se rompe, lo que queda es una profunda sensación de fracaso, frustración y pérdida de propósito.

Desde la perspectiva individual y familiar, los efectos de la deserción son múltiples. En el corto plazo, se manifiestan en la pérdida de autoestima, en el abandono de metas personales, y en la decepción familiar.

A largo plazo, se traducen en una inserción laboral más precaria, limitada a la informalidad, con ingresos inestables y menores posibilidades

de acceder a seguridad social. Según el DANE (2023), los jóvenes entre 18 y 24 años que no completan sus estudios superiores tienen una tasa de informalidad laboral del 70 % y perciben, en promedio, ingresos 40 % inferiores respecto a quienes logran graduarse.

Esta brecha refuerza los ciclos de pobreza y exclusión, particularmente en los sectores rurales, étnicos y de bajos ingresos.

En contextos como el de Buenaventura, donde opera la Universidad del Pacífico, estos impactos se profundizan. La deserción allí no es solo una estadística: es el reflejo de una lucha cotidiana por permanecer en un sistema que no siempre ofrece las condiciones mínimas para la continuidad educativa.

Abandonar los estudios, para muchos estudiantes afrodescendientes y rurales, significa volver a un entorno marcado por la incertidumbre económica, la violencia estructural y la falta de oportunidades reales (UNESCO IESALC, 2022).

A nivel institucional, las consecuencias también son significativas. La deserción afecta de manera directa los indicadores clave que determinan la calidad y la viabilidad de las universidades, como la tasa de retención, el promedio de duración de los estudios y la eficiencia terminal.

Pero también tiene un impacto financiero directo: menos estudiantes matriculados implica menor flujo de recursos por matrículas y menos posibilidades de acceder a financiamiento estatal condicionado por resultados. En el caso de universidades públicas regionales, que ya operan con presupuestos limitados, cada estudiante que abandona representa un desafío para la planificación institucional y la sostenibilidad operativa.

Además, la deserción tiene implicaciones en el tejido universitario. Dificulta la creación de comunidades académicas estables, fragmenta los procesos de formación colectiva y debilita la cultura institucional.

La universidad pues, se resiente cuando se ve forzada a formar estudiantes en tránsito permanente, sin posibilidad de construir vínculos de largo plazo entre docentes, estudiantes y saberes.

Desde una mirada nacional, la deserción universitaria representa un obstáculo para el desarrollo económico y social. La inversión pública que no se traduce en titulaciones completas es una pérdida de eficiencia y de equidad.

Pero más allá del balance contable, lo que está en juego es la posibilidad de construir una sociedad basada en el conocimiento, capaz de

garantizar igualdad de oportunidades reales. Reducir la deserción no es solo una meta educativa: es un compromiso ético con la justicia social.

## **Limitaciones de los enfoques tradicionales**

Frente al fenómeno persistente de la deserción universitaria, las respuestas institucionales han sido, en muchos casos, limitadas y fragmentarias. Aunque en Colombia existen políticas de fomento a la permanencia como las tutorías académicas, becas de sostenimiento, créditos condonables, programas de nivelación y acompañamiento psicosocial, su implementación ha tendido a ser reactiva, desarticulada y de corto alcance.

Estas iniciativas, a menudo bien intencionadas, suelen activarse cuando el estudiante ya ha tomado la decisión de abandonar, lo cual reduce drásticamente su eficacia (Salazar & Vélez, 2021).

Uno de los principales problemas radica en la falta de integración entre las distintas unidades institucionales. Las áreas académicas, administrativas y de bienestar no siempre comparten información de manera sistemática, lo que impide construir una visión holística de cada estudiante.

Las alertas se emiten tarde, cuando las señales de riesgo —como bajo rendimiento, ausencias reiteradas o dificultades financieras— ya se han acumulado sin intervención efectiva. Esta lógica reactiva responde más a una cultura institucional centrada en la corrección que en la prevención (Martínez, 2022).

En la Universidad del Pacífico, estas limitaciones se hacen aún más visibles. Según lo documentado en la investigación base de este libro, la institución no cuenta con un sistema de alertas tempranas ni con una base de datos consolidada que permita hacer seguimiento a las trayectorias estudiantiles.

El área de bienestar universitario carece de herramientas sistemáticas para cruzar variables académicas, socioeconómicas y psicológicas, lo que impide anticipar y atender situaciones de vulnerabilidad antes de que se conviertan en abandono definitivo (González González, 2021).

Esta situación no es exclusiva de una universidad. Es reflejo de una debilidad estructural en la forma en que muchas instituciones de educación superior en Colombia y América Latina comprenden y gestionan la permanencia.

En lugar de apostar por estrategias proactivas basadas en la evidencia, se sigue privilegiando una lógica de contención, con intervenciones

puntuales y escasamente evaluadas. A esto se suma la falta de inversión sostenida en tecnologías de la información y personal capacitado para la analítica institucional (Brunner & Ferrer, 2018).

Frente a este escenario, la incorporación de enfoques basados en ciencia de datos, minería de información y modelos predictivos aparece como una oportunidad disruptiva.

Estas herramientas permiten construir perfiles de riesgo personalizados, identificar patrones invisibles al análisis tradicional y emitir alertas tempranas que faciliten intervenciones oportunas. Como señalan De la Cruz y Pérez (2019), los algoritmos predictivos, cuando se diseñan con criterios éticos y pedagógicos, pueden convertirse en aliados fundamentales para democratizar el éxito académico.

Sin embargo, la adopción de estas tecnologías exige más que infraestructura técnica. Implica una transformación cultural en las universidades: una forma distinta de entender los datos, de trabajar colaborativamente entre áreas, y de asumir que la permanencia no es responsabilidad exclusiva del estudiante, sino un compromiso institucional.

En síntesis, superar las limitaciones actuales requiere pasar de una lógica correctiva a una

lógica predictiva, centrada en el cuidado y en la construcción de trayectorias educativas sostenibles.

### **La urgencia de actuar**

La deserción universitaria, especialmente en contextos históricamente marginados como el del Pacífico colombiano, no puede seguir tratándose como una estadística inevitable ni como una externalidad del sistema educativo.

Cada joven que se retira de las aulas sin culminar su formación representa no solo una pérdida académica, sino una ruptura en un proyecto de vida. Se desvanece la esperanza de progreso individual, pero también se debilita el tejido social que sostiene las aspiraciones colectivas de las comunidades más excluidas del país.

Reducir la deserción no es simplemente una meta administrativa. Es un imperativo ético, una deuda histórica con sectores sociales que durante décadas han sido privados de oportunidades equitativas de acceso y permanencia en la educación superior.

En territorios como Buenaventura, donde convergen pobreza estructural, racismo institucional, violencia y ausencia estatal, garantizar la continuidad educativa implica mucho más que ofrecer cupos: requiere una

política integral de justicia educativa que combine acompañamiento humano, pertinencia territorial y tecnologías al servicio del cuidado (UNESCO IESALC, 2022).

La tarea de las universidades públicas, en particular aquellas con vocación regional, no debe limitarse a facilitar el ingreso. Su responsabilidad ética y política se extiende a construir condiciones reales para que cada estudiante pueda permanecer, aprender y transformarse.

Esto implica revisar modelos pedagógicos, sistemas de evaluación, políticas de bienestar y, sobre todo, la forma en que se conciben las trayectorias estudiantiles. Es necesario abandonar la idea de que el abandono es un “fracaso individual” para entenderlo como síntoma de un sistema que muchas veces excluye a quienes más necesitan permanecer (Brunner & Ferrer, 2018).

Este libro parte precisamente de una experiencia concreta —la Universidad del Pacífico— para proponer una mirada alternativa frente al problema de la deserción. Una mirada que no niega la complejidad de los desafíos, pero que apuesta por soluciones sostenidas en la evidencia, la innovación y el compromiso institucional.

Una de esas soluciones es la incorporación de herramientas de ciencia de datos, inteligencia artificial y minería de información educativa como medios para anticiparse al abandono.

No se trata de reemplazar la dimensión humana por algoritmos. Se trata de potenciarla. De utilizar la tecnología como una aliada para identificar señales de alerta, comprender patrones de riesgo, y orientar recursos de manera eficiente hacia quienes más los necesitan.

Como han demostrado iniciativas similares en América Latina, los sistemas de alerta temprana no solo permiten intervenir a tiempo, sino que generan una cultura de responsabilidad compartida y acompañamiento integral (De la Cruz & Pérez, 2019; Martínez, 2022).

El uso ético y pedagógico de estas herramientas debe estar guiado por principios de equidad, transparencia y sensibilidad cultural. No se trata de predecir el futuro de los estudiantes con frialdad estadística, sino de abrir posibilidades para que el futuro no les sea negado.

Actuar con urgencia, pero también con cuidado. Con tecnología, pero con sentido humano. Con evidencia, pero con esperanza.



## Capítulo 2. Datos para no desertar: el potencial de la analítica predictiva

En las últimas décadas, la educación superior ha experimentado una transformación silenciosa pero profunda: la irrupción del análisis de datos como herramienta para la toma de decisiones.

Lejos de tratarse de una simple tendencia tecnológica, la analítica predictiva se ha convertido en una estrategia crucial frente a uno de los problemas más persistentes y complejos que enfrentan las universidades: la deserción estudiantil.

Durante años, las instituciones de educación superior han implementado programas de acompañamiento, tutorías, becas y servicios de bienestar con el propósito de reducir el abandono. Sin embargo, muchas de estas acciones han sido reactivas y poco focalizadas.

En contraste, la analítica predictiva propone un giro de paradigma: anticipar, intervenir a tiempo y personalizar las estrategias de retención a partir del conocimiento generado por los datos.

Esta capacidad de prever patrones de riesgo en las trayectorias estudiantiles ha demostrado

tener un impacto significativo en la mejora de la permanencia, especialmente cuando se articula con políticas pedagógicas y sociales coherentes (Picciano, 2012; Agasisti & Bowers, 2017).

La potencia de esta herramienta radica en su capacidad de revelar lo que las estadísticas convencionales no alcanzan a mostrar: los puntos críticos del proceso formativo, los factores combinados que aumentan el riesgo de abandono, y las trayectorias no lineales que los estudiantes recorren a lo largo de su vida académica.

Utilizando técnicas de minería de datos, aprendizaje automático y visualización avanzada, las universidades pueden identificar con mayor precisión a aquellos estudiantes que requieren acompañamiento prioritario, y diseñar intervenciones diferenciadas según su perfil y contexto (Romero & Ventura, 2020).

No obstante, la promesa de la analítica predictiva no está exenta de desafíos. Su efectividad depende en gran medida de la calidad, la integridad y la representatividad de los datos utilizados.

También plantea interrogantes éticos sobre la privacidad, la transparencia y la responsabilidad en la toma de decisiones automatizadas. Por ello, comprender cómo funciona, en qué contextos se ha aplicado con éxito y cuáles son sus límites

éticos es fundamental para imaginar una educación más justa, más inteligente y más humana.

La finalidad de este capítulo es justamente ofrecer ese mapa conceptual y práctico. Exploraremos, en primer lugar, los fundamentos técnicos de la minería de datos y el aprendizaje automático aplicados a contextos educativos.

Luego revisaremos casos internacionales donde estas herramientas han transformado la gestión académica y el acompañamiento estudiantil. Finalmente, abordaremos las implicaciones éticas y políticas del uso de datos en educación superior, con el propósito de formular un enfoque de analítica educativa que no solo sea eficaz, sino también pedagógica, crítica y centrada en el bienestar de los estudiantes.

## **Fundamentos de la minería de datos y el aprendizaje automático**

La minería de datos (data mining) es una disciplina que forma parte del campo más amplio de la inteligencia artificial y de las ciencias de la computación. Se refiere al proceso sistemático de explorar grandes volúmenes de datos —muchas veces desestructurados o heterogéneos— con el objetivo de descubrir patrones, correlaciones, tendencias y

estructuras significativas que no son evidentes a simple vista.

En el ámbito educativo, esta técnica se ha consolidado como una herramienta poderosa para comprender y mejorar los procesos de enseñanza y aprendizaje (Romero & Ventura, 2020).

Aplicada a contextos universitarios, la minería de datos permite analizar desde el historial académico de los estudiantes hasta sus interacciones en plataformas de gestión del aprendizaje (LMS), sus características sociodemográficas, indicadores de desempeño, niveles de participación, hábitos de estudio, registros de tutoría, y variables emocionales o conductuales inferidas a partir de encuestas u observaciones digitales (Baker & Siemens, 2014).

A partir de esta información, las instituciones pueden desarrollar modelos que predicen conductas como la deserción, el rezago académico o la probabilidad de éxito en determinadas asignaturas.

El proceso de minería de datos incluye varias etapas, entre ellas: recolección y limpieza de datos, selección de atributos relevantes, transformación de variables, aplicación de técnicas de descubrimiento (como clasificación,

agrupamiento, regresión o detección de anomalías), y validación de resultados.

Técnicas como el *clustering* permiten agrupar estudiantes con características similares; la *clasificación* se emplea para categorizar casos según su riesgo de abandono o éxito; mientras que el *análisis de redes* facilita entender las relaciones sociales y académicas dentro de un grupo estudiantil (Peña-Ayala, 2014).

Paralelamente, el aprendizaje automático (*machine learning*) es una subdisciplina que permite que los sistemas computacionales desarrollen la capacidad de identificar patrones y realizar predicciones o decisiones sin necesidad de ser programados con instrucciones específicas para cada situación.

En lugar de seguir reglas fijas, los modelos aprenden a partir de ejemplos históricos para generalizar comportamientos futuros. Es decir, se entrena un algoritmo con un conjunto de datos (conocido como conjunto de entrenamiento) para que aprenda a “predecir” un resultado en nuevos casos (Mitchell, 1997).

En el campo de la educación, los algoritmos más comúnmente utilizados incluyen la regresión logística (útil para predecir resultados binarios, como si un estudiante desertará o no), los árboles de decisión (que crean una estructura jerárquica de decisiones a partir de

múltiples variables), las máquinas de soporte vectorial (SVM, por sus siglas en inglés), que buscan maximizar la separación entre categorías, y las redes neuronales artificiales, que emulan estructuras del cerebro humano para detectar patrones complejos no lineales (Kotsiantis et al., 2010).

Estos algoritmos pueden trabajar de forma supervisada (cuando conocen previamente las categorías de los datos) o no supervisada (cuando descubren estructuras ocultas sin etiquetas definidas). La capacidad de estos modelos para aprender, adaptarse y mejorar con nuevos datos hace que su aplicación en contextos dinámicos, como la vida universitaria, sea particularmente eficaz.

El valor de estas herramientas radica no solo en su precisión predictiva, sino en su utilidad para orientar decisiones pedagógicas, administrativas y de bienestar estudiantil.

La analítica educativa basada en minería de datos y aprendizaje automático no busca reemplazar el juicio humano, sino complementarlo con evidencia cuantitativa y análisis contextualizado que permita actuar antes de que el abandono se materialice.

## Casos internacionales de uso de algoritmos en educación

El uso de algoritmos y analítica predictiva en la educación superior ha dejado de ser una promesa para convertirse en una realidad tangible en múltiples sistemas universitarios alrededor del mundo.

Países como Estados Unidos, Reino Unido, Australia y Canadá han liderado la adopción de herramientas de minería de datos y aprendizaje automático para predecir comportamientos estudiantiles, evaluar desempeño académico y, sobre todo, mitigar la deserción universitaria.

Uno de los casos más emblemáticos es el de la Georgia State University (GSU), en Estados Unidos. Esta universidad pública, con una matrícula mayoritariamente compuesta por estudiantes de primera generación y minorías étnicas, diseñó un sistema de alerta temprana basado en más de 800 variables académicas, administrativas y comportamentales.

El sistema, que integra información en tiempo real sobre desempeño en cursos, inscripciones, asistencia y uso de recursos institucionales, permite intervenir antes de que los estudiantes acumulen rezagos críticos. Como resultado, GSU logró reducir su tasa de deserción en más del 30 % entre 2012 y 2017 y, lo más relevante, cerrar

significativamente las brechas de logro entre estudiantes blancos y afroamericanos.

En el Reino Unido, diversas instituciones han desarrollado estrategias similares bajo el paraguas del *Learning Analytics Service* impulsado por Jisc, una organización pública que provee infraestructura digital al sector educativo. Este servicio ofrece a las universidades británicas herramientas para monitorear el compromiso estudiantil mediante análisis de datos de los entornos virtuales de aprendizaje, alertas automatizadas, y paneles visuales para profesores y tutores.

Las universidades que han integrado estas soluciones reportan mejoras en la identificación temprana de estudiantes en riesgo y un uso más eficiente de sus recursos de apoyo académico.

En Australia, la Universidad de Nueva Inglaterra implementó un programa de analítica predictiva llamado *Early Alert Program*, que combina algoritmos de *machine learning* con acompañamiento personalizado por parte de tutores humanos. Este modelo ha sido especialmente eficaz en contextos de educación en línea, donde el seguimiento de estudiantes resulta más complejo.

La clave del éxito ha sido la combinación entre el análisis automatizado y la acción humana oportuna y empática (Ifenthaler & Yau, 2020).

En América Latina, aunque con menor sistematicidad y menor inversión en tecnología, también se han desarrollado experiencias significativas. En Chile, el programa PACE (Programa de Acompañamiento y Acceso Efectivo a la Educación Superior) ha incorporado herramientas de análisis de datos para identificar estudiantes en situación de vulnerabilidad académica y diseñar intervenciones diferenciadas de acompañamiento.

La experiencia ha mostrado que la incorporación de información longitudinal permite anticiparse a los momentos críticos del primer año universitario, que suele ser el más sensible en términos de abandono (Valenzuela & González, 2020).

En Brasil, la Universidad Federal de Minas Gerais desarrolló un sistema de modelado predictivo para priorizar la asignación de recursos de bienestar, tales como tutorías, apoyos alimentarios y ayudas económicas. El modelo, entrenado con datos de cohortes anteriores, logró mejorar la focalización de los apoyos y reducir la sobrecarga en los servicios institucionales de atención.

En todos estos casos, el uso de algoritmos ha demostrado ser una herramienta valiosa no solo para aumentar la eficiencia de las intervenciones, sino también para humanizar la

experiencia educativa. Lejos de despersonalizar los procesos, la analítica predictiva ha permitido que las instituciones se acerquen más y mejor a sus estudiantes, interviniendo cuando aún hay tiempo para actuar.

Como señalan Gašević, Dawson y Siemens (2015), cuando se diseña con criterios éticos, pedagógicos y colaborativos, la analítica educativa puede convertirse en una herramienta de justicia académica, ayudando a que nadie quede atrás en su trayecto formativo.

## Ética y riesgos del uso de datos en educación

El uso de herramientas predictivas en la educación superior —aunque con indudable potencial— no está exento de riesgos éticos, sociales y pedagógicos. La creciente dependencia de los algoritmos para anticipar comportamientos estudiantiles como la deserción, el bajo rendimiento o el compromiso académico, ha generado una serie de debates sobre los límites y condiciones del uso de datos en entornos educativos.

Uno de los principales peligros es la estigmatización que puede derivarse del etiquetado automatizado. Cuando un estudiante es clasificado como “de alto riesgo” por un modelo predictivo, esa etiqueta puede condicionar la percepción de docentes, tutores y administrativos, generando una forma de sesgo confirmatorio que margina en lugar de incluir.

Además, si el algoritmo se alimenta de datos históricos cargados de inequidades —como tasas de deserción más altas entre estudiantes de bajos ingresos o comunidades racializadas—, existe el riesgo de que reproduzca esos mismos patrones, perpetuando una lógica de exclusión algorítmica (Prinsloo & Slade, 2017; Eubanks, 2018).

Otro aspecto crítico es el de la privacidad de los datos personales. A medida que las universidades recolectan, cruzan y procesan grandes volúmenes de información académica, administrativa y a veces hasta emocional (por ejemplo, datos de encuestas sobre bienestar o salud mental), aumentan también los riesgos de uso indebido, filtración o explotación comercial de esa información.

La protección de los datos personales de los estudiantes no es solo una exigencia legal, sino un compromiso ético con su autonomía y dignidad (Slade & Prinsloo, 2013).

Surgen así preguntas fundamentales: ¿Quién decide qué variables son relevantes para los modelos predictivos? ¿Con qué criterios se interpreta la información? ¿Qué nivel de intervención es aceptable sobre una trayectoria estudiantil basada en una predicción estadística? ¿Existe margen para el error, la excepción, el contexto?

Ante estos desafíos, diversos marcos éticos internacionales han desarrollado principios orientadores para el uso responsable de la analítica de datos en la educación. Entre los más citados están los propuestos por la Joint Information Systems Committee (Jisc) en Reino Unido, y los de la European Commission's Ethics Guidelines for Trustworthy AI, que subrayan la necesidad de garantizar:

consentimiento informado y voluntario, transparencia y explicabilidad de los algoritmos, supervisión humana significativa, mecanismos de corrección y apelación, y auditorías éticas periódicas sobre los sistemas implementados (Jisc, 2016; European Commission, 2019).

En palabras de Williamson y Eynon (2020), la educación basada en datos no debe ser tecnocrática ni funcionalista. Debe ser una práctica profundamente pedagógica, deliberativa y ética, que promueva el empoderamiento del estudiante, no su vigilancia o disciplinamiento.

El objetivo último de la analítica predictiva no puede ser únicamente la eficiencia institucional, sino el bienestar, la equidad y el acompañamiento efectivo de quienes enfrentan más obstáculos para permanecer en el sistema educativo.

En este sentido, los datos no deben reemplazar el juicio pedagógico ni la relación humana entre docentes, estudiantes y acompañantes. Pero sí pueden convertirse en una herramienta estratégica para visibilizar el riesgo antes de que sea irreversible, y así permitir que las universidades asuman de manera más consciente su responsabilidad en la construcción de trayectorias educativas sostenibles, dignas y transformadoras.



## Capítulo 3. El caso de la Universidad del Pacífico

### Contexto institucional y territorial

La Universidad del Pacífico es una institución pública ubicada en Buenaventura, en el litoral Pacífico colombiano, una de las regiones con mayores desafíos sociales, económicos, ambientales e institucionales del país. Su creación respondió a una deuda histórica con los pueblos afrodescendientes y las comunidades rurales del suroccidente colombiano, históricamente excluidas de los beneficios de la educación superior.

En este sentido, la universidad representa no solo una institución académica, sino un símbolo de resistencia y reivindicación del derecho a la educación como bien público (UNESCO IESALC, 2022).

Buenaventura, su ciudad sede, ha sido catalogada como uno de los municipios con mayor vulnerabilidad estructural en América Latina. Su ubicación estratégica como principal puerto sobre el Pacífico contrasta con los indicadores de pobreza, desigualdad y exclusión.

La presencia del conflicto armado, el narcotráfico, la minería ilegal y el racismo

estructural ha configurado una realidad social compleja que impacta directamente la oferta y la calidad de los servicios públicos, incluida la educación (DANE, 2023; Cruz et al., 2020).

En este contexto, la Universidad del Pacífico enfrenta el reto de formar profesionales en condiciones institucionales adversas. Aunque su misión apunta a transformar el territorio a través del conocimiento, sus recursos son limitados: enfrenta dificultades presupuestales crónicas, rezagos en infraestructura física y tecnológica, escasa conectividad digital en las zonas rurales donde habita buena parte de su estudiantado, y un entorno de inseguridad que obstaculiza tanto la movilidad como la convivencia universitaria (MEN, 2023).

Pese a ello, la institución ha consolidado un papel protagónico como espacio de articulación intercultural, de producción de conocimiento situado y de empoderamiento comunitario.

A través de programas de extensión, proyectos de investigación territorializados y una oferta académica sensible a las realidades étnicas y ambientales del litoral, la Universidad del Pacífico ha contribuido a democratizar el acceso a la educación superior y a fortalecer las capacidades locales para el desarrollo sostenible.

Este carácter transformador no exime a la universidad de los desafíos profundos que enfrenta en términos de permanencia estudiantil. Las condiciones materiales de vida de sus estudiantes, las brechas digitales, las carencias en el sistema de bienestar universitario y la fragmentación institucional son factores que, al converger, condicionan la trayectoria académica de muchos jóvenes.

Por eso, el análisis de su experiencia resulta clave para comprender cómo se expresa la deserción en contextos de alta exclusión, y qué estrategias pueden construirse desde la analítica educativa con enfoque territorial.

### **Recolección y depuración de datos estudiantiles**

Uno de los principales desafíos de la Universidad del Pacífico en materia de gestión académica ha sido la carencia histórica de un sistema integrado de información estudiantil que permita tomar decisiones basadas en evidencia.

La segmentación de los datos entre distintas dependencias (registro, bienestar, finanzas, sistemas) y la falta de interoperabilidad entre plataformas tecnológicas generaron durante años una brecha crítica entre la información

disponible y su utilidad para prevenir fenómenos como la deserción.

Este vacío fue abordado en el marco del proyecto de tesis que dio origen a esta investigación, mediante una estrategia metodológica de consolidación, depuración y preparación de un conjunto de datos confiable y representativo.

El proceso inició con una fase de diálogo interinstitucional para definir los criterios de acceso, anonimización y estructura de los datos, conforme a principios de ética en el tratamiento de información sensible (Slade & Prinsloo, 2013). Posteriormente, se recolectaron más de 13.000 registros históricos correspondientes a estudiantes de pregrado entre los años 2015 y 2020.

El proceso de depuración incluyó la identificación y eliminación de duplicados, el tratamiento de datos faltantes mediante interpolaciones y codificaciones, y la transformación de variables en formatos utilizables para algoritmos de aprendizaje automático.

Las 53 variables iniciales fueron sometidas a un análisis exploratorio para identificar aquellas con mayor valor explicativo, seleccionando finalmente 19 variables para el modelado predictivo. Estas abarcaron dimensiones

académicas (promedio, materias cursadas y perdidas), características personales (edad, género, lugar de procedencia), condiciones económicas (estado de cartera, frecuencia de pagos), y estado final del estudiante (activo, egresado, inactivo).

Asimismo, se aplicaron técnicas de preprocesamiento como la normalización de variables continuas, la discretización por rangos y la codificación de variables categóricas, asegurando la compatibilidad con algoritmos de clasificación como árboles de decisión, regresión logística y random forest.

Para abordar el desbalance entre las clases (desertores vs. no desertores), se utilizó el método SMOTE (Synthetic Minority Over-sampling Technique), lo cual permitió equilibrar el conjunto de entrenamiento sin comprometer la validez estadística del modelo (Chawla et al., 2002).

El resultado fue un dataset robusto, limpio y balanceado, listo para ser usado en tareas de minería de datos y *machine learning*. Más allá de su valor técnico, este conjunto de datos constituye una herramienta estratégica para el diseño de sistemas de alerta temprana y políticas de intervención focalizada, representando un avance significativo hacia una universidad que aprende de sus datos y actúa en función de ellos.

## Diagnóstico inicial de la deserción

El análisis exploratorio realizado a partir del conjunto de datos consolidado permitió construir un diagnóstico empírico sobre la magnitud, distribución y correlatos del fenómeno de la deserción estudiantil en la Universidad del Pacífico.

Uno de los hallazgos más significativos fue que la tasa de deserción acumulada en el segundo semestre de 2020 alcanzó el 15,32 %, una cifra superior al promedio nacional registrado por el SPADIES para universidades públicas regionales, lo que indica una situación particularmente crítica en esta institución (MEN, 2023).

Este fenómeno se manifestó con mayor intensidad en estudiantes de estrato socioeconómico 1, en quienes la precariedad de las condiciones materiales de vida se intersecta con otras variables de vulnerabilidad. Se evidenció una fuerte sobrerrepresentación de mujeres afrodescendientes provenientes de zonas rurales entre los casos de abandono, lo que sugiere una dimensión interseccional que vincula género, etnicidad y territorio como factores agravantes de la exclusión educativa.

Entre los predictores más frecuentes de deserción identificados en el análisis estuvieron el bajo rendimiento académico sostenido, la

acumulación de asignaturas reprobadas, los atrasos en el pago de matrícula, y la residencia en áreas periféricas con limitada conectividad y condiciones de seguridad.

Estas variables no operan de manera aislada, sino que se refuerzan mutuamente en una espiral de riesgo que hace más probable el abandono (González González, 2021).

El patrón territorial fue igualmente revelador. Los estudiantes provenientes de corregimientos ribereños como La Bocana, San Marcos o Zacarías —zonas con escasa infraestructura de transporte, servicios públicos deficientes y bajos niveles de escolaridad de los padres— presentaron tasas de deserción notoriamente más altas.

Esta constatación subraya la importancia de comprender la deserción no solo como un problema individual, sino como una expresión de desigualdades estructurales ancladas en la geografía y la historia del conflicto y la pobreza (DANE, 2023).

El diagnóstico evidenció también la necesidad urgente de fortalecer las estrategias de acompañamiento integral, con especial énfasis en los primeros semestres de formación, que es donde se concentra el mayor porcentaje de retiros.

La falta de redes de apoyo académico, psicológico y económico durante esta fase inicial incrementa el riesgo de deserción temprana, y deja al estudiante expuesto a múltiples vulnerabilidades sin una estructura institucional que lo sostenga (Slade & Prinsloo, 2013).

En suma, el fenómeno de la deserción en la Universidad del Pacífico no puede comprenderse como una suma de decisiones individuales, sino como el resultado de una matriz estructural de exclusión que debe ser enfrentada con estrategias interinstitucionales, políticas públicas diferenciadas y un compromiso genuino con la equidad educativa. Este diagnóstico constituye un primer paso hacia una intervención más justa, focalizada y contextualizada.

## Capítulo 4. Diseño del modelo predictivo

### Elección metodológica: KDD y aprendizaje supervisado

El diseño del modelo predictivo de deserción en la Universidad del Pacífico se fundamentó en el enfoque metodológico conocido como Knowledge Discovery in Databases (KDD), un proceso estructurado, iterativo y orientado a la extracción de conocimiento significativo a partir de grandes volúmenes de datos heterogéneos.

Este enfoque fue elegido no solo por su robustez técnica, sino por su alineación con el propósito del proyecto: generar conocimiento aplicable para mejorar la permanencia estudiantil en contextos de alta vulnerabilidad social (Fayyad et al., 1996).

El proceso de KDD se compone de varias etapas: selección de datos relevantes, preprocesamiento y transformación, minería de datos propiamente dicha, y evaluación e interpretación de los resultados.

Esta estructura permite una integración fluida entre el análisis estadístico y el desarrollo de modelos predictivos, facilitando la retroalimentación entre etapas para optimizar

los resultados en función del contexto institucional.

En este caso, se trató de construir un modelo que no solo tuviera precisión, sino que pudiera ser implementado como herramienta útil para los equipos de bienestar y gestión académica.

Dentro del marco de KDD, se optó por un enfoque de aprendizaje supervisado, ideal cuando se dispone de una variable objetivo previamente conocida —en este caso, el estado final del estudiante: activo, egresado o desertor— y un conjunto de datos etiquetados que permite entrenar algoritmos para predecir nuevos casos.

Este enfoque fue particularmente adecuado porque facilitó el uso de datos históricos institucionales para generar predicciones automatizadas, con base en patrones de comportamiento previamente observados (Mitchell, 1997).

El aprendizaje supervisado permite evaluar el desempeño de diferentes modelos a partir de métricas como precisión, sensibilidad, especificidad, exactitud global y valor F1, lo que facilita comparar distintas técnicas y seleccionar la más apropiada según los objetivos del proyecto.

Además, este enfoque puede adaptarse a nuevas cohortes con ajustes mínimos, lo que ofrece ventajas prácticas en términos de escalabilidad y actualización continua del sistema de alerta temprana.

La elección de esta metodología respondió también a una necesidad estratégica: construir capacidades institucionales en analítica de datos con potencial de ser sostenidas en el tiempo.

El proyecto buscó no solo generar un modelo predictivo técnicamente sólido, sino también establecer las bases metodológicas para una cultura de gestión basada en datos dentro de la universidad, en línea con las recomendaciones de la literatura sobre aprendizaje analítico y sistemas inteligentes en educación (Baker & Siemens, 2014).

Finalmente, al tratarse de un entorno institucional con recursos limitados, era fundamental elegir una metodología compatible con herramientas de software libre y capacidades técnicas disponibles, sin sacrificar la rigurosidad analítica. En ese sentido, el marco KDD y el aprendizaje supervisado ofrecieron una solución eficiente, replicable y apropiada para el contexto de la Universidad del Pacífico.

## **Construcción del dataset: variables, preprocesamiento y balanceo**

La construcción del dataset predictivo fue una de las fases más críticas del proyecto, ya que sobre su calidad dependía directamente la precisión y confiabilidad del modelo final. Se partió de una base de datos institucional compuesta por 13.202 registros correspondientes a estudiantes matriculados entre 2015 y 2020. Esta información fue extraída de múltiples sistemas internos, entre ellos las plataformas de registro académico, gestión financiera, sistemas de admisiones y bienestar universitario, lo que implicó un proceso complejo de consolidación y estandarización.

Durante la fase de limpieza, se identificaron errores comunes como registros duplicados, valores faltantes, formatos heterogéneos en variables categóricas y outliers estadísticos en variables continuas.

Para solucionarlos, se implementaron técnicas de imputación condicional, detección de valores atípicos mediante reglas de Tukey y revisión manual de casos críticos con apoyo de funcionarios institucionales. Esta etapa también incluyó el diseño de un diccionario de datos institucional, fundamental para garantizar la

trazabilidad de cada variable y su correcta interpretación analítica (Peña-Ayala, 2014).

Las variables retenidas para el modelado fueron seleccionadas a partir de criterios de significancia estadística (mediante análisis de correlación, ANOVA y chi-cuadrado) y de relevancia operativa, en función de su capacidad para informar decisiones institucionales.

Se seleccionaron 19 variables independientes, clasificadas en cuatro dimensiones: académicas (promedio acumulado, número de materias vistas y reprobadas), sociodemográficas (edad, sexo, estrato socioeconómico, procedencia geográfica), financieras (estado de cartera, número de pagos realizados, morosidad), y administrativas (programa, cohorte, estado académico final).

En la etapa de preprocesamiento, las variables numéricas fueron normalizadas mediante escalamiento min-max, lo cual evitó que valores extremos influyeran desproporcionadamente en los modelos. Las variables categóricas se transformaron mediante codificación ordinal y one-hot encoding, según su naturaleza ordinal o nominal.

Además, se discretizaron algunas variables continuas como la edad y el promedio

académico, lo cual permitió mejorar la interpretabilidad y el desempeño de algoritmos basados en árboles como Random Forest y Decision Trees (Han, Kamber & Pei, 2012).

Dado que los casos de deserción eran numéricamente inferiores respecto al total de registros, se identificó un problema clásico de desbalance de clases. Para resolverlo, se aplicó la técnica SMOTE (Synthetic Minority Over-sampling Technique), que consiste en generar ejemplos sintéticos de la clase minoritaria a partir de sus vecinos más cercanos.

Esta técnica, ampliamente validada en entornos educativos, permitió mejorar la capacidad del modelo para detectar estudiantes en riesgo sin sacrificar precisión general (Chawla et al., 2002).

En conjunto, este proceso permitió construir un dataset robusto, representativo y éticamente tratado, con alta calidad técnica y relevancia institucional. Su valor no solo reside en su utilidad para el modelo predictivo, sino en su potencial para alimentar futuros sistemas de analítica institucional y toma de decisiones basada en evidencia.

## **Algoritmos utilizados: Random Forest, SVM, Árboles de decisión, etc.**

Con el dataset estructurado y balanceado, se procedió a la fase de modelado predictivo empleando una batería de algoritmos representativos del aprendizaje supervisado. Se seleccionaron técnicas tanto clásicas como avanzadas, con el fin de comparar su capacidad predictiva en el contexto específico de la Universidad del Pacífico.

Los modelos evaluados incluyeron Árboles de Decisión (C4.5), Random Forest, Regresión Logística, Máquinas de Vectores de Soporte (SVM), y Naive Bayes, todos ampliamente utilizados en tareas de clasificación binaria en educación (Kotsiantis et al., 2010).

El entrenamiento de los modelos se realizó utilizando una partición estratificada 70/30 entre conjunto de entrenamiento y conjunto de prueba, asegurando la proporcionalidad de clases en ambos subconjuntos.

Adicionalmente, se aplicó validación cruzada de 10 pliegues para mitigar el riesgo de sobreajuste y garantizar mayor estabilidad estadística de los resultados (Hastie, Tibshirani, & Friedman, 2009).

El algoritmo que arrojó el mejor desempeño fue Random Forest, con una precisión del 86 %,

sensibilidad del 84 %, especificidad del 87 % y una puntuación F1 de 0.85. Este modelo de ensamblado, basado en la combinación de múltiples árboles de decisión aleatorios, mostró una notable robustez ante datos ruidosos y una alta capacidad para detectar interacciones no lineales entre variables.

Además, su bajo riesgo de sobreajuste y su capacidad de manejo de valores faltantes lo convierten en una alternativa idónea para entornos educativos con registros incompletos o heterogéneos (Breiman, 2001).

Los Árboles de Decisión C4.5, si bien más interpretables y visualmente comprensibles para los equipos institucionales, presentaron una menor capacidad de generalización, especialmente en cohortes con alta variabilidad académica y socioeconómica. Aun así, fueron útiles para identificar rutas lógicas de abandono y facilitar la comunicación de resultados a públicos no técnicos.

Las Máquinas de Vectores de Soporte (SVM) ofrecieron un rendimiento competitivo, particularmente en precisión, pero demostraron una sensibilidad mayor a la calidad del preprocesamiento, especialmente en la escala de variables y la separación de clases.

Requieren mayor ajuste de hiperparámetros y su nivel de opacidad algorítmica puede

dificultar su comprensión institucional (Cortes & Vapnik, 1995).

Por su parte, la Regresión Logística mostró un desempeño moderado, pero aportó información valiosa sobre la magnitud y dirección del efecto de cada predictor, lo cual resultó útil para el diseño de estrategias de intervención. Finalmente, Naive Bayes, aunque con menor rendimiento global, fue útil como modelo base por su simplicidad y velocidad de ejecución, lo que puede ser relevante en sistemas en tiempo real con recursos limitados (Zhang, 2004).

La evaluación de estos modelos no se limitó a métricas cuantitativas. Se consideraron criterios cualitativos como la facilidad de implementación, el nivel de transparencia, la compatibilidad con infraestructuras tecnológicas existentes, y la posibilidad de ser usados como base para sistemas de alerta temprana.

Esta evaluación integral asegura que el modelo final no solo sea eficaz en términos predictivos, sino también funcional y sostenible dentro del ecosistema institucional de la Universidad del Pacífico.



## Capítulo 5. Evaluación de modelos y resultados

### Comparación de métricas: precisión, exactitud, F1, ROC

La fase de evaluación comparativa de modelos constituyó un paso esencial en la validación empírica del enfoque predictivo propuesto. Dado que el objetivo era predecir la probabilidad de deserción estudiantil —un fenómeno de naturaleza binaria—, se recurrió a un conjunto de métricas estándar en el campo del aprendizaje supervisado para garantizar una evaluación integral del desempeño de cada algoritmo.

Estas métricas permiten identificar no solo la capacidad global del modelo para hacer predicciones correctas, sino también su eficacia específica en la identificación de estudiantes en riesgo, que es el objetivo más crítico desde el punto de vista institucional.

La primera métrica empleada fue la precisión (accuracy), definida como el número total de predicciones correctas dividido entre el total de observaciones. Si bien esta métrica es ampliamente utilizada, su valor puede resultar engañoso en conjuntos de datos desbalanceados —como ocurre comúnmente en predicciones de

deserción—, donde la clase mayoritaria (estudiantes que no desertan) domina el conjunto y puede inflar artificialmente el rendimiento del modelo.

Por esta razón, se dio mayor peso a métricas como la sensibilidad (recall), que mide la proporción de verdaderos positivos identificados correctamente; la especificidad, que evalúa la capacidad del modelo para identificar correctamente los verdaderos negativos; y el valor F1, que representa la media armónica entre la precisión y la sensibilidad, y resulta especialmente útil cuando existe un desbalance de clases y es importante minimizar tanto los falsos positivos como los falsos negativos (Saito & Rehmsmeier, 2015).

Asimismo, se utilizó el área bajo la curva ROC (AUC-ROC) como métrica resumen del rendimiento general del modelo. Este indicador se construye a partir de la curva ROC, que traza la tasa de verdaderos positivos frente a la tasa de falsos positivos a diferentes umbrales de clasificación.

El AUC-ROC es particularmente valioso porque permite comparar modelos sin depender de un único umbral y porque refleja la capacidad discriminativa global del modelo. Un valor de 1 indica una separación perfecta entre clases, mientras que un valor de 0.5 equivale a un clasificador aleatorio.

Los resultados de esta evaluación comparativa revelaron que el algoritmo Random Forest obtuvo el mejor desempeño general, con un AUC de 0.91, seguido por SVM con 0.87, regresión logística con 0.83, y árboles de decisión con 0.80. Naive Bayes, aunque con un valor AUC más bajo (0.76), presentó una ejecución ágil y baja complejidad computacional, lo que podría justificar su uso en sistemas con limitaciones de hardware o necesidad de respuesta en tiempo real (Han et al., 2012).

Estas diferencias, aunque esperadas debido a la naturaleza de cada algoritmo, confirmaron la superioridad de los métodos de ensamblado — como Random Forest— para problemas complejos con múltiples variables interrelacionadas y clases desbalanceadas.

Además, estas métricas no solo sirvieron para elegir el modelo óptimo, sino también para evidenciar la importancia de seleccionar indicadores que reflejen con precisión el objetivo institucional: reducir la deserción mediante intervenciones oportunas y dirigidas.

### **Selección del mejor modelo**

Tras un análisis detallado de los resultados obtenidos en la fase de evaluación comparativa, se determinó que el algoritmo Random Forest

ofrecía el mejor equilibrio entre rendimiento estadístico, interpretabilidad y viabilidad operativa. Este modelo alcanzó una precisión del 86 %, una sensibilidad del 84 %, y una puntuación F1 de 0.85, lo que indica una alta capacidad para identificar correctamente tanto a los estudiantes en riesgo de deserción como a aquellos con trayectorias estables (Breiman, 2001).

Una de las principales ventajas de Random Forest radica en su estructura de ensamblado de árboles de decisión, lo cual reduce significativamente el riesgo de sobreajuste y mejora la estabilidad de las predicciones.

Asimismo, este algoritmo permite calcular la importancia relativa de cada variable predictora, ofreciendo a los equipos de gestión educativa información clave para el diseño de intervenciones focalizadas (Han, Kamber, & Pei, 2012).

Más allá del rendimiento estadístico, la elección del modelo se fundamentó en criterios de aplicabilidad institucional. Random Forest puede ser implementado fácilmente a través de plataformas de software libre como Scikit-learn (Python) o Weka (Java), herramientas accesibles para universidades con recursos tecnológicos limitados.

Además, su compatibilidad con entornos que presentan datos incompletos, distribuciones no lineales o ruido estadístico lo hacen especialmente adecuado para contextos regionales como el de la Universidad del Pacífico (Scikit-learn developers, 2023).

La capacidad del modelo para adaptarse a nuevas cohortes mediante reentrenamiento periódico, su escalabilidad para integrarse en sistemas más amplios de monitoreo académico, y su potencial para alimentar motores de alerta temprana fueron también determinantes en su selección.

En efecto, el modelo puede ser calibrado para generar alertas basadas en umbrales de riesgo personalizados, ofreciendo a la universidad una herramienta dinámica de gestión de la permanencia estudiantil.

En resumen, la elección de Random Forest no respondió únicamente a criterios de exactitud, sino a una lógica integral de sostenibilidad, utilidad institucional y capacidad transformadora. Su implementación efectiva permitirá a la Universidad del Pacífico avanzar hacia una cultura de prevención y acompañamiento proactivo, orientada por datos y centrada en el bienestar estudiantil.

## **Implicaciones prácticas para la gestión universitaria**

Los hallazgos del proceso de evaluación y validación del modelo predictivo desarrollado en este estudio ofrecen aportes estratégicos para la gestión universitaria en contextos caracterizados por limitaciones institucionales y altos índices de vulnerabilidad, como es el caso de la Universidad del Pacífico.

En particular, evidencian la posibilidad de construir capacidades analíticas locales que permiten transitar de una gestión reactiva a una gestión anticipativa de la permanencia estudiantil.

En primer lugar, el uso de modelos de analítica predictiva permite abandonar los enfoques generalistas que tratan a todos los estudiantes por igual, para adoptar estrategias personalizadas de intervención temprana. Estos modelos pueden identificar perfiles de riesgo con mayor precisión, y con ello priorizar acciones de acompañamiento académico, psicológico y financiero. Como lo plantean Gašević, Dawson y Siemens (2015), el uso responsable de la analítica educativa posibilita una educación más inclusiva y centrada en el estudiante.

En segundo lugar, la identificación de variables predictoras clave —como el promedio

académico, el número de materias reprobadas, la procedencia geográfica y la morosidad financiera— proporciona evidencia directa para rediseñar políticas institucionales. Estas variables no solo explican la deserción, sino que permiten actuar sobre ella, orientando con mayor eficiencia los recursos limitados de las áreas de bienestar, tutoría y permanencia.

En tercer lugar, este modelo ha demostrado ser un catalizador para la transformación institucional hacia una cultura de gestión basada en evidencia. El proceso de construcción del modelo implicó la articulación de diversas áreas—tecnología, planeación, registro, bienestar—, promoviendo un aprendizaje organizacional que trasciende lo técnico. Esto resulta especialmente valioso en universidades regionales donde las decisiones muchas veces se toman con base en intuiciones más que en datos concretos (Ifenthaler & Yau, 2020).

Además, el modelo tiene un alto potencial de replicabilidad y adaptación en otras instituciones públicas de características similares, lo cual lo convierte en un recurso estratégico para fortalecer el sistema de educación superior colombiano. Puede integrarse en plataformas de monitoreo institucional, alimentar sistemas de alerta temprana, y convertirse en una herramienta útil

para procesos de acreditación y rendición de cuentas.

Finalmente, más allá de su dimensión instrumental, este modelo representa una apuesta por la equidad educativa y la justicia territorial. Al articular tecnología con sensibilidad social, permite visibilizar las desigualdades que atraviesan las trayectorias educativas de los estudiantes más excluidos.

En este sentido, no sustituye la acción humana, sino que la complementa, orientándola y fortaleciéndola para que ningún estudiante abandone su proyecto educativo por falta de acompañamiento oportuno y pertinente.

## Capítulo 6. Arquitectura del sistema de alertas tempranas

### Diseño de la arquitectura con modelo C4

Para garantizar la implementación operativa, escalable y mantenible del modelo predictivo en el ecosistema institucional de la Universidad del Pacífico, se adoptó una arquitectura basada en el modelo C4 (Contexto, Contenedores, Componentes y Código), propuesto por Simon Brown (2018).

Esta metodología arquitectónica es ampliamente reconocida por su capacidad de representar visual y jerárquicamente los distintos niveles de un sistema de software, desde una visión macro hasta el detalle del código fuente, facilitando su comprensión tanto para equipos técnicos como para responsables institucionales no especializados.

En el nivel de contexto, se delineó la interacción entre el sistema de alertas tempranas y su entorno, identificando los principales actores: (1) el sistema institucional de gestión académica ACADEMUSOFT, fuente primaria de datos; (2) los usuarios administrativos y académicos responsables del

acompañamiento estudiantil; y (3) los estudiantes, beneficiarios finales del sistema.

Esta capa contextual permitió definir los límites del sistema, los flujos de información bidireccional (input y output), y los canales de comunicación intersistemas, enmarcados en políticas institucionales de privacidad y protección de datos (European Commission, 2019).

A nivel de contenedores, la arquitectura fue segmentada en módulos funcionales clave: (1) un subsistema ETL (Extract, Transform, Load), que realiza la extracción periódica de datos desde diversas fuentes internas, como ACADEMUSOFT, bases de bienestar y registros financieros; (2) un motor de predicción, que ejecuta el modelo Random Forest con parámetros calibrados según cohortes históricas; (3) una base de datos especializada para almacenar resultados de riesgo, umbrales y trazabilidad de alertas; y (4) una interfaz gráfica de usuario (GUI) orientada a los equipos de gestión, desarrollada con tecnologías como Flask o Django, integradas mediante servicios RESTful y arquitecturas de microservicios (Newman, 2015).

En el nivel de componentes, cada contenedor fue descompuesto en submódulos específicos. El módulo ETL, por ejemplo, incluye validadores sintácticos, limpiadores de datos redundantes y

normalizadores de variables. El motor predictivo integra un pipeline de inferencia, control de versiones del modelo (MLOps), y evaluadores de rendimiento.

El generador de alertas incorpora lógica condicional para clasificación de riesgos y herramientas de visualización como Grafana o Plotly. Finalmente, un componente de auditoría y trazabilidad registra cada inferencia realizada, reforzando la transparencia del sistema (Amershi et al., 2019).

Esta arquitectura modular, desacoplada y basada en estándares abiertos favorece la evolución incremental del sistema, su migración a servicios en la nube (por ejemplo, mediante contenedores Docker), y su adaptabilidad a nuevas cohortes o a otros escenarios de predicción (abandono en posgrados, riesgo académico, etc.).

También permite realizar pruebas automatizadas por componente, facilitando el mantenimiento y el escalamiento futuro en términos técnicos y funcionales.

## **Interoperabilidad con ACADEMUSOFT y escenarios de calidad**

Uno de los principales requisitos para la viabilidad del sistema de alertas tempranas fue garantizar su interoperabilidad con

ACADEMUSOFT, la plataforma de gestión académica utilizada por la Universidad del Pacífico.

Esta interoperabilidad debía ser bidireccional, estable, segura y con mínima intervención manual, cumpliendo con estándares de integridad transaccional y trazabilidad.

Para ello, se diseñó una serie de APIs RESTful que permiten la extracción automatizada y segura de datos clave como historial académico, número de matrículas activas, materias reprobadas, promedio acumulado, número de cancelaciones, estado de cartera y variables sociodemográficas básicas.

Estas APIs emplean protocolos HTTPS y autenticación basada en tokens JWT (JSON Web Tokens), lo que garantiza sesiones autenticadas y minimiza vulnerabilidades de seguridad (Fielding, 2000; Jones et al., 2015).

Además, se estableció una política de actualización periódica del dataset — idealmente semanal— a través de trabajos programados (cron jobs) y control de versiones para cada conjunto descargado, asegurando la trazabilidad de la información y la comparación entre cohortes.

Esta integración permite que los datos fluyan desde ACADEMUSOFT hacia el motor predictivo sin necesidad de procesamiento intermedio manual, aumentando así la eficiencia operativa y reduciendo errores humanos.

Para salvaguardar la calidad de los datos, se implementaron validaciones estructurales y semánticas en el módulo ETL, incluyendo verificación de tipos de datos, detección de valores nulos, y consistencia entre registros relacionados (por ejemplo, coincidencia entre estado financiero y matrícula activa).

También se diseñó una lógica de consistencia para que los estudiantes con datos faltantes críticos (como promedio académico o situación de cartera) sean marcados automáticamente para revisión manual, antes de ser procesados por el modelo de riesgo.

En cuanto a la evaluación del rendimiento del sistema, se definieron escenarios de calidad orientados a asegurar tanto la eficiencia técnica como el valor institucional de las predicciones. Las métricas incluyeron: (1) tiempo medio de ejecución del modelo (menos de 2 minutos por cohorte), (2) tasa de falsos positivos y negativos, evaluadas mediante curvas ROC-AUC y matrices de confusión, (3) cobertura institucional (porcentaje de estudiantes con predicción válida), y (4) tasa de actualización de alertas semestral.

Estas métricas permiten no solo monitorear el desempeño técnico del sistema, sino establecer indicadores estratégicos para la mejora continua, y para la toma de decisiones institucional basada en evidencia (Pardo & Siemens, 2014).

## **Visualización de resultados y recomendaciones institucionales**

La interfaz de visualización desarrollada para el sistema de alertas tempranas fue concebida bajo principios de diseño centrado en el usuario, accesibilidad universal y utilidad funcional. Inspirada en modelos de aprendizaje analítico aplicados en entornos universitarios (Ifenthaler & Yau, 2020), esta interfaz se dirige a los equipos de bienestar institucional, coordinación académica, decanaturas y demás responsables del acompañamiento estudiantil, permitiendo una lectura rápida y comprensiva de los niveles de riesgo individuales y colectivos.

El núcleo visual es un dashboard interactivo que organiza la información por niveles jerárquicos: vista general por cohorte, desagregación por programa académico, y acceso individual por estudiante.

Se utilizan visualizaciones dinámicas —como gráficos de barras, matrices de riesgo y mapas de calor— desarrolladas con herramientas

como Plotly, Dash y D3.js, que permiten explorar las predicciones en tiempo real y con posibilidad de filtrar por criterios como sexo, estrato socioeconómico, zona de residencia o antigüedad en la universidad.

Cada estudiante cuenta con un perfil detallado que incluye su trayectoria académica, comportamiento financiero, y principales variables que determinaron su clasificación como caso de riesgo. Esta información no solo permite identificar patrones, sino también humanizar el dato, al comprender el contexto detrás del número. Además, se incorporó una sección de justificación algorítmica mediante técnicas de interpretabilidad como SHAP (SHapley Additive exPlanations), que explican el peso relativo de cada predictor en la clasificación final (Lundberg & Lee, 2017).

El sistema también incluye una función de priorización automática, donde las alertas son clasificadas por criticidad y urgencia, ayudando a los equipos institucionales a asignar los casos según su capacidad operativa. Esta funcionalidad se complementa con un módulo de exportación de reportes en formato PDF o CSV para su uso en comités o reuniones de seguimiento.

A partir de esta infraestructura, se formularon una serie de recomendaciones para asegurar la integración efectiva del sistema en

los procesos institucionales. Entre ellas destacan: (1) constituir comités de riesgo con representación interdisciplinaria; (2) capacitar al personal en el uso ético y crítico de la visualización; (3) establecer protocolos de actuación frente a cada nivel de alerta; y (4) actualizar el modelo predictivo y sus visualizaciones cada semestre para mantener su vigencia.

Estas recomendaciones no solo fortalecen la gestión académica con base en evidencia, sino que fomentan una cultura institucional orientada al cuidado, la equidad y la prevención educativa desde un enfoque tecnopedagógico y sensible al territorio.

## Capítulo 7. Recomendaciones, desafíos y aprendizajes

### Qué debe mejorar la Universidad del Pacífico

El proceso de diseño, desarrollo e implementación del sistema predictivo de deserción en la Universidad del Pacífico ha dejado en evidencia múltiples oportunidades de mejora institucional que trascienden el ámbito puramente tecnológico.

Una de las principales debilidades detectadas es la fragmentación de la información entre distintas dependencias administrativas, lo que dificulta una gestión integral, actualizada y en tiempo real de los datos relevantes para la permanencia estudiantil. Este tipo de segmentación informativa es común en instituciones con estructuras organizativas tradicionales, y constituye un obstáculo para la toma de decisiones basadas en evidencia (Ferguson, 2012).

La consolidación de un sistema de información académico robusto, interoperable y accesible para todos los actores institucionales debe ser una prioridad estratégica. Este sistema no solo debe centralizar los datos académicos, financieros y de bienestar, sino también

permitir su visualización dinámica, trazabilidad histórica y análisis multidimensional.

La implementación de arquitecturas orientadas a servicios (SOA) y el uso de estándares abiertos facilitaría la interoperabilidad con plataformas externas y entre dependencias internas (Romero & Ventura, 2013).

Además, es fundamental fortalecer las capacidades internas en ciencia de datos, inteligencia artificial y análisis institucional. Aunque el modelo predictivo ha sido desarrollado con apoyo externo y utilizando herramientas de código abierto, su sostenibilidad a largo plazo depende de la formación continua y especializada de personal administrativo, técnico y docente.

Esto requiere políticas de capacitación institucional, creación de semilleros de investigación en analítica educativa, y alianzas con universidades, centros de innovación y redes nacionales como Red de Analítica Académica de Colombia (RAAC).

Otra dimensión crítica es la consolidación de una cultura institucional centrada en el acompañamiento estudiantil basado en evidencia. La analítica predictiva no debe ser entendida solo como una herramienta técnica,

sino como un catalizador de transformaciones pedagógicas, sociales y organizativas.

Tal como han planteado Ifenthaler y Yau (2020), la utilidad de los sistemas predictivos se maximiza cuando se articulan con intervenciones concretas en áreas como salud mental, tutorías personalizadas, becas focalizadas y orientación vocacional.

Este cambio cultural debe ser promovido desde la alta dirección, transversalizado en los planes estratégicos institucionales, e interiorizado en todos los niveles de la comunidad universitaria.

En el caso particular de la Universidad del Pacífico —inserta en un territorio históricamente marginado, con población mayoritariamente afrodescendiente y rural— esta transformación debe estar guiada por principios de justicia social, equidad territorial y respeto por la diversidad cultural, reconociendo que los datos, cuando se usan con ética y sensibilidad, pueden ser herramientas poderosas de inclusión.

### **Datos que aún no se recolectan**

Aunque el sistema actual ha avanzado significativamente en la integración de variables académicas, sociodemográficas y financieras, persisten vacíos relevantes que limitan su

capacidad para capturar la complejidad de los factores que inciden en la permanencia estudiantil. En particular, no se recolectan datos relacionados con dimensiones psicosociales, culturales y afectivas, las cuales, según la literatura especializada, son determinantes en el comportamiento académico y la trayectoria vital del estudiante (Siemens & Long, 2011).

Aspectos como el nivel de estrés, la salud mental, las responsabilidades familiares, las trayectorias laborales paralelas, la percepción de discriminación o exclusión, y la calidad de las redes de apoyo emocional y académico, no están sistemáticamente registrados en los sistemas de información institucionales.

Estos factores, aunque subjetivos y de difícil cuantificación, influyen de manera crítica en la toma de decisiones del estudiante respecto a continuar o abandonar su formación.

Para suplir estos vacíos, es necesario adoptar metodologías mixtas que incluyan instrumentos cualitativos como encuestas periódicas, entrevistas semiestructuradas, grupos focales o análisis etnográfico digital.

Asimismo, pueden explorarse fuentes de información no tradicionales, como el análisis de interacción en entornos virtuales de aprendizaje (LMS), métricas de participación en actividades extracurriculares, o integración de registros de

programas estatales de protección social (Prinsloo & Slade, 2016).

Además, es imperativo consolidar marcos éticos robustos para la recolección y tratamiento de estos datos sensibles. Esto incluye el diseño de protocolos de consentimiento informado, mecanismos de anonimización, y la garantía de que la información recolectada se utilizará únicamente para fines de acompañamiento y mejora institucional, no para sanción ni exclusión.

La incorporación de principios como la minimización del daño, la justicia algorítmica y la explicabilidad debe ser un eje rector del sistema (Willis, Slade & Prinsloo, 2016).

Otro vacío significativo está relacionado con la ausencia de mecanismos para registrar y analizar fenómenos de microdeserción, como el abandono intermitente de asignaturas, el ausentismo parcial o los retrasos en los procesos de matrícula.

Estos eventos, muchas veces invisibilizados en los reportes de deserción formal, pueden ser señales tempranas de abandono definitivo y deben ser incorporados al sistema como indicadores de riesgo latente (Kuh et al., 2006).

Integrar estos datos en el modelo predictivo no solo mejoraría la precisión técnica, sino que

ampliaría la comprensión institucional sobre las múltiples formas que adopta la exclusión educativa en contextos vulnerables.

Potencial de replicabilidad en otras universidades públicas

Uno de los aprendizajes más valiosos de este proyecto es su potencial de réplica en otras instituciones de educación superior, especialmente en universidades públicas regionales que comparten características estructurales similares: altos niveles de vulnerabilidad socioeconómica, baja inversión tecnológica y limitada capacidad institucional para la toma de decisiones basadas en datos.

El enfoque metodológico adoptado — centrado en el uso de herramientas de código abierto, el diseño modular y la interoperabilidad con sistemas preexistentes como ACADEMUSOFT— convierte a esta experiencia en un referente adaptable y escalable (Macfadyen & Dawson, 2012).

La clave para garantizar la replicabilidad reside en lograr una articulación efectiva entre rigor técnico y sensibilidad institucional. La experiencia demuestra que un modelo predictivo no puede imponerse como un producto externo, sino que debe co-construirse con los actores locales, respondiendo a su

contexto sociocultural, organizativo y pedagógico.

En este sentido, se recomienda que todo proceso de implementación comience con un diagnóstico participativo, donde estudiantes, docentes, funcionarios administrativos y equipos de bienestar definan las prioridades, variables contextuales y condiciones operativas para el éxito del sistema (Ferguson et al., 2016).

Además, esta clase de sistemas no solo responde a un imperativo de eficiencia operativa, sino que puede ser integrado con las políticas de aseguramiento de la calidad, la autorregulación institucional y la evaluación externa exigidas por los organismos nacionales e internacionales de acreditación.

La capacidad de estos modelos para generar evidencia objetiva, trazable y replicable, los convierte en insumos valiosos para la rendición de cuentas y el diseño de estrategias de mejoramiento continuo (Ifenthaler & Yau, 2020).

La posibilidad de replicar esta iniciativa se ve fortalecida por el creciente ecosistema de comunidades académicas dedicadas a la analítica del aprendizaje y al uso ético de datos educativos. Organizaciones como SoLAR (Society for Learning Analytics Research) y redes latinoamericanas como RAAC (Red de

Analítica Académica de Colombia) ofrecen espacios para la cooperación interinstitucional, el intercambio de buenas prácticas y la consolidación de estándares técnicos y éticos.

Por tanto, el modelo desarrollado en la Universidad del Pacífico no solo representa un avance local, sino una oportunidad nacional. Su replicabilidad está condicionada no por limitaciones tecnológicas, sino por la voluntad institucional de abrazar una cultura de gestión basada en datos, de compromiso con la equidad educativa, y de apuesta por una universidad más inteligente, humana y territorialmente pertinente.

## Conclusión

### **El algoritmo como herramienta, no como oráculo**

El modelo predictivo de deserción estudiantil desarrollado en la Universidad del Pacífico no debe entenderse como un oráculo infalible, sino como una herramienta al servicio de la comprensión institucional y del diseño de políticas proactivas de permanencia. Su valor radica en su capacidad para sistematizar grandes volúmenes de datos, identificar patrones invisibles a simple vista, y orientar decisiones fundamentadas en evidencia empírica. Sin embargo, su poder predictivo no sustituye la complejidad del juicio humano ni las dinámicas sociales que atraviesan la vida estudiantil.

Como ha advertido la literatura crítica en el campo de la analítica del aprendizaje, la implementación de algoritmos en educación debe acompañarse de una reflexión ética profunda sobre sus alcances, límites y riesgos. Los modelos predictivos pueden amplificar sesgos existentes si no se diseñan y aplican con cuidado; también pueden generar estigmas si se transforman en mecanismos de clasificación inamovibles. Por ello, deben ser usados como brújulas, no como sentencias, y articulados con

espacios de deliberación pedagógica, acompañamiento humanizado y acción transformadora (Slade & Prinsloo, 2013).

Además, su implementación debe ir acompañada de una estrategia institucional que incluya capacitación permanente, revisión continua de las variables utilizadas, mecanismos de gobernanza de datos, y diálogo constante con los sujetos involucrados. El algoritmo no es un fin en sí mismo, sino un medio para potenciar el proyecto educativo, hacer más visible lo invisible y permitir que ninguna señal temprana de abandono pase desapercibida.

En contextos de alta vulnerabilidad, como el Pacífico colombiano, el uso de ciencia de datos debe contribuir a dignificar las trayectorias estudiantiles, no a reproducir la exclusión. Su verdadero aporte radica en traducir el conocimiento algorítmico en acción pedagógica y en abrir caminos institucionales que reconozcan a cada estudiante no como un dato aislado, sino como una historia en proceso.

### **La permanencia como un compromiso institucional y social**

La lucha contra la deserción no puede quedar confinada a los márgenes de una base de datos o al escritorio de un analista. Es, ante todo, un compromiso ético, político y social que interpela

a múltiples actores: las instituciones de educación superior, los gobiernos, las organizaciones de la sociedad civil, las familias y los propios estudiantes. La permanencia no es solo un indicador técnico ni una meta operativa; es una manifestación de justicia educativa, equidad territorial y derecho al futuro.

Acompañar a un estudiante hasta su graduación implica mucho más que monitorear su rendimiento académico. Significa construir entornos educativos que sean emocionalmente seguros, culturalmente pertinentes y económicamente viables.

Supone, además, transformar las condiciones estructurales que condicionan el abandono: la pobreza, la violencia, la desatención del Estado, la falta de conectividad y las múltiples formas de exclusión que afectan a los jóvenes en Colombia y América Latina (Tinto, 2012).

En contextos como el del litoral Pacífico, donde convergen la desigualdad étnica, la marginación geográfica y la fragilidad institucional, la permanencia estudiantil debe ser asumida como una causa colectiva y un objetivo de justicia redistributiva. No basta con contar con algoritmos precisos; se requiere de políticas afirmativas, acompañamientos psicosociales, apoyo económico sostenible y una voluntad institucional explícita de cuidar, incluir y transformar.

Las universidades, especialmente las públicas, deben reconocer que su rol va más allá de la formación académica. Son actores territoriales con capacidad de incidir en las dinámicas sociales y de reconfigurar oportunidades vitales.

Por ello, fortalecer sus capacidades analíticas debe ir de la mano con una renovación de su vocación transformadora. La ciencia de datos, cuando se pone al servicio de la equidad, puede ser una herramienta poderosa para visibilizar trayectorias ocultas, anticipar riesgos y abrir caminos de futuro.

Como han sostenido expertos en educación superior, la permanencia debe ser entendida como una construcción colectiva, resultado de la interacción entre condiciones institucionales, relaciones interpersonales y trayectorias individuales (Thomas, 2012). En ese sentido, los modelos predictivos no son fines en sí mismos, sino instrumentos que deben integrarse en un ecosistema de cuidado, corresponsabilidad y compromiso ético.

### **Próximos pasos: política pública basada en evidencia**

El trabajo realizado en este libro no solo ofrece una herramienta útil para la gestión institucional, sino una plataforma para la

transformación del sistema educativo superior desde una perspectiva estratégica, técnica y ética.

En un país como Colombia, donde las brechas territoriales, sociales y económicas siguen marcando el acceso y la permanencia en la educación superior, resulta urgente avanzar hacia políticas públicas que estén fundamentadas en evidencia robusta y que respondan a los contextos reales de las instituciones.

Es indispensable que el Estado colombiano — especialmente a través del Ministerio de Educación Nacional, el ICETEX, el DNP y el SENA— impulse el desarrollo de sistemas nacionales de monitoreo y alerta temprana que permitan anticipar y mitigar riesgos de deserción con enfoques diferenciales.

Estas plataformas deben estar articuladas con los sistemas de aseguramiento de la calidad y con las estrategias regionales de desarrollo educativo. En este sentido, es clave promover la interoperabilidad entre bases de datos, el fortalecimiento de la infraestructura tecnológica de las universidades, y la creación de estándares éticos para el manejo de datos educativos (Slade & Prinsloo, 2013).

Asimismo, se debe consolidar una política pública de permanencia que no se limite a

acciones reactivas o asistencialistas, sino que aborde las causas estructurales de la exclusión educativa. Esto implica articular programas intersectoriales entre educación, salud, transporte, conectividad, cultura y protección social.

También exige reconocer la diversidad de trayectorias, necesidades y desafíos de los estudiantes de zonas rurales, étnicas y periféricas, con políticas específicas que garanticen condiciones dignas para el estudio, incluyendo el acceso a vivienda, alimentación, acompañamiento psicosocial y servicios de salud mental (Tinto, 2012).

A nivel de gobernanza, es fundamental fomentar la colaboración entre universidades públicas, en particular las regionales, para el desarrollo de soluciones compartidas en analítica de datos, innovación pedagógica y evaluación institucional. El impulso de redes como la Red de Analítica Académica de Colombia (RAAC) puede contribuir a sistematizar experiencias, generar conocimiento colectivo y elevar el nivel técnico de los equipos institucionales.

Finalmente, es momento de consolidar un modelo de educación superior que valore el dato como insumo para la justicia, y no como un instrumento de control o estandarización. Un modelo donde la analítica educativa no sustituya

la relación humana, sino que la complemente; y donde el diseño algorítmico no reemplace el juicio pedagógico, sino que lo potencie.

Este libro es, en ese sentido, un llamado a construir desde la universidad un nuevo pacto educativo: más inclusivo, más ético, más inteligente. Un pacto donde los datos hablen, sí, pero donde se escuchen también las voces de los estudiantes. Un pacto donde el derecho a aprender sea garantizado no solo con diplomas, sino con trayectorias sostenidas, dignas y acompañadas hasta la graduación.



## Bibliografía

Agasisti, T., & Bowers, A. J. (2017). Data analytics and decision-making in education: Towards the educational data scientist. *European Journal of Education*, 52(4), 463–476. <https://doi.org/10.1111/ejed.12220>

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 61–75). Springer.

Baker, R. S., & Siemens, G. (2014). Educational data mining and learning analytics. In K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (2nd ed., pp. 253–274). Cambridge University Press.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

Brunner, J. J., & Ferrer, G. (2018). *Educación superior en América Latina: ¿Transformación o inercia?* Fondo de Cultura Económica.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357.

DANE. (2023). *Boletín técnico: Indicadores de pobreza multidimensional*. Departamento

Administrativo Nacional de Estadística.  
<https://www.dane.gov.co>

De la Cruz, R., & Pérez, J. (2019). Analítica de datos aplicada a la permanencia estudiantil: retos y oportunidades. *Revista Educación y Sociedad*, 30(1), 75–94.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37–54.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6), 304–317.  
<https://doi.org/10.1504/IJTEL.2012.051816>

Ferguson, R., Macfadyen, L., Clow, D., Tynan, B., Alexander, S., & Dawson, S. (2016). Setting learning analytics in context: overcoming the barriers to large-scale adoption. *Journal of Learning Analytics*, 3(1), 120–144.

González, M. A. (2021). *Diseño de un modelo predictivo para la prevención en la deserción de estudiantes de la Universidad del Pacífico* [Tesis de maestría, Universidad de San Buenaventura].

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.

Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*, 68, 1961–1990. <https://doi.org/10.1007/s11423-020-09788-z>

Jisc. (2016). *Code of Practice for Learning Analytics: A Literature Review and Framework*. <https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics>

MEN (Ministerio de Educación Nacional). (2023). *Sistema para la Prevención de la Deserción en Instituciones de Educación Superior (SPADIES)*. <https://www.mineduacion.gov.co>

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, 41(4), 1432–1462. <https://doi.org/10.1016/j.eswa.2013.08.042>

Prinsloo, P., & Slade, S. (2016). Student vulnerability, agency, and learning analytics: An exploration. *Journal of Learning Analytics*, 3(1), 159–182.

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data*

*Mining and Knowledge Discovery*, 10(3), e1355.  
<https://doi.org/10.1002/widm.1355>

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5), 30–40.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510–1529.  
<https://doi.org/10.1177/0002764213479366>

Thomas, L. (2012). *Building student engagement and belonging in Higher Education at a time of change: Final report from the What Works? Student Retention & Success programme*. Higher Education Academy.

Tinto, V. (2012). Enhancing student success: Taking the classroom seriously. *The International Journal of the First Year in Higher Education*, 3(1), 1–8.  
<https://doi.org/10.5204/intjfyhe.v3i1.119>

UNESCO IESALC. (2022). *Informe sobre la educación superior en América Latina y el Caribe*. Instituto Internacional para la Educación Superior en América Latina y el Caribe.  
<https://www.iesalc.unesco.org>

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology*, 45(3), 223–235.

<https://doi.org/10.1080/17439884.2020.1798995>



## Sobre el autor

**Manuel Alberto González González** es Ingeniero de Sistemas, Especialista en Administración de la Información y Bases de Datos, así como en Procesos de Desarrollo de Software. Su formación académica se consolida con una Maestría en Ingeniería de Software, que le ha permitido desarrollar una perspectiva amplia, crítica y propositiva sobre los retos y oportunidades del mundo digital.

A lo largo de más de una década, ha ejercido como docente universitario, acompañando la formación de nuevas generaciones de profesionales en áreas clave como la ingeniería, la tecnología y el desarrollo de sistemas. En el aula, ha sabido combinar la rigurosidad técnica con la sensibilidad pedagógica, formando no solo ingenieros competentes, sino también pensadores comprometidos con el desarrollo de soluciones éticas, sostenibles y contextualmente relevantes.

Paralelamente, su trayectoria en el sector productivo ha sido igual de sólida: más de diez años de experiencia profesional le han

permitido participar en múltiples proyectos de diseño, implementación y gestión de soluciones informáticas aplicadas a sectores organizacionales diversos. Esta doble pertenencia —a la academia y a la industria— ha nutrido su visión integral del quehacer tecnológico, articulando teoría y práctica con una mirada crítica y propositiva.

Este libro es una muestra clara de su compromiso con el conocimiento situado y útil. Nace de la preocupación por las altas tasas de deserción en la educación superior, pero también de la convicción de que los algoritmos y la ciencia de datos, usados con ética y propósito, pueden ser herramientas poderosas para transformar esa realidad. Así, su obra no solo ofrece soluciones técnicas, sino que plantea un enfoque de ingeniería de software con sentido social, orientado al fortalecimiento de trayectorias educativas y a la promoción de la equidad en el acceso y permanencia universitaria.