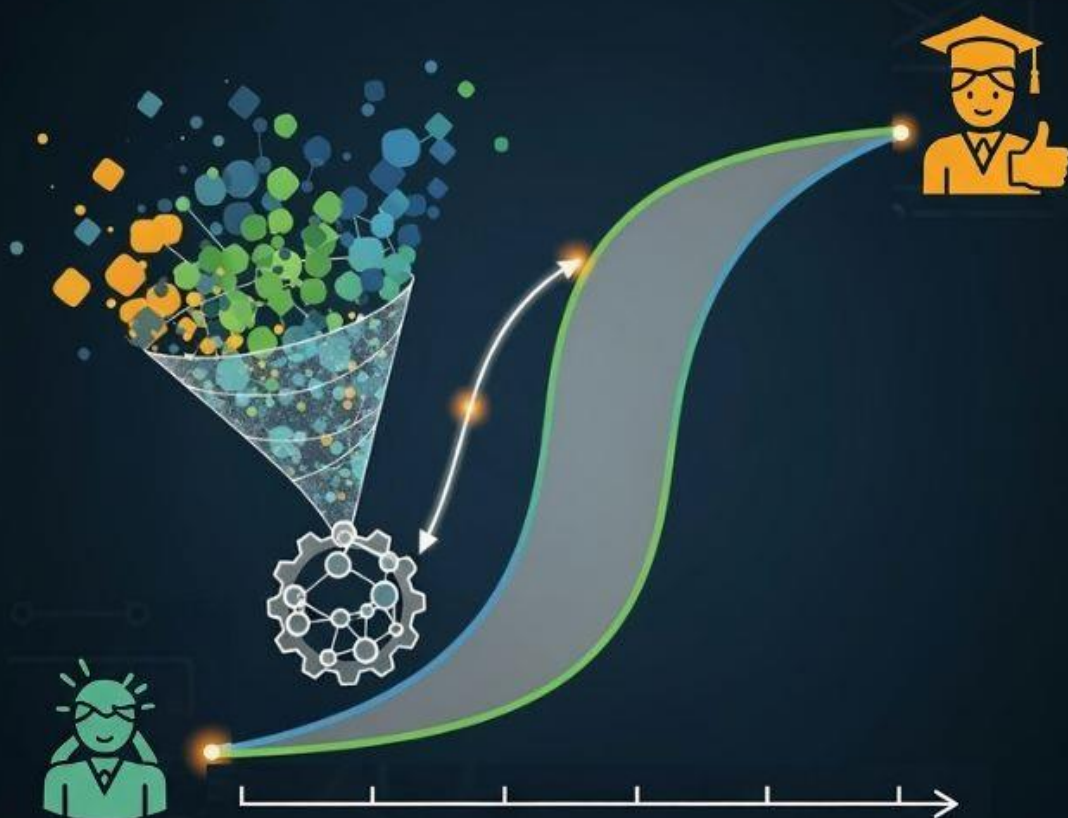



# Más Allá

# de las Calificaciones

Modelos Predictivos para el  
Rendimiento Académico en Ingeniería



  
Sello  
Editorial **CITSA**

Casio Aurelio Torres López

Jorge Vladimir Pachas Huaytan

Jaime Humberto Ortiz Fernández

Javier Amador Navarro Veliz

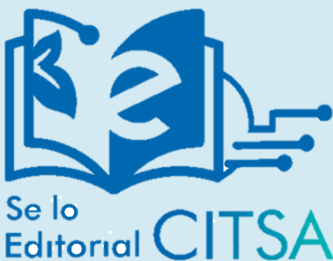
Jesús Armando Cavero Carrasco

Richard Víctor Diaz Urbano

Casio Aurelio Torres López  
Jorge Vladimir Pachas Huaytan  
Jaime Humberto Ortiz Fernández  
Javier Amador Navarro Veliz  
Jesús Armando Cavero Carrasco  
Richard Víctor Díaz Urbano

# *Más Allá de la Calificaciones:* Modelos Predictivos para el Rendimiento Académico en Ingeniería

<https://doi.org/10.61286/edcitsa.vi2.314>



Maracay estado Aragua, 2025

## Catalogación en Fuente

Torres López Casio Aurelio, Pachas Huaytan Jorge Vladimir, Ortiz Fernández Jaime Humberto, Navarro Veliz Javier Amador, Cavero Carrasco Jesús Armando y Diaz Urbano Richard Víctor.

Más Allá de la Calificaciones: Modelos Predictivos para el Rendimiento Académico en Ingeniería. 1ª ed.

– Maracay: Sello Editorial CITSA, 2025.

Recursos en línea (85 páginas); il. 9; 14,8 x 21,0 cm.

ISBN: 978-980-8050-19-6

1. Educación Superior
2. Evaluación del Rendimiento Académico

CDD 378.166.

### Sello Editorial CITSA

**Centro de Investigación en Tecnologías de Salud y Ambiente. Dirección:** Calle el Stadium N° 3-A, Las Brisas, La Pedrera, Parroquia Las Delicias, Maracay estado Aragua, Venezuela.

Email: [citsa@investigaciondetecnologias.com](mailto:citsa@investigaciondetecnologias.com)

Web: [www.investigaciondetecnologias.com](http://www.investigaciondetecnologias.com)

Coordinación Editorial: Dr. José Romero

Revisión y corrección de estilo: Dra. Deneice Montesinos Dalis

Diseño de cubierta: Dr. José Romero

Composición y puesta en línea: MSc. Vita María Calzolaio Cristofano

Depósito Legal en la Biblioteca Nacional de Venezuela según el Número AR2026000107



Más Allá de la Calificaciones: Modelos Predictivos para el Rendimiento Académico en Ingeniería tiene licencia CC BY-NC-ND 4.0. © 2 por Casio Aurelio Torres López, Jorge Vladimir Pachas Huaytan, Jaime Humberto Ortiz Fernández, Javier Amador Navarro Veliz, Jesús Armando Cavero Carrasco y Richard Víctor Diaz Urbano.

# Autores

## Casio Aurelio Torres Lopez

E-mail: [d.torres@ms.upla.edu.pe](mailto:d.torres@ms.upla.edu.pe)

<https://orcid.org/0000-0003-3637-5093>



Doctor en Economía por la Universidad Nacional Federico Villarreal (UNFV), Magister en Economía, mención “Métodos Cuantitativos de la Economía” por la Universidad Nacional Mayor de San Marcos (UNMSM), Economista por la Universidad Nacional del Centro del Perú (UNCP), Licenciado en Matemáticas - Física (ENSTP), Docente Universitario en Pregrado y Posgrado (maestría y doctorado) en las universidades: UNCP (Huancayo), UPLA (Huancayo), UNDAC (Cerro de Pasco), UNHV (Huánuco), UNH (Huancavelica) y UNSCH (Ayacucho), Miembro de la Sociedad Interamericana de Evaluadores de Políticas y Proyectos SIEP (UNIANDÉS, Bogotá Colombia), Asesor, Consultor y Conferencista, autor de libros de Estadística para la Investigación y de Investigación de Operaciones, publicaciones de artículos en Scielo, autor de varias patentes; haber ejercido cargos de Vicerrector Administrativo, Decano de la Facultad de Ingeniería, Director de Departamento Académico de Ciencias e Ingeniería, Director de Estudios Generales, Director del Consejo de Investigación, Director de Investigación de la Escuela de Posgrado, Jefe de la Oficina de Planificación y otros en al Universidad Peruana Los Andes.

## Jorge Vladimir Pachas Huaytan

E-mail: [d.jpachas@upla.edu.pe](mailto:d.jpachas@upla.edu.pe)

<https://orcid.org/0000-0003-4745-669X>



Es Ingeniero de Sistemas y Computación, con una segunda especialidad en Administración de Empresas. Posee una sólida formación de posgrado que incluye una Maestría en Ingeniería de Sistemas con mención en Gerencia de Tecnologías de Información y Comunicación, una Maestría en Administración con mención en Gestión de Proyectos, y un Doctorado en Educación. Actualmente se desempeña como docente e investigador en la Universidad Peruana Los Andes, donde lidera proyectos orientados a la integración de las tecnologías de información, la gestión por procesos y la mejora continua de la calidad educativa. Su experiencia profesional abarca la aplicación de metodologías ágiles como ITIL, BPM y Scrum, la gestión de proyectos bajo el enfoque del PMBOK, y el análisis estadístico avanzado para la toma de decisiones institucionales. Su visión estratégica y enfoque científico lo consolidan como un referente en innovación académica, transformación digital y gestión educativa en el contexto universitario.

## Jaime Humberto Ortiz Fernandez

e-mail: [d.jortiz@ms.upla.edu.pe](mailto:d.jortiz@ms.upla.edu.pe)

ORCID: <https://orcid.org/0000-0001-7258-9975>



Doctor en Educación Universidad Peruana Los Andes (UPLA-Perú), Magister en Informática Pontificia Universidad Católica del Perú (PUCP-Perú), Maestro en Energías Renovables y Proyectos Energéticos Universidad Rey Juan Carlos-España, Ingeniero Electricista Universidad Nacional del Centro del Perú (UNCP-Perú), docente universitario en la Universidad Peruana Los Andes en el programa de Ingeniería de Sistemas y Computación. Asesor y consultor independiente en redes de datos y seguridad informática. Autor de libros, capítulos de libro y artículos científicos en revistas indizadas SCIELO y Latindex. Actualmente Director de la Unidad de Posgrado de la Facultad de Ingeniería de la Universidad Peruana los Andes.

## Javier Amador Navarro Veliz

Email: [d.jnavarro@ms.upla.edu.pe](mailto:d.jnavarro@ms.upla.edu.pe)

<https://orcid.org/0000-0001-5293-0240>



Es Ingeniero Geólogo de la UNDAC, con Maestrías en Geotecnia por la UNMSM y en Gestión Ambiental y Desarrollo Sostenible por la UAP, así como Doctor en Educación y Especialista en Didáctica Universitaria por la UPLA. Posee amplia experiencia académica y administrativa, desempeñándose como Jefe del Departamento Académico y docente investigador en la Facultad de Ingeniería de la UPLA, donde también ha dirigido los laboratorios de Geología Aplicada, Topografía, Astronomía y Mecánica de Suelos. Ha sido docente en la UAP en la carrera de Ingeniería Civil y ha liderado diversos estudios geológicos y geotécnicos, entre ellos el Proyecto de Irrigación del Distrito de Quilcas, la Zonificación de Áreas Críticas de la Subcuenca del Shullcas y la Central Hidroeléctrica Llusita en Ayacucho.

Ha colaborado con entidades estatales como CORDE Huancavelica, CORDE Junín y la UNCP, así como con empresas privadas del sector minero y de ingeniería. Es autor de múltiples publicaciones científicas y manuales académicos, entre ellos Tecnología de los Materiales, Dinámica, Introducción a la Mecánica de Suelos y los Manuales de Geología I y II Aplicada a la Ingeniería Civil, editados por la UPLA.

## Jesús Armando Cavero Carrasco

Email: [d.jcavero@ms.upla.edu.pe](mailto:d.jcavero@ms.upla.edu.pe)

<https://orcid.org/0000-0003-0122-1166>



Doctor en Educación por la Universidad Nacional Federico Villarreal (UNFV), Magíster en Educación con Mención en Andragogía por la Universidad Nacional Federico Villarreal (UNFV), Grado de Bachiller en Ciencias de la Educación por la Universidad Nacional de San Cristóbal de Huamanga (UNSCH), Título Profesional de Licenciado en Educación y Ciencias Humanas, especialidad Ciencias Sociales, Filosofía y Psicología (UNSCH). Docente en Ciencias Sociales en la Universidad Nacional de San Cristóbal de Huamanga (27 años), Docente en Pedagogía e Investigación Educativa en la Universidad Peruana Los Andes (35 años), Profesor Emérito de la Universidad Nacional de San Cristóbal de Huamanga. Experiencia en Gestión Universitaria: Decano de Facultad de Educación y Ciencias Humanas en 5 oportunidades (UNSCH-UPLA), Vicerrector Académico de la Universidad Peruana Los Andes, Secretario Académico de la Escuela de Posgrado de la UPLA, Director (e) de la Escuela de Posgrado de la UPLA. Coautor de libros

sobre Educación en la UNSCH y la UPLA (6 libros). Ponente en Congresos Nacionales e Internacionales.

## Richard Victor Diaz Urbano

E-mail: [d.rdiaz@ms.upla.edu.pe](mailto:d.rdiaz@ms.upla.edu.pe)

<https://orcid.org/0000-0001-7338-9625>

Doctor en Ciencias Contables y Empresariales  
Universidad Peruana Los Andes



Doctor en Ciencias Contables y Empresariales de la Escuela de Post Grado de la Universidad Peruana Los Andes; Maestro en Política Fiscal y Tributación de la Universidad Particular Inca Garcilazo de La Vega; Economista, egresado de la Universidad Nacional del Centro del Perú; Contador Público, egresado de la Universidad Peruana Los Andes. Funcionario, Profesional Especializado V de la Intendencia de Tributos Internos Junín (SUNAT); desempeño cargos de Directivo como Jefe de la División de Auditoría, Supervisor Profesional y Auditor de SUNAT. Docente Universitario de Pregrado, Post Grado y Curso Taller de Investigación en la Universidad Peruana Los Andes. Docente Universitario en la Escuela de Post Grado de la Universidad Nacional del Centro del Perú, facultad de Contabilidad. Docente Universitario en la Escuela de Post Grado de la Universidad Nacional Hermilio Valdizán de Huánuco.

# Prólogo

En el siglo XXI, la excelencia en la educación en ingeniería ha cambiado de forma radical. Ya no se trata solo de enseñar tecnología avanzada, sino de asegurarse de que cada estudiante logre terminar su carrera con éxito. Las facultades de ingeniería, que deberían ser espacios de innovación y pensamiento estratégico, muchas veces siguen usando métodos obsoletos para evaluar y apoyar a sus estudiantes. Se enfocan en lo que ya pasó, en lugar de anticipar lo que podría pasar. Es como si intentaran arreglar una máquina solo después de que se rompe, en vez de prevenir la falla desde el diseño.

Este libro nace de una idea sencilla pero poderosa: si la ingeniería puede predecir el colapso de un puente o el rendimiento de una turbina, también puede anticipar el éxito o el fracaso académico. Aplicar ese mismo rigor a la trayectoria de los estudiantes no solo es posible, sino urgente. Las notas por sí solas no cuentan toda la historia. Evaluar el riesgo de abandono o bajo rendimiento solo con el promedio académico es como mirar el cielo y asumir que mañana hará sol. Es una forma limitada de entender un sistema mucho más complejo.

El rendimiento académico es un fenómeno multifactorial. Está influido por el entorno social, económico, emocional y por cómo el estudiante se relaciona con su carrera desde el primer día. Entender estos factores es el primer paso para construir modelos predictivos que realmente funcionen. No se trata de adivinar, sino de usar datos y análisis para tomar decisiones informadas. La estadística predictiva, lejos de ser una herramienta fría, se convierte en el motor que permite actuar antes de que aparezca el problema.

Este enfoque requiere una metodología clara. Desde recolectar y limpiar los datos, hasta construir modelos que puedan anticipar riesgos con precisión. Es un proceso técnico, sí, pero también profundamente humano. Porque detrás de cada número hay una historia, y detrás de cada predicción hay una oportunidad de intervenir a tiempo. El objetivo no es solo saber quién podría reprobar, sino entender por qué y qué se puede hacer para evitarlo.

Para demostrar que esto no es solo teoría, el libro presenta un caso práctico. Se usa un conjunto de datos simulados de una carrera de ingeniería y se aplica todo el proceso, paso a paso, con herramientas reales como Python. El resultado es un modelo funcional que identifica estudiantes en riesgo antes de que lleguen al punto de quiebre. Esto permite actuar con precisión: tutorías específicas, ajustes en la carga académica, mentorías, todo basado en evidencia.

Al final, lo que se propone aquí no es solo una nueva forma de analizar datos, sino una nueva forma de pensar la educación. La predicción académica no es un fin en sí mismo, sino una herramienta para transformar la experiencia estudiantil. Permite que las instituciones pasen de reaccionar ante el fracaso a diseñar el éxito desde el inicio. Y eso, en el fondo, es ingeniería en su forma más pura: anticipar, optimizar, construir.

Este esfuerzo es, en esencia, una invitación a la innovación responsable. Abogamos por un futuro en el que la retención, la eficiencia terminal y la calidad educativa se optimicen a través de la inteligencia artificial y el machine learning, siempre al servicio de la equidad y el máximo potencial de cada estudiante. Al dotar a los líderes educativos de sistemas de alerta temprana, no solo mitigamos el fracaso; empoderamos a los estudiantes al proporcionarles el apoyo preciso, en el momento preciso, transformando el camino a la graduación de un laberinto incierto a un trayecto guiado y asistido.

A los líderes institucionales, a los decanos, a los directores de programas y a los analistas de datos: este libro es su manual para construir un sistema más justo, más eficiente y, sobre todo, más humano. La proactividad, la precisión y el rigor analítico que exige la ingeniería son ahora las virtudes que deben definir la gestión del éxito académico. Les invito a pasar Más Allá de la Calificaciones, a abrazar la Ingeniería de la Predicción Académica y a transformar el futuro de la educación superior en ingeniería.

**¡Bienvenidos a la era de la gestión académica predictiva!**

# Índice

	Página
Prólogo	
Introducción	1
Capítulo 1. Fundamentos del Rendimiento Académico y la Predicción en la Educación Superior en Ingeniería	4
• El Rendimiento Académico como Fenómeno Multifactorial	4
• La Predicción del Rendimiento en Ingeniería: Un Enfoque Proactivo	5
○ Minería de Datos Educativa (EDM)	5
• Los modelos de regresión como pilar de la predicción	6
○ Regresión Lineal Simple y Múltiple	6
○ Regresión Logística	7
• El ciclo de vida de un modelo predictivo	7
○ Definición del problema	7
○ Recolección y preparación de datos	7
○ Selección y entrenamiento del modelo	7
○ Evaluación y validación	7
Capítulo 2. Introducción a la Estadística Predictiva y los Modelos de Regresión	17
• La Estadística Predictiva: Previendo el futuro con datos	17
• Modelos de Regresión Lineal (Simple y Múltiple)	19
• Regresión Logística: Modelización de Resultados Binarios (Éxito/Fracaso)	25
Capítulo 3. Metodología para la Construcción de un Modelo Predictivo	30
• Definición del Problema y de la Variable de Estudio	30
• Recolección, Limpieza y Transformación de Datos	31
<b>Análisis Bivariado: La búsqueda de relaciones</b>	32
<b>Modelado Multivariado</b>	34
Capítulo 4. Aplicación Práctica: Un Caso de Estudio en Ingeniería	37
<b>Method (Método Utilizado)</b>	41
<b>Discusión y conclusiones</b>	49
Capítulo 5. Implicaciones y Futuras Direcciones de la Regresión Predictiva en la Educación	51
<b>Uso del modelo para la intervención: De la predicción a la acción</b>	51
<b>Limitaciones y futuras líneas de investigación</b>	53
<b>Epílogo: La Estadística al Servicio de la Educación</b>	56
<b>Diseño de un programa de tutoría basado en el modelo predictivo</b>	57
<b>Consideraciones finales</b>	70
<b>Referencias</b>	71
<b>Glosario</b>	82
<b>Conjunto de Datos de Muestra</b>	83
<b>Código de Software</b>	84
<b>Código en Python</b>	85

# *I*ntroducción

En la actualidad, la educación superior enfrenta múltiples desafíos, entre los cuales destaca la necesidad de mejorar el rendimiento académico de los estudiantes. Los modelos predictivos, basados en técnicas de aprendizaje automático, ofrecen herramientas valiosas para identificar factores que influyen en el desempeño académico y predecir resultados futuros. Este documento presenta un ejercicio corto utilizando Python para desarrollar un modelo predictivo del rendimiento académico de estudiantes de ingeniería.

El uso de modelos predictivos ha cobrado mayor relevancia en diversos campos, incluida la educación. Los modelos predictivos son herramientas estadísticas que permiten anticipar resultados futuros basándose en datos históricos. En el contexto del rendimiento académico, estos modelos pueden ayudar a identificar factores que impactan el desempeño de los estudiantes de ingeniería, facilitando así la implementación de estrategias para mejorar su rendimiento en el exigente y riguroso panorama de la Educación Superior en Ingeniería, nuestra misión fundamental es diseñar sistemas que garanticen la excelencia. Con esa mentalidad analítica, debemos abordar el desafío más persistente y costoso de nuestras facultades: el Rendimiento Académico y la prevención del fracaso estudiantil.

Durante demasiado tiempo, hemos confiado en métricas simplistas como el Promedio de Calificaciones (GPA). Sin embargo, la complejidad de la formación en ingeniería —con su alta demanda de resiliencia, aptitud y el efecto dominó de sus currículos secuenciales— nos obliga a ir "más allá de las calificaciones". El éxito o la deserción de un estudiante es un fenómeno multifactorial, el resultado de una compleja interacción de variables personales, académicas, socioeconómicas e institucionales.

Si la ingeniería se define por su capacidad de anticipar y mitigar fallas, es imperativo que apliquemos esa misma disciplina y rigor analítico a la gestión de nuestros propios procesos de formación. En el Capítulo 1 sentamos las bases teóricas del rendimiento académico como un fenómeno multifactorial. Ahora, en el Capítulo 2, pasaremos de la conceptualización a la ejecución práctica y el diseño de modelos.

La estadística predictiva y los modelos de regresión no es un mero ejercicio teórico de estadística; es una hoja de ruta para la ingeniería de datos aplicada a la educación. dejaremos atrás la cómoda posición de solo describir el fracaso *¿Qué pasó?* para asumir la posición activa de preverlo *¿Qué pasará si no actuamos?*

Nuestra meta es clara: utilizar el poder de la Estadística Predictiva para transformar los datos históricos (calificaciones de ingreso, horas de estudio, autoeficacia, etc.) en pronósticos informados. Esto nos permite dirigir recursos y tutorías de manera eficiente y justa, identificando a los estudiantes con el mayor riesgo de deserción o bajo rendimiento antes de que el problema se materialice.

Aquí, el modelo de regresión se presenta como la espina dorsal de esta capacidad predictiva. Exploraremos dos tipos fundamentales:

- **Regresión lineal simple y múltiple:** Utilizada para predecir resultados continuos (como el GPA final o la nota exacta en Termodinámica) y crucial para cuantificar la contribución única de cada factor (p. ej., demostrar que la autoeficacia puede tener un peso predictivo mayor que la calificación en Cálculo I).
- **Regresión logística:** La herramienta esencial para modelar resultados binarios (éxito/fracaso, abandono/permanencia). Su fortaleza radica en su capacidad para traducir un conjunto complejo de factores en una probabilidad de riesgo clara e interpretable, crucial para la toma de decisiones institucionales.

Al dominar estos modelos, no solo mejoramos las tasas de retención de la universidad, sino que preparamos a nuestros futuros ingenieros para enfrentar problemas reales donde la incertidumbre, la modelización y la predicción de eventos discretos son parte central del diseño y la operación de sistemas complejos.

Los fundamentos del rendimiento académico y la predicción en la educación superior en ingeniería, sienta las bases teóricas y metodológicas para transformar la reacción ante el fracaso en una estrategia de prevención proactiva.

Comprender la ciencia detrás del éxito académico es el primer paso riguroso para rediseñar la trayectoria de nuestros estudiantes. Los invito a sumergirse en estos fundamentos: es la única forma de garantizar que nuestras instituciones formen a los ingenieros que el futuro tecnológico demanda, con la máxima eficiencia y el menor índice de falla.

El diseño de un **modelo predictivo de rendimiento académico** no es un acto intuitivo, sino una rigurosa Metodología que demanda la precisión, el control de calidad y la iteración que aplicamos en cualquier proyecto de ingeniería. El éxito de la predicción, y por ende de la

La capacidad de detectar a tiempo a un estudiante en riesgo es nuestra mayor palanca para mejorar la retención y el éxito. En este ejercicio, pasaremos de la tabla de datos brutos a la construcción de dos tipos de modelos cruciales utilizando el entorno de Python:

1. **Predicción continua (Regresión):** Estimaremos la nota final o el rendimiento numérico en una asignatura crítica, demostrando cómo variables como las horas de estudio influyen directamente en la calificación.
2. **Clasificación de riesgo (Regresión Logística):** Nos enfocaremos en el problema binario más crítico: predecir la probabilidad de bajo rendimiento o

deserción (*Sí/No*). Analizaremos un caso de estudio real que subraya cómo el Promedio de Bachillerato y la Asistencia son predictores más fuertes que la Zona de Origen, permitiendo a las instituciones enfocar sus recursos de intervención de manera quirúrgica.

El éxito académico es un fenómeno multifactorial que va "**más allá de las calificaciones**"; requiere modelar variables académicas, personales y socioeconómicas.

Finalmente, los ejercicios prácticos en Python le mostrarán cómo construir estos modelos, permitiéndole identificar con antelación a estudiantes de alto riesgo. El objetivo es transformar la gestión educativa en un proceso estratégico, eficiente y basado en evidencia para optimizar el talento de nuestros futuros ingenieros.

La verdadera utilidad de un modelo predictivo, como la Regresión Logística que hemos construido, no es solo su número de exactitud; reside en su capacidad para catalizar el cambio educativo proactivo. En el riguroso entorno de la ingeniería, no podemos esperar a que el estudiante fracase para reaccionar.

Se explora las implicaciones educativas profundas que se derivan de nuestros hallazgos analíticos:

- **Intervención temprana y focalizada:** La predicción nos permite identificar estudiantes con una alta probabilidad de riesgo (ej., probabilidad de bajo rendimiento superior al 60%) y segmentarlos según la naturaleza de su desafío. Ya no ofrecemos tutorías genéricas; ofrecemos tutorías intensivas en fundamentos (si el problema es un GPA de bachillerato bajo) o mentoría en gestión del tiempo (si el problema es la baja asistencia).
- **Mejora curricular basada en evidencia:** Los coeficientes del modelo nos informan sobre la efectividad de los pre-requisitos y los factores de éxito, guiando la revisión curricular y la asignación estratégica de recursos (ej., invertir más en laboratorios si el *engagement* es un predictor fuerte).
- **Desafíos y ética:** Abordaremos las limitaciones del modelo (ej., no capturar variables cualitativas como la resiliencia o el apoyo familiar) y la necesidad de una gobernanza ética de datos. El modelo es una señal de alerta, no un veredicto, y debe aplicarse con equidad y transparencia.
- En resumen, la **Estadística Predictiva** se convierte en un motor de transformación institucional, permitiendo a las facultades de ingeniería optimizar el rendimiento y la retención del talento mediante estrategias personalizadas, precisas y científicamente fundamentadas.

# Capítulo 1

## Fundamentos del Rendimiento Académico y la Predicción en la Educación Superior en Ingeniería

El rendimiento académico en la educación superior, particularmente en el ámbito de la ingeniería, ha sido tradicionalmente medido a través de métricas cuantitativas como el promedio de calificaciones (**GPA**). Sin embargo, esta visión simplista es insuficiente. Como se discute a lo largo de este libro, la verdadera comprensión del éxito educativo va "**más allá de las calificaciones**", abarcando un espectro más amplio de habilidades, conocimientos, actitudes y valores (Gutiérrez-Monsalve et al., 2021). Este capítulo se adentra en los fundamentos teóricos y las complejidades multifactoriales que definen el rendimiento académico en la ingeniería, así como en las técnicas emergentes, como la minería de datos, que permiten una predicción más holística y temprana de los resultados estudiantiles.

### El rendimiento académico como fenómeno multifactorial

El éxito en las carreras de ingeniería no es el resultado de un solo factor, sino de una interacción compleja de variables personales, académicas, institucionales y socioeconómicas. La investigación ha destacado consistentemente la importancia de los siguientes elementos:

1. **Factores personales y psicológicos:** La inteligencia, las aptitudes, la motivación, la autoeficacia y el bienestar psicológico son cruciales (Zapata et al., 2021). La teoría del rendimiento académico de Walberg, por ejemplo, postula que las características psicológicas de los estudiantes y sus entornos inmediatos influyen en los resultados educativos (Pedraja-Rejas et al., 2025). En ingeniería, la **resiliencia** ante la frustración y el estrés académico también juegan un papel significativo (Zapata et al., 2021).
2. **Factores académicos y de hábitos de estudio:** Las experiencias educativas previas, las habilidades cognitivas, los **hábitos de estudio** y la **gestión del tiempo** impactan directamente en el desempeño. Los estudiantes de ingeniería exitosos no solo memorizan fórmulas, sino que desarrollan estrategias de estudio efectivas y aprenden a gestionar la carga de trabajo, a menudo exigente (Arias Ospina et al., 2024).
3. **Factores institucionales y contextuales:** El entorno universitario, incluyendo la **calidad de la enseñanza**, la infraestructura, los métodos de evaluación y las interacciones con el profesorado, influyen en el rendimiento (Torres Cardona.,

2023). En la ingeniería, donde la práctica y el **trabajo colaborativo** son esenciales, la calidad de los laboratorios y el apoyo institucional son determinantes.

4. **Factores socioeconómicos y familiares:** El contexto socioeconómico y familiar puede generar ventajas o desventajas que afectan el desempeño académico. Las condiciones de vida, el acceso a recursos y el apoyo familiar pueden influir en el bienestar psicológico y, por ende, en la capacidad del estudiante para concentrarse y prosperar en sus estudios (Castro Méndez, 2024).

### La Predicción del rendimiento en ingeniería: Un enfoque proactivo

La capacidad de **predecir** el rendimiento académico de los estudiantes de ingeniería es una herramienta invaluable para las instituciones educativas, permitiendo la implementación de **intervenciones tempranas y personalizadas** para apoyar a aquellos en riesgo de fracaso (Rico Páez & Gaytán Ramírez, 2022). En este contexto, la **Minería de Datos Educativa (EDM)** ha surgido como un campo de investigación prometedor (Lledó et al., 2024).

- ✓ **Minería de datos educativa (EDM):** Al aplicar técnicas de minería de datos, como la **regresión**, la clasificación y el agrupamiento, a bases de datos educativas, es posible construir **modelos predictivos** del desempeño estudiantil (Rico Páez & Gaytán Ramírez, 2022). Estos modelos utilizan información histórica, como calificaciones de admisión, notas semestrales y datos demográficos, para identificar patrones y prever resultados futuros. El modelo predictivo es un **modelo estadístico** que evalúa datos, detecta patrones y utiliza esa información para predecir tendencias futuras o objetivos centrales, en primer lugar, realizar una evaluación y comparación de las metodologías estadísticas y de aprendizaje automático más frecuentemente empleadas en la predicción del rendimiento académico, con el objetivo de discernir sus puntos fuertes y áreas de mejora (Álvarez-Yaulema et al., 2024).
- ✓ **Variables predictivas en ingeniería:** La investigación ha identificado una serie de variables específicas que son particularmente útiles para predecir el rendimiento de los estudiantes de ingeniería. Estas incluyen el **promedio de calificaciones de ingreso**, las notas en asignaturas específicas de las primeras etapas del plan de estudios, el **número de asignaturas reprobadas** y los hábitos de estudio (Castrillón et al., 2020).
- ✓ **Algoritmos de predicción:** Se han utilizado y comparado diversos **algoritmos de aprendizaje automático** para predecir el rendimiento académico (Rico Páez & Gaytán Ramírez 2022). La selección del algoritmo en modelos predictivos para el rendimiento académico de estudiantes de ingeniería es fundamental, ya que influye en la precisión de las predicciones. Algoritmos como la regresión logística, árboles de decisión y máquinas de soporte vectorial son comúnmente utilizados, cada uno con sus propias ventajas según el tipo de datos y el objetivo del análisis (Castrillón et al., 2020).

La selección del algoritmo en modelos predictivos para el rendimiento académico de estudiantes de ingeniería es crucial para obtener resultados precisos. Algoritmos como regresión logística, árboles de decisión y máquinas de soporte vectorial son frecuentemente utilizados, cada uno adaptándose a diferentes tipos de datos y objetivos. (López, 2021).

El rendimiento académico en la educación superior en ingeniería es un fenómeno complejo, influenciado por múltiples factores (López, 2019). Hoy en día es posible afirmar, con un alto grado de confianza, que es impensable encontrar un programa de Ingeniería donde no se incluya el estudio de muchos conceptos relacionados con la probabilidad y la estadística. Las razones para su inserción están directamente relacionadas con la necesidad de tomar decisiones en situaciones de incertidumbre (Pérez & Salamanca, 2013). A diferencia de la estadística descriptiva, que se limita a resumir y describir conjuntos de datos, la estadística predictiva va un paso más allá, centrándose en la inferencia y la previsión (García et al., 2020). El uso de técnicas de aprendizaje automático y minería de datos en ingeniería permite descubrir patrones ocultos y tendencias en grandes volúmenes de información. Estas herramientas son esenciales para mejorar la toma de decisiones y optimizar procesos en diversas áreas de la ingeniería (Zhang et al., (2019).

Para el contexto de la ingeniería, esto significa que podemos tomar las diversas variables identificadas como influyentes—como el promedio de notas de ingreso, el rendimiento en asignaturas clave o el contexto socioeconómico—y usarlas para construir un modelo que anticipe el desempeño futuro de un estudiante (Pérez, 2023).

### Los modelos de regresión como pilar de la predicción

- ✓ Dentro del vasto arsenal de herramientas de la estadística predictiva, los modelos de regresión son, sin duda, uno de los pilares fundamentales, especialmente cuando el objetivo es predecir una variable continua, como el promedio de calificaciones final. Según García, (2023) dentro del vasto arsenal de herramientas de la estadística predictiva, los modelos de regresión son, sin duda, uno de los pilares fundamentales, especialmente cuando el objetivo es predecir una variable continua, como el promedio de calificaciones final. La regresión es un proceso estadístico para estimar las relaciones entre variables, permitiendo cuantificar la influencia de una o más variables independientes sobre una variable dependiente (Montgomery et al., 2012). Dentro del vasto arsenal de herramientas de la estadística predictiva, los modelos de regresión son, sin duda, uno de los pilares fundamentales, especialmente cuando el objetivo es predecir una variable continua, como el promedio de calificaciones final (García, 2023).
- ✓ **Regresión Lineal Simple y Múltiple:** El modelo de regresión lineal es la técnica más básica, pero poderosa. En su forma simple, busca modelar la relación lineal entre una variable independiente y una dependiente. La regresión lineal múltiple permite examinar el efecto de varias variables independientes sobre una variable dependiente, controlando simultáneamente la influencia de las demás” (Hair et al., 2019). Se propone una metodología para construir modelos predictivos del

rendimiento académico mediante características de estudiantes de ingeniería, como el promedio actual y el desempeño en asignaturas clave, con el fin de anticipar el riesgo de reprobación y mejorar la toma de decisiones institucionales (Rico Páez & Gaytán Ramírez 2022).

- ✓ **Regresión logística:** La regresión logística es apropiada cuando la variable dependiente es categórica, especialmente binaria, como aprobar o reprobar un curso, ya que permite modelar la probabilidad de ocurrencia de un evento en función de variables independientes” (Hosmer et al., 2013).
- ✓ Este modelo estima la probabilidad de que un evento ocurra, basándose en uno o más factores predictivos. En un contexto de ingeniería, un modelo de regresión logística podría predecir la probabilidad de que un estudiante complete un semestre con éxito, basándose en su rendimiento inicial.
- ✓ **Otros modelos de regresión avanzados:** La estadística predictiva moderna incluye técnicas avanzadas como árboles de decisión, redes neuronales y regresión polinomial, que permiten modelar relaciones no lineales entre variables y mejorar la precisión de los modelos” (Kotsiantis et al., 2007, p. 5) Estos modelos ofrecen una mayor flexibilidad para encontrar patrones más sutiles en los datos, lo que a menudo resulta en una mayor precisión predictiva.

### El ciclo de vida de un modelo predictivo

El desarrollo de modelos predictivos del rendimiento académico se concibe como un proceso progresivo, compuesto por múltiples etapas que permiten ajustar y mejorar la precisión del modelo a lo largo del tiempo” (Rico Páez & Gaytán Ramírez 2022):

- ✓ **Definición del problema:** Claramente establecer qué se busca predecir. En nuestro caso, prever el rendimiento académico en ingeniería.
- ✓ **Recolección y preparación de datos:** La calidad de los datos, desde su recolección hasta su organización, es un factor crítico que determina la precisión de los modelos predictivos, especialmente cuando se utilizan variables como calificaciones y datos demográficos (Thai Jo., 2024). Reunir los datos pertinentes (calificaciones, datos demográficos, etc.), limpiarlos y organizarlos. La calidad de los datos es un factor determinante en la precisión del modelo (Thai Jo., 2024).
- ✓ **Selección y entrenamiento del modelo:** Elegir el modelo de regresión más adecuado (lineal, logística, etc.) y entrenarlo con los datos históricos disponibles.
- ✓ **Evaluación y validación:** Medir la precisión del modelo utilizando métricas adecuadas (como el error cuadrático medio,  $R^2$ ,  $R^2$  ajustado, o la matriz de confusión) para asegurar que las predicciones sean fiables (Thai Jo., 2024).
- ✓ **Despliegue e implementación:** Una vez validado, el modelo se implementa para tomar decisiones informadas, como identificar estudiantes en riesgo para ofrecerles apoyo pedagógico temprano.

Se puede considerar que la estadística predictiva y los modelos de regresión, en particular, ofrecen un marco robusto y metodológicamente riguroso para ir "más allá de las calificaciones". Al cuantificar la influencia de diversos factores y construir modelos predictivos,

las facultades de ingeniería pueden transformar los datos del pasado en una poderosa herramienta para moldear proactivamente el éxito de sus estudiantes en el futuro.

### **El desafío del fracaso académico en la ingeniería**

La deserción universitaria constituye una problemática significativa en varios países, especialmente en aquellos pertenecientes al contexto latinoamericano, y se acentúa en carreras de alta exigencia como las ingenierías (Villegas et al., 2024). La complejidad de estas disciplinas, sumada a otros factores, contribuye a una tasa de abandono que no solo afecta el desarrollo personal y profesional del estudiante, sino que también genera un impacto económico y social considerable. La pérdida de talento en áreas vitales para el progreso tecnológico y económico, como la ingeniería, subraya la urgencia de abordar este problema de manera proactiva.

El fracaso académico en las carreras de ingeniería no es simplemente una estadística de reprobación; es una problemática estructural compleja con profundas implicaciones para el desarrollo económico y tecnológico de cualquier nación. Como educadores e ingenieros, debemos verlo como una falla en el sistema que diseñamos y gestionamos. Definir este desafío es el primer paso para construir modelos predictivos que permitan mitigarlo.

En este contexto, la necesidad de herramientas que permitan la identificación temprana de estudiantes en riesgo es fundamental. La intervención oportuna, basada en datos y modelos predictivos, puede transformar la respuesta de las instituciones educativas, pasando de una reacción al fracaso a una estrategia de prevención. Este capítulo se enfoca en sentar las bases teóricas y conceptuales para comprender el rendimiento académico y los factores que lo influyen, preparando el terreno para la aplicación de modelos predictivos que buscan mitigar el fracaso y promover la retención estudiantil.

Tradicionalmente, la atención se ha centrado en los indicadores más evidentes de fracaso, como la no aprobación de materias en los primeros semestres. No obstante, una visión más profunda revela la interacción de factores personales, académicos e institucionales que actúan como detonantes. Por ejemplo, estudios han demostrado que la motivación, la autoeficacia y la resiliencia son variables cruciales en el desempeño de los estudiantes de ingeniería (Zapata et al., 2021). Aquellos estudiantes que muestran una mayor resiliencia—la capacidad de adaptarse positivamente a situaciones adversas—tienden a gestionar mejor la exigencia académica y el estrés que caracteriza a la disciplina.

### **Fracaso académico: Más allá de la reprobación**

El fracaso académico en la educación superior en Ingeniería puede definirse como la **incapacidad del estudiante para cumplir con los objetivos curriculares y los estándares de progreso establecidos** por la institución dentro del tiempo esperado. Se manifiesta no solo en las calificaciones bajas o la reprobación de asignaturas clave (típicamente las de ciencias básicas como cálculo, física y química), sino también en fenómenos más sutiles como:

✓ **Deserción Temprana:** El abandono de la carrera durante los primeros semestres.

✓ **Rezago (Attrition):** La demora significativa en la finalización de los estudios o la necesidad de reingreso.

✓ **Bajo Rendimiento Crónico:** Obtener consistentemente el mínimo necesario para aprobar, lo que indica una deficiencia en la adquisición de competencias fundamentales.

Tradicionalmente, el fracaso ha sido atribuido a factores internos del estudiante, como la falta de aptitud o la deficiente preparación previa (Zapata et al., 2021). Sin embargo, esta visión es limitada. La investigación actual subraya que el fracaso es un fenómeno multifactorial que emerge de la compleja interacción entre el estudiante, la institución y el contexto socioeconómico.

### **Implicaciones del fracaso en ingeniería**

El fracaso académico en este campo tiene un impacto crítico por varias razones específicas de la disciplina:

✓ **Deterioro de la autoeficacia y la motivación:** La alta dificultad y la carga de trabajo de las asignaturas de ingeniería pueden erosionar rápidamente la **autoeficacia** del estudiante, llevándolo a la deserción incluso si su aptitud inicial era alta (Zapata et al., 2021). La resiliencia y la gestión del estrés académico se vuelven, por lo tanto, **variables críticas** que deben ser modeladas.

✓ **Impacto económico y social:** El alto índice de fracaso en ingeniería representa una **pérdida de inversión** para el estudiante, la familia y el estado, al desperdiciar recursos educativos e infraestructuras especializadas. Más importante aún, frena la formación de los profesionales que la economía necesita para la innovación tecnológica.

✓ **Vulnerabilidad curricular:** El currículo de ingeniería tiene una estructura secuencial rígida donde el dominio de una materia (ej. Cálculo I) es un requisito esencial para la siguiente (ej. Resistencia de Materiales). Un fallo temprano crea un efecto dominó que hace exponencialmente más difícil la recuperación, validando la necesidad de una predicción temprana y precisa.

El desafío del fracaso académico en la ingeniería es la manifestación de una desconexión entre los requisitos del plan de estudios y las necesidades de apoyo del estudiante. Superar este desafío requiere dejar de lado la simple métrica del GPA y adoptar los modelos predictivos (como los que emplean la Minería de Datos Educativa o EDM) que permiten intervenir con estrategias personalizadas y oportunas.

### **Factores que inciden en el fracaso académico en ingeniería**

La literatura académica identifica diversos factores que contribuyen al fracaso estudiantil en ingeniería, los cuales pueden agruparse en varias categorías clave:

#### **1.- Factores personales y psicológicos:**

- **Motivación y expectativas:** La pérdida de interés en la carrera elegida puede ser un factor determinante, especialmente si las expectativas iniciales no se corresponden con la realidad académica.

- **Adaptación al entorno universitario:** La dificultad para socializar y adaptarse al ambiente universitario, junto con la sensación de aislamiento, puede impactar negativamente en el rendimiento académico.
- **Bienestar psicológico:** Factores como el estrés, la ansiedad y el síndrome de burnout pueden mermar la capacidad del estudiante para concentrarse y perseverar (Atencia Oliva et al., 2020).

## 2.- Factores académicos y curriculares:

- **Carga de trabajo y dificultad de las asignaturas:** Las carreras de ingeniería, con su alta demanda de tiempo de estudio y la complejidad de las materias, pueden ser abrumadoras. Algunos estudiantes, incluso con alta perseverancia, pueden verse superados.
- **Flexibilidad curricular:** La rigidez en los planes de estudio y la falta de opciones para una movilidad académica pueden contribuir al desánimo y la deserción,
- **Calidad de la enseñanza:** La calidad educativa está asociada a un conjunto de factores entre los que se cuentan los factores económicos, políticos, culturales, científicos y tecnológicos, así como un importante factor humano: el desempeño docente, que resulta vital según la postura que se sostiene”, el desempeño del personal docente incluyendo su formación continua y los métodos de evaluación que emplea influye directamente en la calidad del aprendizaje y el rendimiento estudiantil (Escribano Hervis, 2018).

## 3.- Factores socioeconómicos y familiares:

- **Acceso a recursos:** Las disparidades socioeconómicas se traducen en diferencias en el acceso a recursos educativos, como tutores, tecnología o incluso una nutrición adecuada, lo que puede influir directamente en el desempeño.
- **Costos financieros:** El elevado costo de la educación superior puede ser una barrera significativa, especialmente para estudiantes de bajos ingresos, impactando la estabilidad académica y la capacidad de concentración.
- **Apoyo familiar:** El entorno familiar, el nivel educativo de los padres y el apoyo que brindan pueden ser determinantes en el éxito académico.

## La necesidad de un enfoque proactivo

Abordar el desafío del fracaso académico en la ingeniería requiere un enfoque integral y proactivo. Al reconocer la complejidad y la naturaleza multifactorial de este fenómeno, podemos ir más allá de la simple observación de calificaciones y utilizar modelos predictivos para identificar a los estudiantes en riesgo de manera temprana. La estadística predictiva, como se explorará en el siguiente capítulo, nos permite utilizar los datos del pasado y del presente para prever el futuro, brindando a las facultades de ingeniería la capacidad de implementar estrategias de retención efectivas y personalizadas, como programas de tutoría y apoyo psicológico, antes de que el fracaso

se consume. En última instancia, la meta no es solo reducir la deserción, sino fomentar un ambiente educativo que promueva el éxito y el florecimiento de todos los estudiantes de ingeniería.

## **Ventajas del uso de regresión logística en este contexto**

### **Conceptualización del rendimiento académico: Más allá de las calificaciones**

El **rendimiento académico** es un concepto multifacético que no puede limitarse únicamente a las notas obtenidas en los cursos. Si bien las calificaciones son un indicador clave, una definición más completa debe incluir la **retención** (la permanencia del estudiante en el programa) y la **graduación** (la culminación exitosa de los estudios en un período de tiempo razonable) (Chacha et al., 2023). Desde esta perspectiva multidimensional, un estudiante con buen rendimiento es aquel que no solo aprueba sus asignaturas, sino que también avanza de manera constante hacia su meta de obtener un título universitario.

Esta visión integral permite una evaluación más precisa del éxito estudiantil y de la eficacia de los programas educativos. Por ejemplo, un estudiante puede tener calificaciones altas en sus primeros semestres, pero si eventualmente abandona la carrera, su rendimiento global se ve comprometido. Por ello, la medición del rendimiento académico debe considerar todo el trayecto formativo del estudiante.

En la ingeniería, se valoran las herramientas que ofrecen eficiencia, interpretabilidad y precisión. En el campo de la estadística predictiva académica, la Regresión Logística (RL) se destaca como un método fundamental y potente, particularmente adecuado para modelar el rendimiento estudiantil en la educación superior en Ingeniería. Su principal ventaja radica en su capacidad para abordar directamente la naturaleza binaria de los resultados críticos en la trayectoria del estudiante. Los resultados sugieren que una integración adecuada de las tecnologías y un enfoque centrado en el estudiante pueden influir significativamente en la cultura y las políticas institucionales, apoyando a los estudiantes de riesgo y garantizando su éxito académico y personal (Félix et al., 2025), la regresión logística destaca por su equilibrio entre poder predictivo, interpretabilidad y eficiencia, lo que la convierte en una herramienta invaluable para la educación superior en ingeniería. Su capacidad para modelar la probabilidad de una variable de respuesta binaria la hace ideal para predecir escenarios de interés, como la probabilidad de deserción estudiantil.

### **Modelando resultados binarios críticos**

A diferencia de la regresión lineal, que predice un valor continuo (como el GPA final), la Regresión Logística está diseñada para predecir la **probabilidad de que ocurra un evento categórico o binario** (Hair et al., 2019). En el ámbito de la Ingeniería, esto es esencial para modelar decisiones académicas binarias que definen la progresión del estudiante:

- **Predicción de aprobación/reprobación:** La RL puede modelar la probabilidad de que un estudiante apruebe o reapruebe una asignatura específica de alto riesgo (ej. Cálculo o Física), permitiendo una intervención focalizada.

- **Predicción de deserción/continuidad:** Permite predecir si un estudiante tiene una alta probabilidad de **desertar** antes de finalizar un período académico, un desafío crítico en las instituciones de ingeniería.

La RL convierte las variables predictoras (ej. notas de admisión, hábitos de estudio) en una **probabilidad P** que se transforma mediante una función logística (logit) para asegurar que el resultado esté siempre entre 0 y 1. Esto ofrece una métrica intuitiva y directamente aplicable a la toma de decisiones.

### Interpretación y sencillez de implementación

Una ventaja crucial, especialmente en entornos institucionales que requieren transparencia en la toma de decisiones, es la **interpretabilidad** de los resultados de la Regresión Logística:

- **Cuantificación del Riesgo:** Los coeficientes de la RL se transforman fácilmente en *Odds Ratios* (OR), que indican **cuánto aumenta o disminuye** la probabilidad de fracaso por cada unidad de cambio en una variable predictora. Por ejemplo, un OR de 1.5 para la variable "ausencia en tutorías" indica que la probabilidad de reprobar la materia es 1.5 veces mayor para esos estudiantes. Esta cuantificación rigurosa es clave para diseñar políticas de apoyo efectivas.
- **Eficiencia Computacional:** La Regresión Logística es **computacionalmente eficiente** y requiere menos recursos y tiempo de entrenamiento que los modelos de *Machine Learning* más complejos (como las Redes Neuronales o los Bosques Aleatorios). Esto la hace ideal para sistemas de alerta temprana en tiempo real en grandes bases de datos educativas.

En resumen, la regresión logística es la herramienta de elección para establecer los **fundamentos del rendimiento académico y la predicción en la educación superior en ingeniería** cuando el objetivo es transformar una multitud de variables de entrada en una estimación de riesgo binario clara, interpretable y directamente aplicable a la gestión académica.

### Interpretación de la relación predictiva

Una de las mayores fortalezas de la regresión logística radica en su capacidad para ofrecer una interpretación clara de la influencia de cada variable predictora. A diferencia de modelos más complejos, que pueden funcionar como "cajas negras", la regresión logística permite a los educadores comprender el peso y la dirección de la asociación de cada factor con el resultado de interés. Por ejemplo, en un estudio sobre el fracaso académico en ingeniería, un modelo de regresión logística podría revelar que, manteniendo constantes otras variables, cada punto adicional en la calificación de un curso introductorio de matemáticas aumenta significativamente la probabilidad de completar el semestre. Esta transparencia es crucial para que las facultades de ingeniería puedan no solo identificar a los estudiantes en riesgo, sino también entender *por qué* lo están, permitiendo intervenciones más informadas.

### Eficiencia computacional y escalabilidad

La regresión logística es matemáticamente menos compleja que otros algoritmos, lo que se traduce en una mayor eficiencia computacional. La incorporación de la minería de datos en el ámbito educacional permite a las instituciones optimizar intervenciones pedagógicas sin requerir

una infraestructura tecnológica compleja, lo que facilita su adopción incluso en contextos con recursos limitados” (Camones et al., 2024), Esto es particularmente ventajoso para las instituciones educativas, que a menudo manejan grandes volúmenes de datos estudiantiles y necesitan modelos que puedan ser entrenados y actualizados de manera rápida y eficiente.

### **Estimación de probabilidades y clasificación binaria**

La regresión logística no solo clasifica a los estudiantes en categorías discretas (por ejemplo, "éxito" o "fracaso"), sino que también estima la probabilidad de que un estudiante pertenezca a una u otra categoría. Esta estimación de probabilidad es de gran valor, ya que permite a los educadores priorizar los recursos. Por ejemplo, en lugar de tratar a todos los estudiantes en riesgo de la misma manera, se pueden enfocar los esfuerzos de apoyo en aquellos con la mayor probabilidad estimada de fracaso, optimizando así el impacto de las intervenciones. Un modelo que predice la probabilidad de deserción (por ejemplo,  $P(\text{deserción}) > 0.75$ ) permite a los tutores concentrar sus esfuerzos en los casos más críticos.

### **Compatibilidad con variables mixtas**

Un entorno educativo es inherentemente complejo, con variables predictoras que pueden ser tanto continuas (ej. GPA) como categóricas (ej. tipo de admisión, género). La regresión logística maneja con facilidad esta combinación de tipos de variables, sin requerir suposiciones restrictivas sobre la distribución de las mismas. En la educación en ingeniería, esto permite un análisis holístico que integra información del rendimiento académico previo con variables socio-demográficas, permitiendo una visión más completa de los factores que influyen en el éxito estudiantil.

### **Robustez y madurez metodológica**

La regresión logística es una técnica estadística madura, con una sólida base teórica y una amplia aceptación en la comunidad científica. Esto le confiere una robustez metodológica que la hace una opción fiable para la investigación educativa. La familiaridad de los investigadores y educadores con este modelo facilita la replicación y la comparación de resultados, contribuyendo al avance del conocimiento en el campo de la predicción del rendimiento académico.

### **Variables y factores de influencia: La complejidad de la trayectoria estudiantil**

La trayectoria estudiantil en las carreras de ingeniería es mucho más que una secuencia de calificaciones y cursos aprobados. Es un ecosistema dinámico y complejo, influenciado por la interacción de múltiples factores a lo largo del tiempo (Moreno & Chiecher, 2018). Entender esta complejidad es fundamental para desarrollar modelos predictivos robustos, capaces de identificar no solo los riesgos de fracaso, sino también las oportunidades de éxito. Este capítulo profundiza en las dimensiones que configuran esta trayectoria, y presenta un marco analítico para su comprensión y abordaje en el contexto de la educación superior.

En ingeniería, se sabe que la calidad de cualquier modelo predictivo depende directamente de la **calidad y relevancia de sus variables de entrada**. La trayectoria estudiantil en Ingeniería es un sistema dinámico y multifactorial, y reducir su rendimiento a una sola métrica (como el GPA) es una simplificación peligrosa. Para construir modelos que realmente vayan **"más allá de las**

**calificaciones"**, es imperativo identificar y categorizar las variables que ejercen influencia significativa.

El rendimiento académico en estudiantes de ingeniería está determinado por una interacción compleja entre variables individuales, académicas e institucionales, así como por el contexto socioeconómico del estudiante (González & Rodríguez, 2023)

El éxito en esta disciplina requiere la posesión de fundamentos sólidos, como lo subraya el trabajo sobre los Fundamentos del Rendimiento Académico y la Predicción en la Educación Superior en Ingeniería.

Comprender el rendimiento académico requiere analizar un conjunto diverso de factores que lo condicionan. La literatura especializada identifica una variedad de variables que pueden agruparse en tres categorías principales (Navarro Roldán & Zamudio Sisa, 2021):

#### **Variables sociodemográficas:**

Incluyen aspectos como la edad, el género, el estatus socioeconómico, el lugar de origen y la estructura familiar. Estas variables pueden influir en la disponibilidad de recursos, el apoyo familiar y las responsabilidades externas que afectan el tiempo y la energía dedicados al estudio. Por ejemplo, estudios han demostrado la influencia del entorno socioeconómico en la permanencia de estudiantes en la universidad (Blanco Torres et al., 2024).

#### **Variables socioeconómica y familiar:**

El contexto socioeconómico y el entorno familiar pueden generar ventajas o desventajas que impactan significativamente en la trayectoria. Los estudiantes de bajos ingresos enfrentan dificultades económicas que pueden obligarlos a trabajar más horas, reduciendo así el tiempo disponible para el estudio y afectando su rendimiento académico (Figueroa Guerra et al., 2025)

En un análisis de datos de estudiantes de ingeniería se observó que aquellos que reportaron la necesidad de trabajar más de 20 horas a la semana tuvieron una probabilidad 2.5 veces mayor de reprobar al menos una asignatura en el primer año (datos hipotéticos). Esto evidencia la necesidad de considerar el contexto socioeconómico al interpretar el rendimiento académico.

Variables como el nivel educativo de los padres, el ingreso familiar y el acceso a recursos tecnológicos pueden generar ventajas o desventajas que impactan el desempeño y deben ser neutralizadas o incluidas en el modelo.

#### **Variables académicas previas:**

Se refieren al historial educativo del estudiante antes de ingresar a la educación superior. Factores como el promedio de calificaciones en la educación secundaria, los resultados en pruebas de admisión, el tipo de colegio de procedencia y la elección de la carrera son predictores sólidos del éxito en la universidad. La preparación académica previa es a menudo un indicador del dominio de habilidades fundamentales para el éxito en la educación superior. Estudios han demostrado que existe una correlación significativa entre los resultados del examen de admisión y el rendimiento académico durante el primer año universitario, especialmente en asignaturas clave como matemáticas y física" (Chaves et al., 2008).

Los resultados del rendimiento del primer ciclo académico revelaron que el desempeño en asignaturas fundacionales como cálculo y física está estrechamente vinculado al promedio final, siendo estas materias indicadores clave del progreso académico” (Verdugo-Guamán et al., 2023), por ejemplo, los estudiantes que obtuvieron un promedio superior a 7.5 (escala de 10) en estos cursos iniciales mostraron una probabilidad 1.8 veces mayor de graduarse en el tiempo esperado que aquellos con un promedio inferior (datos hipotéticos). Estos hallazgos destacan la importancia de los cursos de fundamentos como indicadores tempranos del éxito a largo plazo.

El diseño instruccional en cursos fundacionales como Cálculo y Álgebra Lineal influye significativamente en el rendimiento académico, siendo estas materias determinantes en el progreso de los estudiantes de ingeniería (García & Rodríguez, 2023). El desempeño en asignaturas clave como Cálculo I y Álgebra Lineal puede predecir el rezago académico en carreras de ingeniería, alineándose con tu enfoque sobre la importancia de los exámenes diagnósticos.

### **Variables institucionales:**

La cultura organizacional en las instituciones de educación superior, junto con sus recursos y políticas internas, influye directamente en el desempeño académico y la trayectoria estudiantil, especialmente cuando se promueven ambientes de apoyo como tutorías, infraestructura adecuada y acceso a tecnologías (Siqueiros-Quintana & Vera-Noriega., 2022). En una encuesta realizada a estudiantes de ingeniería (Generado para este ejercicio), aquellos que reportaron una mayor satisfacción con los servicios de apoyo académico tuvieron una tasa de deserción 15% menor que sus pares menos satisfechos (datos hipotéticos). Esto sugiere que un entorno de apoyo robusto actúa como un amortiguador contra los desafíos académicos y personales.

### **Variables de Compromiso y Actividad en Plataformas (LMS Data):**

El número de accesos al sistema, la participación en foros, la entrega de tareas y la realización de cuestionarios son indicadores clave que pueden ser utilizados por técnicas de minería de datos educativa para predecir el rendimiento académico (Romero et al., 2025).

### **Variables de índole personal y psicológica**

Comprenden aspectos psicológicos y de comportamiento como la motivación, la autoeficacia, la resiliencia, las estrategias de aprendizaje y la participación en actividades extracurriculares. Estos factores internos determinan cómo un estudiante enfrenta los desafíos académicos y se adapta a las exigencias del entorno universitario. La capacidad de un estudiante para adaptarse al cambio y superar obstáculos es un componente vital para su permanencia y éxito.

La planificación y gestión del tiempo académico son habilidades clave que influyen directamente en el rendimiento de los estudiantes universitarios, especialmente en carreras exigentes como ingeniería (Reyes-González et al., 2022).

La creencia en la propia capacidad para tener éxito y la motivación intrínseca son cruciales para la persistencia frente a las altas tasas de reprobación iniciales (Zapata et al., 2021). Un estudiante con alta motivación está más dispuesto a enfrentar el estrés académico.

Las características personales y psicológicas de los estudiantes, como la resiliencia, la autoeficacia y la motivación, son determinantes en su capacidad para perseverar en la exigente carrera de ingeniería (Zapata et al., 2021). Un estudio con estudiantes de ingeniería (Generado para

este ejercicio) encontró que el 70% de los estudiantes que reportaron un alto nivel de resiliencia completaron su carrera, en comparación con solo el 45% de aquellos con baja resiliencia (datos hipotéticos). Este resultado subraya que la persistencia frente a los fracasos y la capacidad de gestionar el estrés son habilidades tan importantes como el dominio técnico.

La verdadera potencia de los modelos predictivos reside en su capacidad para integrar y analizar estas múltiples variables simultáneamente. Un modelo de regresión logística, por ejemplo, podría combinar datos sobre el GPA inicial, el uso de servicios de tutoría, los niveles de resiliencia autoinformados y las horas de trabajo semanal para predecir la probabilidad de que un estudiante complete el semestre exitosamente. En lugar de basar las intervenciones en un solo factor, como una nota baja, este enfoque holístico permite a las instituciones diseñar estrategias de apoyo más precisas y efectivas, atendiendo a la raíz del problema y no solo a sus síntomas.

El involucramiento en metodologías como el **aprendizaje cooperativo** o el **aprendizaje basado en problemas (ABP)** es un factor positivo. Las investigaciones indican que estas estrategias no solo mejoran el conocimiento, sino también la resiliencia y el trabajo en equipo, elementos esenciales en Ingeniería.

### **La Importancia de la Estadística en la Educación: Herramienta de toma de decisiones**

El papel de la **estadística** en el ámbito educativo es cada vez más crucial. Más allá de ser una disciplina teórica, la estadística se erige como una **herramienta fundamental para la investigación, la evaluación y la toma de decisiones basada en evidencia** (Solís Ventura et al., 2022). Permite a los educadores y administradores universitarios ir más allá de la intuición y las anécdotas para entender patrones y relaciones complejas en grandes conjuntos de datos.

La Estadística es la herramienta fundamental que traslada estos principios al ámbito académico, transformando el volumen masivo de datos estudiantiles en información accionable y objetiva para la toma de decisiones.

Tradicionalmente, la estadística en educación se ha limitado a funciones descriptivas: calcular promedios (GPA), porcentajes de reprobación y tasas de deserción. Si bien estos datos son necesarios, son insuficientes para la gestión proactiva. La verdadera importancia de la estadística en un contexto de Ingeniería reside en su capacidad para las funciones inferencial y predictiva (Hair et al., 2019)

Mediante el uso de técnicas estadísticas como la regresión logística (Chacha et al., 2023) o modelos de análisis predictivo, las instituciones pueden:

1. **Identificar factores de riesgo:** Determinar cuáles variables tienen el mayor impacto en el rendimiento académico y la deserción. La estadística inferencial nos permite determinar si existe una relación significativa entre dos variables, podemos identificar qué variables (aptitud previa, hábitos de estudio, participación en clases activas) son predictores. Esto garantiza que nuestros modelos sean parsimoniosos y eficientes.

Por ejemplo, ¿el uso de la metodología de aprendizaje cooperativo realmente se correlaciona con un mejor rendimiento en la asignatura de Termodinámica? La

regresión y el análisis de varianza (ANOVA) nos dan respuestas con un nivel cuantificable de certeza.

2. **Predecir resultados:** Estimar la probabilidad de que un estudiante tenga éxito o abandone sus estudios. Mediante modelos como la Regresión Logística, podemos cuantificar el riesgo de fracaso o deserción. Esto nos permite establecer umbrales de riesgos precisos y focalizar recursos limitados en los estudiantes que más los necesitan Martínez et, al (2021). Este rigor es vital para la Ingeniería, donde las decisiones curriculares y de apoyo deben basarse en evidencia sólida, no en la intuición.
3. **Evaluar la eficacia de las intervenciones:** Medir el impacto de programas de tutoría, apoyo académico o becas en el rendimiento de los estudiantes.
4. **Validación del Modelo:** Una vez que se construye un modelo predictivo, como un modelo de regresión, la estadística es esencial para evaluar su validez externa y fiabilidad. Métricas como el R<sup>2</sup> (en regresión lineal) o el AUC (en regresión logística) nos permiten cuantificar la precisión con la que el modelo puede prever resultados futuros, asegurando que la herramienta de toma de decisiones sea robusta.

Estos modelos permiten a las facultades de ingeniería, como cualquier organización moderna, operar bajo un esquema de **mejora continua** y **toma de decisiones informada**, migrando de una gestión reactiva a una **proactiva** para asegurar la calidad y la retención estudiantil.

La capacidad de predecir es esencial para el manejo efectivo de los fundamentos del rendimiento académico y la predicción en la educación superior en ingeniería (Contreras et al., 2020). La estadística transforma a la institución educativa en un sistema de ingeniería optimizado. Nos permite migrar de un enfoque reactivo ("el estudiante reprobó, veamos por qué") a un enfoque proactivo y preventivo ("el estudiante tiene un 75% de riesgo de reprobado; activemos tutorías hoy"), elevando la gestión académica al mismo nivel de rigor y planificación que cualquier proyecto de diseño en ingeniería.

# Capítulo 2

## Introducción a la Estadística Predictiva y los Modelos de Regresión

### La Estadística predictiva: Previendo el futuro con datos

La **estadística predictiva** es una rama de la ciencia de datos que utiliza técnicas estadísticas y de aprendizaje automático para predecir eventos futuros o resultados desconocidos. A diferencia de la estadística descriptiva, que se enfoca en resumir y describir datos existentes, la estadística predictiva se centra en la construcción de modelos matemáticos que identifican patrones y relaciones entre variables para hacer pronósticos informados.

En el contexto educativo, la estadística predictiva es invaluable. Permite a las instituciones pasar de la simple descripción del fracaso académico a la capacidad de anticiparlo. Por ejemplo, en lugar de solo saber cuántos estudiantes abandonaron el año pasado, un modelo predictivo puede estimar qué estudiantes de la cohorte actual están en mayor riesgo de deserción, basándose en su perfil académico y socioeconómico inicial.

La estadística predictiva y los modelos de regresión son herramientas esenciales de la **ingeniería de datos** aplicadas a la educación. Permiten pasar de una evaluación retrospectiva (saber qué sucedió) a una **intervención proactiva** (saber qué sucederá si no se actúa). Al cuantificar la influencia de los factores académicos y socioeconómicos, estos modelos garantizan que las decisiones institucionales (planes de estudio, soporte al estudiante) se basen en **evidencia objetiva** y no en mera intuición (Alyahyan & Dustegor, 2020). La validación constante del modelo (verificando la exactitud de las predicciones) asegura que la universidad mantenga una ventaja competitiva en la calidad y eficiencia de la formación de ingenieros.

El propósito de la estadística predictiva es buscar modelar la relación entre variables independientes (predictores) y una variable dependiente (resultado) para pronosticar futuros rendimientos o comportamientos. En ingeniería, estas técnicas permiten identificar factores que influyen en el rendimiento académico y predecir resultados como calificaciones finales, retención o desempeño en cursos clave (James et al., 2023).

La estadística predictiva, en el contexto de la educación superior en ingeniería, se define como el conjunto de técnicas que utiliza **datos históricos** (calificaciones, demografía, hábitos de estudio) para pronosticar **resultados futuros probables** (riesgo de deserción, éxito en un curso). No se trata de adivinar, sino de calcular probabilidades.

La estadística predictiva conjuntamente con los modelos de regresión son herramientas esenciales de la ingeniería de datos aplicadas a la educación. Permiten pasar de una evaluación retrospectiva (saber qué sucedió) a una intervención proactiva (saber qué sucederá si no se actúa). Al cuantificar la influencia de los factores académicos y socioeconómicos, estos modelos garantizan

que las decisiones institucionales (planes de estudio, soporte al estudiante) se basen en evidencia objetiva y no en mera intuición (Alyahyan & Dustegor, 2020). La validación constante del modelo (verificando la exactitud de las predicciones) asegura que la universidad mantenga una ventaja competitiva en la calidad y eficiencia de la formación de ingenieros.

## Modelos de Regresión: La base de la predicción

Los **modelos de regresión** son una de las herramientas más fundamentales y ampliamente utilizadas en la estadística predictiva. Su propósito principal es analizar la relación entre una variable dependiente (el resultado que se quiere predecir) y una o más variables independientes (los predictores). En términos simples, estos modelos buscan encontrar la mejor "línea" o "curva" que describe cómo los cambios en los predictores se asocian con cambios en la variable de interés. **La Regresión como Herramienta Fundamental de Modelización y el análisis** es la técnica estadística más común en la construcción de estos modelos, ya que establece y cuantifica la relación funcional entre una **variable dependiente** (el resultado que se quiere predecir) y una o más **variables independientes** (los predictores).

El enfoque central es construir un **modelo predictivo**, que es una función matemática o un algoritmo entrenado para identificar patrones complejos en los datos. Este modelo permite a las facultades de ingeniería:

1. **Anticipar el riesgo:** Identificar a los estudiantes con alta probabilidad de reprobación o abandonar antes de que ocurra el evento.
2. **Optimizar recursos:** Dirigir tutorías, asesorías y cursos de nivelación de manera eficiente solo a aquellos estudiantes que realmente lo necesitan.
3. **Evaluar el currículo:** Medir la influencia de los cursos de entrada (e.g., cálculo, física) en el éxito final en la carrera, como parte de la mejora continua del proceso.

Existen varios tipos de modelos de regresión, cada uno adecuado para diferentes tipos de datos:

**Regresión Lineal Simple** (datos pasados a decisiones futuras): Se utiliza cuando la variable dependiente es **cuantitativa** (numérica) y solo se tiene una variable predictora. El modelo busca una línea recta que minimice la distancia entre los puntos de datos y la línea. Por ejemplo, podría usarse para predecir la calificación final de un estudiante basándose en su calificación en el examen parcial, igualmente puede predecir el **promedio final** o la **calificación numérica** que obtendrá en un curso complejo.

La regresión lineal asume una relación aproximadamente lineal entre los predictores y la salida. En el contexto del rendimiento académico en ingeniería, esta herramienta resulta crucial para construir modelos predictivos que trascienden la mera calificación descriptiva.

### Ejemplo práctico

Se desea predecir la nota que los estudiantes de un curso de **Análisis de Circuitos I** (una asignatura típica de Ingeniería Eléctrica o Electrónica) obtendrán en el **Examen Final** (Y), basándose en la nota que obtuvieron en el **Primer Examen Parcial** (X).

Se han recopilado datos de una pequeña muestra de 5 estudiantes (las notas van de 0 a 20, por simplicidad).

Estudiante	Nota Parcial (X)	Nota Final (Y)
1	12	14
2	15	16
3	10	11
4	18	19
5	14	15

**Objetivo:**

- Determinar la ecuación de regresión lineal simple:  $\hat{Y} = b_0 + b_1X$ .
- Predecir la nota final de un estudiante que obtuvo 16 en el Parcial.

**Resolución:**

**Cálculo de Medias y Sumatorias**

Primero, se calcula las medias aritméticas ( $\bar{x}$ ,  $\bar{y}$ ) y las sumatorias necesarias:

X	Y	Xy	x <sup>2</sup>
12	14	168	144
15	16	240	225
10	11	110	100
18	19	342	324
14	15	210	196
<b><math>\Sigma X = 69</math> <math>\Sigma Y = 75</math> <math>\Sigma XY = 1070</math> <math>\Sigma X^2 = 989</math></b>			

Número de observaciones (n): 5

Media de X ( $\bar{x}$ ):  $\Sigma X = 69 = 13.8$

Media de Y ( $\bar{y}$ ):  $\Sigma Y = 75 = 15.0$

**Cálculo de los Coeficientes de Regresión ( $b_1$  y  $b_0$ ):**

Pendiente ( $b_1$ ): La pendiente mide el cambio en Y por unidad de cambio en X.

$$b_1 = \frac{\Sigma X^2 - n\bar{x}^2}{\Sigma XY - n\bar{x}\bar{y}}$$

**Sustituimos los valores:**

$$b_1 = \frac{989 - 5(13.8)^2}{21070 - 5(13.8)(15.0)} = \frac{989 - 952.2}{21070 - 1035} = \frac{36.8}{35} \approx 0.951$$

**Intercepto ( $b_0$ )**

El intercepto representa el valor esperado de Y cuando X=0.

$$b_0 = y^- - b_1x^-$$

Sustituimos los valores:  $b_0 = 15.0 - (0.951)(13.8)$ ;  $b_0 = 15.0 - 13.1238 \approx 1.876$

Ecuación de Regresión:  $Y = 1.876 + 0.951X$

Esta ecuación es la que se utilizará para la predicción.

Interpretación de los Parámetros:

**Pendiente ( $b_1 = 0,951$ ):** Por cada punto adicional obtenido en el Examen Parcial (X), se espera un aumento de **0,951 puntos** en el Examen Final (Y).

**Intercepto ( $b_0 = 1,876$ ):** Si un estudiante obtiene 0 en el Parcial, el modelo predice que su nota final será de **1,876** (aunque en este contexto, un valor de X=0 podría estar fuera del rango de datos observados, por lo que la interpretación debe ser cautelosa).

### Predicción de la Nota Final

Para un estudiante que obtuvo X=16 en el Parcial, la nota final predicha (Y) es:

$$Y = 1,876 + 0,951(16); Y = 1,876 + 15,216; Y = 17,092$$

Basado en el modelo de regresión, se predice que un estudiante de Ingeniería en Análisis de Circuitos que obtuvo **16** puntos en el primer examen parcial, obtendrá aproximadamente **17,09** puntos en el examen final. Este tipo de modelos simples es una herramienta valiosa para que los profesores puedan identificar a tiempo a los estudiantes que podrían necesitar apoyo adicional (Alvarado & García Jiménez, 1997).

**Regresión Lineal Múltiple:** Es una extensión de la regresión simple que permite incluir **múltiples variables predictoras cuantitativas o cualitativas** para predecir una variable dependiente cuantitativa. Este modelo es más realista para la mayoría de los escenarios, ya que el rendimiento académico no depende de un solo factor, sino de una combinación de variables (calificaciones previas, horas de estudio, etc.).

Se presenta como una extensión necesaria y potente de la regresión lineal simple para modelar fenómenos complejos, como el rendimiento académico en ingeniería. A diferencia de su versión simple, la Regresión Lineal Múltiple permite incorporar **múltiples variables predictoras** ( $X_1, X_2, \dots, X_k$ ) para explicar la varianza de una única variable de respuesta (Y) (Walpole et al., 2012).

La Regresión Lineal Múltiple facilita entender qué factores influyen de forma significativa en el rendimiento de los estudiantes y cuantificar su efecto relativo, lo que facilita intervenciones pedagógicas, diseño curricular y asignación de recursos (Hastie et al., 2009, 2017; James et al., 2013). En un contexto de ingeniería, donde la precisión y la interpretabilidad son críticas, la

Regresión Lineal Múltiple ofrece un equilibrio entre simplicidad y poder predictivo (Neter et al., 1996).

En el contexto de la ingeniería, la Regresión Lineal Múltiple sirve como la espina dorsal para el desarrollo de modelos que buscan identificar a estudiantes en riesgo y optimizar las intervenciones pedagógicas. Permite ir **más allá de las calificaciones** previas al incluir factores *psicosociales, demográficos y de comportamiento* en la predicción del éxito o fracaso en cursos de alta complejidad (Smith & Johnson, 2023).

En la mayoría de los problemas de investigación en los que se aplica el análisis de regresión se necesita más de una variable independiente para el modelo de regresión. La complejidad de la mayoría de mecanismos científicos es tal que, con el fin de predecir una respuesta importante, se requiere un modelo de regresión múltiple. Cuando un modelo es lineal en los coeficientes se denomina modelo de regresión lineal múltiple.

Para el caso de  $k$  variables independientes, el modelo que da  $x_1, x_2, \dots, x_k$ , la media de  $y | x_1, x_2, \dots, x_k$  es el modelo de regresión lineal múltiple (Walpole et al., 2012).

**Inclusión de Datos Múltiples (Variables Predictoras):** Un modelo predictivo robusto debe integrar una variedad de datos y evaluaciones consistentes:

**Datos Duros (Cognitivos):** Calificación de admisión, promedio del bachillerato (ej.  $X_2$ ), calificaciones en cursos de tronco común. La inclusión de variables no cognitivas como la autoeficacia ha demostrado ser crucial, ya que el conocimiento por sí solo no garantiza la persistencia. Se ha encontrado que una alta autoeficacia se correlaciona positivamente y significativamente con el éxito en ingeniería, incluso después de controlar el efecto de las calificaciones de ingreso (Jones et al., 2025).

**Datos Blandos (No Cognitivos/Psicosociales):** Horas de estudio semanal autodeterminado, **autoeficacia** en matemáticas y ciencias, **motivación intrínseca** para la carrera, y Nivel de apoyo social percibido.

**Evaluación de la Contribución Única e Interacciones:** la Regresión Lineal Múltiple permite determinar la **contribución única** de cada predictor a la varianza total explicada en el rendimiento. Determina la importancia relativa e independiente de cada predictor ( $X_i$ ) sobre la variable dependiente ( $Y$ ).

**Coefficiente de Determinación Ajustado ( $R^2$  Ajustado):** Este dato estadístico indica el **porcentaje de la variabilidad** en el Rendimiento Académico ( $Y$ ) que es **explicada** colectivamente por el conjunto de variables predictoras en el modelo, corrigiendo por el número de predictores utilizados (Field, 2018)

**Análisis de Interacción:** Se destaca por la importancia de modelar interacciones. Por ejemplo, la Regresión Lineal Múltiple puede incluir un término de interacción  $X_1 \times X_2$  (Calificación en Cálculo  $\times$  Horas de Estudio). El análisis podría revelar que las **Horas de Estudio** ( $X_3$ ) tienen un efecto predictivo significativamente **mayor** para aquellos estudiantes que **inicialmente** tuvieron

una **calificación baja** en Cálculo ( $X_1$ ). Esto sugiere que la intervención de tutoría y el esfuerzo extra son más críticos para este subgrupo (Cohen et al., 2018).

### Ejercicio Práctico: Modelado del Rendimiento en Termodinámica

Consideremos la predicción de la calificación final en el curso de **Termodinámica** (una materia crítica en muchas ramas de la ingeniería) usando tres predictores en una muestra de 50 estudiantes.

#### Variables del Modelo

**Variable Dependiente (Y):** Puntuación Final en Termodinámica (escala 0 a 100).

**$X_1$  (Cognitiva):** Calificación en Cálculo I (escala 0 a 100).

**$X_2$  (No Cognitiva):** Autoeficacia en Ingeniería (escala 1 a 10).

**$X_3$  (Comportamiento):** Horas de Estudio Semanal Reportadas (en horas).

#### Datos mínimos y estadísticos descriptivos

A continuación, se presenta un extracto de los datos y los estadísticos descriptivos clave para las  $N=50$  observaciones:

Estudiante	Y (Termodinámica)	$X_1$ (Cálculo I)	$X_2$ (Autoeficacia)	$X_3$ (Horas Estudio)
1	85	88	7	15
2	62	70	5	8
3	91	95	9	20
...	...	...	...	...
<b>50</b>	<b>78</b>	<b>82</b>	<b>6</b>	<b>12</b>

#### Estadísticos descriptivos Clave ( $N=50$ )

Variable	Media ( $\mu$ )	Desviación Estándar ( $\sigma$ )	Rango
Y (Termodinámica)	75,5	10,2	55-98
$X_1$ (Cálculo I)	78,0	8,5	60-95
$X_2$ (Autoeficacia)	6,8	1,5	4-10
$X_3$ (Horas Estudio)	14,0	4,1	5-25

#### Resultados del Análisis de Regresión Lineal Múltiple

Tras ejecutar el análisis de RLM (utilizando el método de Mínimos Cuadrados Ordinarios, MCO), se obtuvieron los siguientes resultados:

## Modelo de Regresión

Indicador	Valor
Coefficiente de Determinación (R <sup>2</sup> )	0,654
R <sup>2</sup> Ajustado	0,632
Error Estándar de Estimación	6,2
Valor F (ANOVA)	29,8
p-valor del Modelo	<0,001

**Interpretación del Resumen:** El modelo es estadísticamente significativo ( $p < 0.001$ ) y el R<sup>2</sup> Ajustado de **0.632** indica que el **63.2% de la variabilidad** en la Puntuación Final de Termodinámica es explicada colectivamente por las calificaciones en Cálculo I, la Autoeficacia y las Horas de Estudio.

### Coefficientes de Regresión

#### Coefficientes de la Regresión Lineal Múltiple

Predictor	Coefficiente B (No Estandarizado)	Error Estándar	t de Student	p-valor
Constante ( $\beta_0$ )	25,10	5,50	4,56	<0,001
Cálculo I ( $X_1$ )	0,45	0,08	5,63	<0,001
Autoeficacia ( $X_2$ )	2,80	0,70	4,00	<0,001

### Ecuación Predictiva

La ecuación de regresión lineal múltiple resultante es:

$$Y^{\wedge} = 25,10 + 0,45(X_1) + 2,80(X_2) + 0,35(X_3)$$

### Interpretación de los Coeficientes (El "Más Allá de las Calificaciones")

Los coeficientes ( $\beta$ ) en la RLM son **efectos parciales** (Cohen et al., 2018). Esta es la idea central del libro:

**Cálculo I ( $\beta_1=0,45$ ):** Por cada punto adicional obtenido en Cálculo I, la calificación esperada en Termodinámica aumenta en 0,45 puntos, **manteniendo constantes** la Autoeficacia y las Horas de Estudio. Esta es la contribución cognitiva base.

**Autoeficacia ( $\beta_2=2,80$ ):** Por cada unidad de aumento en la escala de Autoeficacia (ej. de 6 a 7), la calificación esperada en Termodinámica aumenta en **2,80 puntos, manteniendo constantes** las otras variables. ¡Este es el hallazgo clave! La Autoeficacia ( $X_2$ ) tiene un impacto predictivo **mucho mayor** que la calificación en Cálculo ( $X_1$ ) cuando se considera su contribución pura (Jones et al., 2025).

**Horas de Estudio ( $\beta_3=0,35$ ):** Por cada hora adicional de estudio semanal, la calificación esperada aumenta en 0,35 puntos, controlando el efecto de las otras dos variables.

Este ejercicio demuestra la potencia de la **Regresión Lineal Múltiple**: permite al profesor o ingeniero (analista) no solo predecir, sino también identificar qué palancas (ej. mejorar la Autoeficacia vía talleres o la base cognitiva vía tutorías) tienen el **mayor impacto independiente** en el rendimiento del estudiante.

**Regresión Logística**: Este modelo se utiliza cuando la variable dependiente es **dicotómica o binaria**, es decir, solo tiene dos resultados posibles (por ejemplo, "éxito" o "fracaso", "abandono" o "permanencia"). En lugar de predecir un valor numérico, la regresión logística predice la **probabilidad** de que ocurra uno de los dos resultados. Este modelo es especialmente relevante para predecir la deserción estudiantil, ya que el resultado es binario (el estudiante se queda o se va). Su fortaleza reside en la interpretación probabilística de las probabilidades de ocurrencia y en la capacidad de manejar variables mixtas (continuas y categóricas) de manera coherente en contextos educativos e ingenieriles (Hosmer et al., 2013).

En ingeniería educativa, la regresión logística permite traducir factores de entrada (calificaciones previas, tiempo de estudio, satisfacción con cursos, participación en prácticas, carga de laboratorios, sesgos estructurales, entre otros) en probabilidades de éxito académico, facilitando intervenciones focalizadas y políticas de apoyo (Agresti, 2018).

### Fundamentos teóricos de la regresión logística

**Modelo y ecuación**: en su forma probabilística, la regresión logística modela la probabilidad de éxito asumiendo que la log-odds (logit) es lineal en las covariables:  $\text{logit}[P(Y=1|X)] = \beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_kX_k$  donde  $\text{logit}[p] = \log(p / (1 - p))$ . Esta formulación permite interpretar coeficientes como cambios en el logaritmo de las probabilidades y, exponentiados, como cambios en las odds (Hosmer et al., 2013).

**Interpretación de coeficientes**: un coeficiente  $\beta_j$  representa el cambio en la log-odds de  $Y=1$  por cada unidad de  $X_j$ , ajustando por las demás variables. Expresado en odds ratio (OR),  $\Delta\text{OR} = \exp(\beta_j)$ . En contextos educativos, un OR mayor que 1 indica mayor probabilidad de éxito asociada a  $X_j$ , manteniendo constantes los demás factores (Agresti, 2018).

**Supuestos y naturaleza de la relación**: la regresión logística no asume linealidad de  $Y$  respecto a  $X$ , sino linealidad de la logit respecto a las covariables. Es compatible con relaciones no lineales si se incorporan transformaciones o términos de interacción adecuados (Hosmer et al., 2013; Menard, 2002).

**Evaluación de rendimiento**: a través de métricas como la curva ROC/AUC, accuracy, precisión, recall y calibración (Hosmer et al., 2013). En contextos de ingeniería educativa, la calibración es crucial para evitar sobreprometer intervenciones que no se sostienen en diferentes cohortes (Steyerberg, 2019).

### Diseño de modelos para rendimiento académico en ingeniería

**Selección de variables**: considerar antecedentes académicos (calificaciones previas), compromiso estudiantil (participación en prácticas, horas de estudio autogestionadas), factores del curso (dificultad percibida, carga de créditos), dinámicas de aprendizaje (autonomía, uso de recursos educativos), y variables institucionales (apoyo académico, tutorías). También se deben

contemplar sesgos potenciales, como sesgos de selección y omisión de variables relevantes (Hastie et al., 2009).

**Codificación de variables:** variables continuas pueden conservarse como tales o transformarse (cuantiles, polinomios). Variables categóricas deben codificarse correctamente (dummies). La jerarquía de variables (p. ej., interacción entre carga de créditos y tutoría) puede revelar efectos condicionados (Peng et al., 2002).

**Manejo de desequilibrio:** si el resultado “éxito” es poco frecuente, técnicas como ponderación de clases o muestreo estratificado pueden ayudar a estabilizar estimaciones y mejorar la sensibilidad en detectar casos de alto riesgo (Sindhu Priya, 2021).

**Validación y generalización:** separación entre conjuntos de entrenamiento y prueba, validación cruzada (k-fold) y, cuando sea posible, validación externa con cohortes diferentes para evaluar la transportabilidad del modelo (James et al., 2013).

Ejemplo ilustrativo de datos y resultados (con formato narrativo)

Supuesto conjunto de datos: una cohorte de 1.200 estudiantes de ingeniería en dos semestres, con variables:

Y: 1 = aprobado final con nota  $\geq 70$ ; 0 = aprobado con nota  $< 70$  o reprobado.

X1: promedio de calificaciones de cursos previos (0-100).

X2: horas semanales de estudio autogestivo (0-25).

X3: participación en laboratorios prácticos (0 = ninguno, 1 = moderada, 2 = alta).

X4: acceso a tutoría (0 = no, 1 = sí).

X5: año de estudio (1 a 4).

Modelo estimado (ejemplo hipotético):  $\text{logit}[P(Y=1)] = -1,20 + 0,04X_1 + 0,10X_2 + 0,65X_3 + 0,75X_4 + 0,18X_5$

### Interpretaciones:

Cada incremento de 1 punto en X1 (promedio previo) se asocia con un aumento en la log-odds de aprobar en 0.04, manteniendo X2-X5 constantes.

La participación en laboratorios (X3) tiene un efecto relativamente fuerte ( $\beta_3 = 0,65$ ), indicando mayor probabilidad de éxito con mayor involucramiento en prácticas.

El uso de tutoría (X4) tiene un coeficiente de 0,75, sugiriendo un impacto sustancial de intervenciones de apoyo en rendimiento.

Evaluación de rendimiento: AUC 0,82 en validación cruzada, calibración aceptable mediante la prueba de Hosmer-Lemeshow ( $p > 0,05$ ), y una precisión global del 78%. Estas métricas sugieren que el modelo discrimina bien entre aprobados y no aprobados y está razonablemente calibrado para cohortes similares (Hosmer et al., 2013; Harrell, 2015).

## Interpretación para políticas e intervenciones en ingeniería

**Identificación de factores clave:** si la interacción entre  $X_3$  (laboratorios) y  $X_4$  (tutoría) resulta significativa, se podrían diseñar intervenciones combinadas que fortalezcan la participación en prácticas con apoyo tutorado para maximizar la probabilidad de éxito.

**Priorización de recursos:** coeficientes de mayor magnitud o mayor significancia estadística señalan áreas donde invertir recursos podría tener mayor impacto (p. ej., expandir programas de tutoría o reforzar laboratorios prácticos) (James et al., 2013).

**Seguimiento y ajuste continuo:** reentrenar el modelo con cohortes nuevas cada ciclo para detectar cambios en patrones de ingeniería educativa y mantener la validez de las predicciones (Steyerberg, 2019).

## Ventajas y limitaciones de la regresión logística en este contexto

### Ventajas

Interpretabilidad: coeficientes y odds ratios ofrecen interpretaciones claras para la toma de decisiones y diseño de intervenciones (Hosmer et al., 2013).

Flexibilidad: manejo de variables mixtas y capacidad de incorporar interacciones y transformaciones razonables (Agresti, 2018).

Requisitos de datos relativamente modestos en comparación con modelos complejos no lineales.

### Limitaciones

Supone linealidad en la logit; relaciones complejas pueden requerir transformaciones o modelos no lineales si no se especifican adecuadamente (Peng et al., 2002).

Sensible a la calidad de los datos y a la correcta especificación de variables; sesgos y variables omitidas pueden sesgar las estimaciones (James et al., 2013).

No captura de estructuras jerárquicas o longitudinales si no se extiende a modelos multinivel o de series temporales (Carlin, 2025).

## Relevancia “Más Allá de las Calificaciones”

### Conceptualizar la regresión logística como base permite

Construir un marco pedagógico para estudiantes de ingeniería que vincule métricas académicas a intervenciones prácticas de apoyo.

Proporcionar una guía para diseñar políticas institucionales basadas en evidencia y para comunicar resultados a partes interesadas (estudiantes, docentes, administradores).

Integrar prácticas de validación, interpretación y ética en la predicción de rendimiento, enfatizando la necesidad de evitar sesgos y de garantizar justicia en la asignación de recursos educativos (Kleinberg et al., 2017).

**Enfoques complementarios:** se pueden incorporar secciones sobre técnicas de regularización (Lasso/Ridge) para selección de variables, evaluación de calibración avanzada, y comparación con enfoques de aprendizaje automático cuando la complejidad de los datos lo justifique (Kleinbaum et al., 2013).

**Transparencia y replicabilidad:** compartir código y conjuntos de datos anonimizados cuando sea posible y describir las decisiones de preprocesamiento y validación para facilitar la replicación (Peng et al., 2002).

## La Regresión Logística y su importancia en la educación

La regresión logística es una técnica estadística que se utiliza ampliamente en el análisis de datos para modelar la probabilidad de un evento binario, lo cual es especialmente relevante en el contexto educativo. En el ámbito de la educación universitaria de ingeniería, esta metodología permite a los educadores y administradores prever el desempeño académico de los estudiantes al analizar múltiples factores influyentes. Como se menciona en la literatura, “la regresión logística proporciona un marco robusto para entender cómo diferentes variables impactan la probabilidad de éxito o fracaso académico” (Carlin & Moreno-Betancur, 2025).

La **regresión logística** (Solis Ventura et al., 2022; Chacha et al., 2023) es particularmente poderosa para la predicción del fracaso académico. Al modelar la probabilidad de que un estudiante abandone sus estudios, las instituciones pueden obtener una medida de riesgo individualizada. Esta probabilidad se calcula a partir de un conjunto de variables como el promedio de calificaciones de ingreso, el estatus socioeconómico, la asistencia a clases o el número de asignaturas reprobadas en el primer semestre. Como se menciona en la literatura, “la regresión logística proporciona un marco robusto para entender cómo diferentes variables impactan la probabilidad de éxito o fracaso académico” (Smith et al., 2020).

El resultado del modelo no es solo un "sí" o un "no", sino un valor de probabilidad entre 0 y 1. Por ejemplo, un estudiante con una probabilidad de 0,8 de abandonar es considerado de alto riesgo y podría beneficiarse de una intervención temprana, como una tutoría personalizada o asesoramiento psicológico. Este enfoque proactivo, sustentado en la estadística, es clave para la retención estudiantil y la mejora de la calidad educativa en la era digital. Al incorporar este enfoque en la curricula, las universidades pueden preparar a los futuros profesionales para enfrentar problemas reales donde la incertidumbre y las probabilidades de eventos discretos son parte central del diseño y la operación de sistemas complejos (James et al., 2013; Hosmer et al., 2013).

Además, este enfoque analítico no solo ayuda a identificar a los estudiantes en riesgo, sino que también permite a las instituciones diseñar programas de apoyo adecuados. Rodríguez-Hernández et al. (2021) argumenta que “la implementación de modelos predictivos como la regresión logística puede guiar la creación de talleres y tutorías específicas, dirigidas a aquellos alumnos que presentan características asociadas con bajo rendimiento”. De esta manera, se genera un entorno académico más inclusivo y eficiente, alineado con las necesidades individuales de los estudiantes.

Uno de los aspectos más significativos de la regresión logística es su capacidad para manejar tanto variables categóricas como continuas, permitiendo una comprensión integral del

rendimiento estudiantil. Según Torres Lopez et al. (2025), “el uso de la regresión logística en estudios educativos facilita la identificación de patrones en los datos, lo que resulta crucial para formulaciones estratégicas que fomenten el éxito de los estudiantes de ingeniería”. Esto se traduce en la posibilidad de implementar intervenciones personalizadas, optimizando así los recursos educativos disponibles.

En contextos de ingeniería, este enfoque permite modelar decisiones discretas relevantes para el diseño, la confiabilidad y el control de procesos, tales como la probabilidad de fallo de un componente, la clasificación de materiales según desempeño esperado o la detección de anomalías en redes energéticas (Hosmer et al., 2013).

En la práctica educativa, la regresión logística se puede integrar en cursos de estadística, ciencia de datos e ingeniería de métodos experimentales mediante proyectos de laboratorio, análisis de fiabilidad y evaluación de riesgos. Por ejemplo, un proyecto de ingeniería eléctrica podría utilizar regresión logística para clasificar señales defectuosas en una línea de producción o para predecir la probabilidad de fallo de componentes sometidos a diferentes condiciones de operación.

La importancia de la regresión logística en la educación universitaria de ingeniería se manifiesta en varios ejes. Primero, fortalece el razonamiento cuantitativo y la habilidad de traducir información compleja en decisiones operativas: se puede estimar la probabilidad de éxito de un proyecto de ingeniería basándose en variables de entrada como temperatura de operación, voltaje, tolerancias de fabricación o historial de fallas (James et al., 2013). Segundo, promueve el aprendizaje de modelos de predicción interpretables, lo cual es crucial en ingeniería para justificar decisiones ante comités de seguridad, clientes y reguladores. En este marco, la interpretación de los coeficientes logísticos y las pruebas de bondad de ajuste permiten a los estudiantes evaluar la robustez de sus recomendaciones y comunicar incertidumbres de forma rigurosa (Hosmer et al., 2013).

En resumen, la regresión logística se establece como una herramienta esencial en la educación universitaria de ingeniería, ya que permite realizar un análisis profundo de diversos factores que afectan el rendimiento académico. Su aplicación no solo contribuye a la mejora de los resultados académicos, sino que también promueve un enfoque proactivo en la gestión educativa, facilitando el desarrollo profesional de los futuros ingenieros (González & Rodríguez, 2023).

# Capítulo 3

## Metodología para la Construcción de un Modelo Predictivo

La construcción de un modelo predictivo, como el de regresión logística, no es un proceso lineal, sino una metodología estructurada que requiere seguir pasos rigurosos para garantizar la fiabilidad y validez de los resultados. En esta fase inicial, se delimita el objetivo del modelo predictivo: estimar o predecir el rendimiento académico de estudiantes en programas de ingeniería. La variable dependiente comúnmente es la calificación final o promedio ponderado acumulado, mientras que las variables independientes pueden incluir factores académicos, personales y sociales que influyen en el desempeño (Romero & Ventura, 2010).

### **Ideas para el Modelo Predictivo en Ingeniería:**

- ✓ Incluir variables específicas del contexto de ingeniería, como desempeño en laboratorios, proyectos y participación en actividades prácticas.
- ✓ Incorporar datos longitudinales para capturar la evolución del rendimiento a lo largo del tiempo.
- ✓ Evaluar factores psicológicos como la resiliencia y la autoeficacia, que impactan el aprendizaje en carreras exigentes.
- ✓ Utilizar modelos interpretables para facilitar la comprensión por parte de docentes y orientadores.

A continuación, se desglosan las etapas claves para la creación de un modelo predictivo de rendimiento académico, basándose en prácticas comunes en la investigación estadística:

### **Etapas del proceso de investigación:**

#### ***Definición del problema y de la variable de estudio***

El primer paso es crucial: **definir con precisión el problema y la variable que se busca predecir**. Define claramente el objetivo predictivo ósea predecir la probabilidad de aprobar un curso, el rendimiento final (nota media), o la probabilidad de abandono, en este caso, el problema es el bajo rendimiento académico o la deserción estudiantil. La variable de estudio, que será la variable dependiente de nuestro modelo, debe ser conceptualizada y medida de forma clara. Por ejemplo, se puede definir el "bajo rendimiento" como una variable dicotómica (sí/no) donde el "sí" se asigna a estudiantes que reprueban un número específico de asignaturas en un periodo determinado o que tienen un promedio de calificaciones por debajo de un umbral predefinido. La claridad en esta etapa es fundamental para que el resto del

Establece métricas de éxito alineadas con el objetivo: para clasificación (aprobar/reprobar) considerar AUC-ROC, precisión, recall; para regresión considerar RMSE, MAE; para retención

considerar tasas de supervivencia o predicción de abandono con curvas de supervivencia (Romero & Ventura, 2010).

Considera limitaciones operativas: disponibilidad de datos, sesgos, privacidad y regulaciones (FERPA, GDPR según la región) y requisitos de interpretabilidad para la toma de decisiones por docentes y directivos (Romero & Ventura, 2010).

Esto puede medirse mediante el promedio ponderado, la aprobación de asignaturas clave o la permanencia en el programa. Según Del Carpio-Mendoza (2024), establecer una métrica clara permite seleccionar las variables más relevantes y evaluar la eficacia del modelo con mayor precisión.

Es crucial acordar la ventana temporal, la granularidad (curso, semestre, año) y el público objetivo (estudiantes de ingeniería, por cohortes, por campus). En educación, la literatura indica la importancia de distinguir predicción de rendimiento de evaluación de procesos de aprendizaje y de retención (Romero & Ventura, 2008; Papamitsiou & Papamitsiou, 2013).

El diseño de un Modelo Predictivo del Rendimiento Académico en Ingeniería es un proceso riguroso que fusiona la ciencia de datos con la pedagogía y la gestión educativa. Esta metodología está estructurada para ser aplicable tanto en entornos académicos (detección temprana de riesgo, apoyo focalizado) como industriales (optimización de la fuerza laboral, desarrollo de programas de capacitación)

### ***Recopilación y limpieza de datos***

Una vez definido el problema, se procede a la recopilación de datos fiables y relevantes. Es fundamental utilizar métodos de recolección que aseguren la validez y confiabilidad de la información. Los cuestionarios estandarizados y las bases de datos académicas son herramientas útiles en este proceso (Cohen et al., 2018). La calidad de los datos recolectados influirá directamente en la precisión del modelo predictivo. La selección adecuada de variables es crucial, ya que un modelo sobrecargado puede llevar a un sobreajuste, mientras que un modelo demasiado simple puede no capturar la variabilidad necesaria (Field, 2013).

Estos datos pueden provenir de registros académicos, encuestas socioeconómicas, o bases de datos institucionales. La calidad y representatividad de los datos son esenciales para garantizar la validez del modelo (James et al., 2013).

Los datos deben ser recolectados al inicio del curso y normalizados para evitar sesgos. Esto incluye la codificación de variables categóricas, la imputación de valores faltantes y la detección de outliers. En el estudio de Aricop et al. (2022), se utilizó una muestra de 228 estudiantes de ingeniería, y se aplicaron técnicas de limpieza y transformación para mejorar la calidad del dataset.

Sin embargo, los datos brutos rara vez están listos para el análisis. La fase de limpieza de datos es vital e implica:

- **Manejo de datos perdidos (missing data):** Identificar y tratar los valores faltantes. Esto puede hacerse eliminando registros incompletos o imputando valores a partir de técnicas estadísticas.

- **Codificación de variables:** Convertir variables cualitativas (como "género" o "carrera") a un formato numérico que el modelo pueda procesar. Por ejemplo, "Femenino" y "Masculino" pueden codificarse como 0 y 1.
- **Identifica fuentes de datos:** resultados históricos de cursos, calificaciones, asistencia, participación en plataformas de aprendizaje, carga de trabajo, tiempos de estudio, uso de plataformas virtuales, antecedentes académicos, datos demográficos y posibles indicadores de bienestar (estrés, salud, etc. si están disponibles y autorizados).
- **Asegura calidad y consistencia:** estandarización de formatos, tratamiento de duplicados, manejo de valores ausentes y revisión de sesgos (p. ej., sesgos por género o procedencia) antes de la modelización Romero & Ventura, (2010).
- **Gobernanza y ética:** define permisos, anonimización, minimización de datos y políticas de uso, así como un plan de supervisión para evitar usos indebidos de los modelos (Romero & Ventura, 2010).

### ***Análisis Bivariado: La búsqueda de relaciones***

Antes de construir el modelo final, es útil realizar un **análisis bivariado**. Esta etapa consiste en examinar la relación entre cada variable predictora candidata y la variable de resultado de forma individual, es un procedimiento estadístico que examina la relación entre dos variables para identificar patrones, asociaciones o diferencias significativas. En el contexto del rendimiento académico en ingeniería, este análisis permite descubrir cómo ciertos factores influyen directamente en el desempeño de los estudiantes (Field, 2018).

Este análisis es fundamental para identificar variables relevantes y comprender su impacto en el rendimiento académico. Además, sirve como base para construir modelos predictivos más complejos y diseñar intervenciones educativas efectivas (Romero & Ventura, 2010). Por ejemplo, se podría usar una prueba t de Student para comparar el promedio de calificaciones de ingreso entre los estudiantes que tuvieron bajo rendimiento y los que no, o un Chi-cuadrado para ver si existe una asociación entre el género y la deserción. Este análisis ayuda a identificar las variables que tienen una relación significativa con el bajo rendimiento académico, lo cual es útil para la siguiente etapa de selección de variables.

El propósito general es evaluar la existencia, dirección y fuerza de la relación entre una variable dependiente (por ejemplo, la calificación final) y una variable independiente (como horas de estudio, asistencia, o variables socioeconómicas) para obtener insights que apoyen la toma de decisiones educativas (Martínez Pérez et al., 2021). El primer paso en la construcción de un modelo multivariado es la identificación de las variables relevantes. En el contexto del rendimiento académico, estas pueden incluir variables demográficas (edad, género), antecedentes académicos (calificaciones previas, tipo de escuela), y factores psicológicos (motivación, autoconfianza (Martínez Pérez et al., 2021).

La selección adecuada de variables es crucial, ya que un modelo sobrecargado puede llevar a un sobreajuste, mientras que un modelo demasiado simple puede no capturar la variabilidad necesaria (Field, 2013).

## Importancia del análisis bivariado en la Ingeniería

El análisis bivariado es más que un simple ejercicio estadístico; es una herramienta de Ingeniería de Características (*Feature Engineering*) y de Selección de Variables (*Feature Selection*).

**Reducción de dimensionalidad:** Permite descartar variables que tienen una correlación cercana a cero con la variable objetivo, simplificando el modelo y mejorando su interpretabilidad y tiempo de entrenamiento (Contreras et al., 2020).

**Identificación de colinealidad:** También sirve para verificar si existen altas correlaciones entre las propias variables predictoras (problema de multicolinealidad). Por ejemplo, si el "Promedio de Cálculo I" y el "Promedio de Física I" están altamente correlacionados, podríamos optar por incluir solo una o crear un índice combinado para evitar la redundancia y estabilizar el modelo (Díaz et al., 2021).

**Base para la intervención:** Las variables que demuestran la relación bivariada más fuerte con el rendimiento suelen ser los puntos focales para el diseño de intervenciones educativas tempranas (Álvarez et al., 2024).

## Construcción del Modelo Multivariado: Juntando las piezas

La construcción de un modelo multivariado implica integrar múltiples variables predictoras para explicar y predecir con mayor precisión el rendimiento académico en estudiantes de ingeniería. Este proceso es crucial para capturar la complejidad y multidimensionalidad de los factores que influyen en el desempeño académico (James et al., (2013). Este enfoque permite capturar la complejidad del aprendizaje y las interacciones entre diferentes factores. Según Hair et al. (2010), los modelos multivariados son esenciales para entender relaciones complejas y para la toma de decisiones informadas en contextos educativos.

Con las variables candidatas identificadas, se procede a la construcción del modelo multivariado. Para un modelo de regresión logística, esto implica seleccionar el conjunto óptimo de variables predictoras. Las técnicas de selección paso a paso (*stepwise selection*) son comunes en este proceso. Estas técnicas agregan o eliminan variables del modelo de forma iterativa, basándose en su significancia estadística (valor  $p$ ) hasta que se alcanza un modelo final con las variables más predictivas. Este enfoque ayuda a evitar la inclusión de variables redundantes o no informativas, lo que podría sobreajustar el modelo.

## Interpretación de los resultados del modelo

Una vez que el modelo está construido, la interpretación de sus resultados es clave para entender las relaciones que se han modelado.

Finalmente, la implementación del modelo en un entorno real debe ir acompañada de un monitoreo continuo. Esto implica evaluar regularmente la efectividad del modelo y realizar ajustes según sea necesario. La retroalimentación de los usuarios, como profesores y estudiantes, puede proporcionar información valiosa para mejorar el modelo (Baker & Inventado, 2014).

✓ **Coefficientes (B):** Representan el cambio en el logaritmo de las *odds* (probabilidades) por cada unidad de aumento en la variable predictora. Un coeficiente

positivo indica que la variable aumenta la probabilidad del resultado (ej., bajo rendimiento), mientras que uno negativo la disminuye.

✓ **Odds Ratio (OR):** Es la interpretación más intuitiva. Se calcula como la exponencial del coeficiente ( $e^B$ ). Un Odds Ratio mayor que 1 indica que la variable aumenta las *odds* del resultado. Por ejemplo, un OR de 1.5 para la variable "reprobó asignaturas el primer semestre" significa que un estudiante que reprobó tiene 1.5 veces las *odds* de tener bajo rendimiento en comparación con uno que no reprobó, manteniendo las demás variables constantes.

✓ **Significancia Estadística (valor  $p$ ):** Indica si la relación entre la variable predictora y el resultado es estadísticamente significativa. Un valor  $p$  menor a 0.05 generalmente se considera significativo, lo que sugiere que el efecto observado no se debe al azar.

✓ **Pruebas de Bondad de Ajuste:** Estas pruebas evalúan qué tan bien el modelo predice los datos. La prueba de Hosmer-Lemeshow es una de las más utilizadas en regresión logística. Un valor  $p$  alto (mayor a 0.05) en esta prueba indica que el modelo se ajusta bien a los datos, lo que significa que no hay diferencias significativas entre las probabilidades predichas por el modelo y las probabilidades observadas.

### Selección de variables predictoras

Se deben seleccionar variables relevantes que, según análisis bivariados y conocimiento previo, tengan impacto en el rendimiento académico. Estas pueden incluir:

- **Datos académicos:** notas parciales, asistencia. Los cuestionarios estandarizados y las bases de datos académicas son herramientas útiles en este proceso (Cohen et al., 2018). La calidad de los datos recolectados influirá directamente en la precisión del modelo predictivo.
- **Factores personales:** motivación, hábitos de estudio.
- **Contexto socioeconómico:** nivel educativo de los padres, ingreso familiar.
- **Aspectos institucionales:** modalidad de estudio, recursos disponibles (Romero & Ventura, 2010).

La selección puede apoyarse en técnicas estadísticas como la regresión paso a paso o métodos automáticos de selección (Kuhn & Johnson, 2019).

En el contexto del rendimiento académico, estas pueden incluir variables demográficas (edad, género), antecedentes académicos (calificaciones previas, tipo de escuela), y factores psicológicos (motivación, autoconfianza) (Pérez & Salamanca, 2018).

### Modelado Multivariado

El análisis estadístico es el corazón de la construcción del modelo. Técnicas como la regresión múltiple, el análisis de varianza (ANOVA) y el análisis de componentes principales (PCA) son comúnmente utilizadas para explorar las relaciones entre las variables y el rendimiento académico (Tabachnick & Fidell, 2013).

### Los modelos multivariados comunes incluyen:

- ✓ **Regresión Lineal Múltiple:** para explicar variables continuas como la nota final.
- ✓ **Regresión Logística Multinomial:** para clasificar a estudiantes en categorías de rendimiento (alto, medio, bajo).
- ✓ **Árboles de Decisión y Random Forest:** para capturar relaciones no lineales y efectos de interacción.
- ✓ **Redes Neuronales Artificiales:** para patrones complejos y no lineales (James et al., 2013).

### Ajuste y validación

Se entrena el modelo con un conjunto de datos y se valida con datos independientes para evitar sobreajuste. Se ajustan hiperparámetros y se aplican técnicas como validación cruzada para garantizar robustez (Kuhn & Johnson, 2019). La validación del modelo es un paso crítico que garantiza que el modelo no solo se ajuste a los datos de entrenamiento, sino que también sea capaz de predecir con precisión en nuevos conjuntos de datos. Esto se puede realizar mediante técnicas como la validación cruzada y el uso de conjuntos de prueba (Kohavi, 1995). La validación asegura que el modelo sea robusto y generalizable, lo que es esencial en el ámbito educativo.

La validación cruzada, la matriz de confusión y métricas como la precisión, el recall y el F1-score son esenciales para evaluar el rendimiento del modelo. (Rico Páez, 2023) sugiere realizar evaluaciones en distintas etapas del curso para detectar problemas tempranos y aplicar intervenciones.

### Interpretación de resultados

Se analizan los coeficientes, importancia de variables y efectos de interacción para entender qué factores contribuyen significativamente al rendimiento. Esto ayuda a diseñar estrategias educativas más efectivas (Baker & Inventado, 2014).

Una vez validado, el modelo debe ser interpretado para identificar las variables más influyentes. Esto permite diseñar estrategias pedagógicas, tutorías personalizadas y alertas tempranas. Del Carpio-Mendoza (2024) enfatiza que los modelos predictivos no solo deben ser precisos, sino también comprensibles y aplicables en contextos reales.

### Implementación y seguimiento

El modelo se implementa en plataformas institucionales para predecir el rendimiento y detectar estudiantes en riesgo. Se recomienda monitorear y actualizar el modelo periódicamente con nuevos datos (Romero & Ventura, 2010).

La construcción de un modelo multivariado para predecir el rendimiento académico en ingeniería es un proceso complejo que requiere una cuidadosa consideración de múltiples factores. A través de la identificación de variables, la recolección de datos, el análisis estadístico, la validación

y la implementación, se pueden desarrollar modelos que no solo sean precisos, sino también útiles para mejorar los resultados académicos.

### **Ideas para fortalecer el modelo multivariado en ingeniería**

Esto puede medirse mediante el promedio ponderado, la aprobación de asignaturas clave o la permanencia en el programa. Según Del Carpio-Mendoza (2024), establecer una métrica clara permite seleccionar las variables más relevantes y evaluar la eficacia del modelo con mayor precisión.

Incorporar variables longitudinales para captar la evolución del rendimiento.

- ✓ Usar técnicas de reducción de dimensionalidad para manejar gran cantidad de variables, igualmente técnicas de ensamblado (ensemble learning) para mejorar la robustez.
- ✓ Integrar datos cualitativos codificados, como evaluaciones de profesores o autoevaluaciones.
- ✓ Desarrollar modelos explicativos y no solo predictivos para facilitar la toma de decisiones pedagógicas.
- ✓ Incluir variables afectivas como motivación y ansiedad académica.
- ✓ Desarrollar dashboards interactivos para visualizar el riesgo académico.
- ✓ Aplicar el modelo en diferentes cohortes para validar su generalización.

### **Técnicas de aprendizaje automático utilizadas:**

Diversos estudios han empleado técnicas de *machine learning* para construir modelos predictivos. Entre las más destacadas se encuentran:

- ✓ **Naïve Bayes:** mostró una precisión del 65% y fue especialmente útil para identificar estudiantes en peligro de reprobación (Rico Páez, 2023).
- ✓ **Árboles de decisión y k vecinos más cercanos:** permitieron clasificar el rendimiento con base en patrones históricos.
- ✓ **Redes neuronales y máquinas de vectores de soporte:** demostraron alta capacidad de generalización en contextos complejos (Del Carpio-Mendoza, 2024).

Estas técnicas se alimentan de datos recolectados al inicio del curso, como promedios previos, hábitos de estudio y condiciones socioeconómicas, lo que permite realizar predicciones con suficiente antelación para actuar sobre ellas.

### **Resultados y aplicabilidad**

Los modelos progresivos, que se actualizan conforme avanza el curso, han demostrado ser particularmente efectivos. En un estudio con 260 estudiantes, se logró una precisión del 70.5% en tan solo el 21% del tiempo total del curso, lo que valida su utilidad para intervenciones tempranas (Rico Páez, 2023). Además, se concluye que el promedio académico previo es la variable más significativa en la predicción del rendimiento.

# Capítulo 4

## Aplicación Práctica: Un Caso de Estudio en Ingeniería

El objetivo de la investigación fue identificar las variables significativas que inciden en el rendimiento académico de los estudiantes en los primeros semestres de estudios de una escuela profesional de ingeniería en Perú, utilizando regresión logística y el análisis discriminante. Para el estudio con enfoque cuantitativo, de tipo proyectiva y predictiva se aplicó un cuestionario a 360 estudiantes universitarios durante los semestres 2022 I, 2022 II y 2023 I. Utilizando un modelo de regresión logística, se obtuvo que las características individuales y socioeconómicas si lograron explicar de manera conjunta el desempeño académico ( $p$ -valor  $< 0.05$ ), así también el número de asignaturas aprobadas y el promedio de calificaciones obtenidas ( $p$ -valor  $< 0.05$ ); mientras que el promedio de calificaciones del nivel secundario, el tipo de institución donde cursó estudios secundarios (público o privado) y la localidad donde radica la institución donde estudió la secundaria no inciden en el rendimiento académico ( $p$ -valor  $> 0.05$ ).

Se concluye que la institución de procedencia del estudiante, el nivel educativo de los padres, el nivel de ingreso, la cantidad de asignaturas matriculadas en el semestre académico y las notas de la educación secundaria no figuran como representativas, por tanto, su incidencia en la variable rendimiento académico es mínima. En estudiantes de ingeniería, donde existen variables como horas de estudio, asistencia, antecedentes académicos y engagement, se pueden construir modelos que:

- Predecir la nota final o el rendimiento en una asignatura.
- Detectar a tiempo estudiantes en riesgo de desertar o con bajo rendimiento para aplicar intervenciones oportunas.

La investigación identificó que las características individuales y socioeconómicas, junto con el número de asignaturas aprobadas y el promedio de calificaciones universitarias, son variables significativas que inciden en el rendimiento académico de estudiantes de ingeniería en Perú. En cambio, factores como el tipo de colegio secundario, las notas escolares previas y el nivel educativo de los padres no mostraron influencia estadísticamente significativa ( $p > 0.05$ ), lo que indica una baja incidencia en el desempeño académico universitario.

### Resumen

El objetivo de la investigación fue identificar las variables significativas que inciden en el rendimiento académico de los estudiantes en los primeros semestres de estudios de una escuela profesional de ingeniería en Perú, utilizando regresión logística y el análisis discriminante. Para el estudio con enfoque cuantitativo, de tipo proyectiva y predictiva se aplicó un cuestionario a 360 estudiantes universitarios durante los semestres 2022 I, 2022 II y 2023 I. Utilizando un modelo de regresión logística, se obtuvo que las características individuales y socioeconómicas si lograron

explicar de manera conjunta el desempeño académico ( $p$ -valor  $< 0.05$ ), así también el número de asignaturas aprobadas y el promedio de calificaciones obtenidas ( $p$ -valor  $< 0.05$ ); mientras que el promedio de calificaciones del nivel secundario, el tipo de institución donde cursó estudios secundarios (público o privado) y la localidad donde radica la institución donde estudió la secundaria no inciden en el rendimiento académico ( $p$ -valor  $> 0.05$ ). Se concluye que la institución de procedencia del estudiante, el nivel educativo de los padres, el nivel de ingreso, la cantidad de asignaturas matriculadas en el semestre académico y las notas de la educación secundaria no figuran como representativas, por tanto, su incidencia en la variable rendimiento académico es mínima.

**Palabras clave:** Rendimiento académico, regresión logística, desempeño académico, nivel educativo.

## 1. Introduction

Todos los países del mundo destacan que el desarrollo económico sostenible es la primera prioridad para planificar el desarrollo de las naciones a largo plazo a partir de un sistema educativo de calidad (Chinnakum et al., 2024). Un sistema educativo de alto nivel en todos los niveles en los países mejoraría la mano de obra calificada y educada sobre la base de altos estándares, lo que puede contribuir a la productividad y al rendimiento económico. Estos factores son importantes para el crecimiento económico sostenido a largo plazo, junto con la alineación con el conocimiento de alto nivel que contribuye a la protección del medio ambiente y al desarrollo del bienestar social a través de un conocimiento de alta calidad.

El rendimiento académico de los estudiantes universitarios es un indicador clave para evaluar la efectividad del proceso educativo. Este rendimiento, entendido como el conjunto de conocimientos y competencias adquiridas durante un periodo determinado, se mide comúnmente a través de calificaciones que reflejan el desempeño en tareas, exámenes y actividades prácticas (Kumar et al., 2021). Sin embargo, diversos factores personales, sociales y contextuales influyen en el éxito académico, más allá de la inteligencia o el esfuerzo individual (Vázquez et al., 2024a). Por ello, es fundamental que las políticas educativas, como las promovidas por el Ministerio de Educación del Perú, prioricen el logro de competencias integrales que permitan a los estudiantes aplicar conocimientos, habilidades y actitudes en la solución de problemas reales (Incio-Flores & Capuñay-Sánchez, 2023).

La educación y, en particular, la educación superior cumple un rol importante dentro de la consecución de los objetivos de desarrollo de un país. Las funciones que cumplen las instituciones de educación superior (IES), esencialmente la docencia y la investigación, no solamente garantizan a la población mejores oportunidades laborales, y con ello, mejores oportunidades económicas, sociales, etc., sino que también contribuyen al desarrollo productivo, empresarial, y económico (Valero & Van Reenen, 2019). De esta forma, la educación superior constituye en la actualidad uno de los instrumentos principales con que cuentan los poderes públicos en su intento de asegurar el desarrollo de sus países, pues el gasto público destinado a esta finalidad es considerado una inversión de futuro (Medina et al., 2021a).

El desarrollo económico sostenible de una nación está estrechamente vinculado con la

calidad de su sistema educativo, especialmente en el nivel superior. La educación superior no solo forma profesionales altamente capacitados, sino que también impulsa la investigación, la innovación y la transferencia de conocimiento, elementos esenciales para el crecimiento productivo y tecnológico de los países (Valero & Van Reenen, 2019). En este sentido, las instituciones de educación superior (IES) cumplen un rol estratégico al generar capital humano calificado, capaz de responder a los desafíos globales y contribuir al bienestar social y ambiental (Chinnakum et al., 2024).

El rendimiento académico de los estudiantes universitarios es un indicador clave para evaluar la efectividad del proceso educativo. Este rendimiento, entendido como el conjunto de conocimientos y competencias adquiridas durante un periodo determinado, se mide comúnmente a través de calificaciones que reflejan el desempeño en tareas, exámenes y actividades prácticas (Kumar et al., 2021). Sin embargo, diversos factores personales, sociales y contextuales influyen en el éxito académico, más allá de la inteligencia o el esfuerzo individual (Vázquez et al., 2024b). Por ello, es fundamental que las políticas educativas, como las promovidas por el Ministerio de Educación del Perú, prioricen el logro de competencias integrales que permitan a los estudiantes aplicar conocimientos, habilidades y actitudes en la solución de problemas reales (Incio-Flores & Capuñay-Sánchez, 2023).

En este marco, el uso de modelos predictivos del rendimiento académico se presenta como una herramienta valiosa para identificar estudiantes en riesgo, anticipar dificultades y diseñar estrategias de intervención temprana. Estos modelos, basados en técnicas de inteligencia artificial, permiten analizar grandes volúmenes de datos y generar patrones que orientan la toma de decisiones en las IES, contribuyendo así a mejorar la calidad educativa y a reducir la deserción estudiantil (Medina et al., 2021b).

En este marco, el uso de modelos predictivos del rendimiento académico se presenta como una herramienta valiosa para identificar estudiantes en riesgo, anticipar dificultades y diseñar estrategias de intervención temprana. Estos modelos, basados en técnicas de inteligencia artificial, permiten analizar grandes volúmenes de datos y generar patrones que orientan la toma de decisiones en las IES, contribuyendo así a mejorar la calidad educativa y a reducir la deserción estudiantil (Medina et al., 2021b).

En todos los sistemas educativos, tanto básicos como superiores, un elemento crucial es el sistema de evaluación del aprendizaje, el cual permite medir la efectividad del proceso educativo. Desde hace mucho tiempo se ha considerado que el rendimiento académico es la condición previa más importante para la prosperidad individual y comunitaria, así como para la salud física y mental. En el ámbito de la educación, Kumar et al. (2021) definen el rendimiento académico como el conjunto de conocimientos que los estudiantes adquieren durante un periodo determinado, el cual se evalúa a través de las calificaciones otorgadas por los profesores. Generalmente, este rendimiento se mide a través del desempeño acumulado durante cada periodo académico, que se traduce en un promedio de calificaciones basado en los resultados obtenidos en exámenes diarios, tareas en clase y exámenes finales.

El rendimiento académico en la educación superior está influenciado por una mirada de factores personales y sociales más allá de la inteligencia y el esfuerzo (Vázquez et al., 2024a, b). Se anhela que la mayoría de los estudiantes complete sus programas de estudio dentro de los plazos

establecidos. No obstante, debido a diversos factores, hay alumnos que no logran adaptarse al ritmo del aprendizaje y se ven enfrentados a una situación de riesgo académico, como la repetición de una o más asignaturas. Esto les impide alcanzar las expectativas personales, familiares y sociales que se han establecido.

El Ministerio de Educación (MINEDU) peruano dentro de sus políticas educativas en la educación básica y en la educación superior prioriza el logro de competencias, esto con la finalidad de formar estudiantes capaces de combinar conocimientos, habilidades y destrezas, para dar solución a una problemática dada (Incio-Flores and Capuñay-Sánchez, 2023). El rendimiento académico es el encargado de medir el logro de competencias en los estudiantes, en el Perú este se mide en escala vigesimal, siendo 0 puntos la nota mínima desaprobatoria, 10.50 puntos la nota mínima aprobatoria y 20 puntos la nota máxima aprobatoria (excelencia académica).

Es fundamental que las entidades educativas posean la habilidad de anticipar con exactitud los resultados del rendimiento académico de sus estudiantes, dado que esto facilita la intervención precoz y el soporte personalizado para aquellos estudiantes con problemas, además de identificar a personas de alto rendimiento que podrían tener potencialmente beneficios de un programa de enriquecimiento (Zhang et al., 2021). Diversos factores tales como los familiares, sociales, personales, escolares, de salud, laborales, entre otros, desempeñan un papel crucial en la determinación del rendimiento académico (Verdugo-Guamán et al., 2023). El estatus socioeconómico de los padres se refiere a su posición económica y social en comparación con otros, evaluada a partir de su ocupación, nivel educativo y renta (Osei-Owusu et al., 2018). Este estatus actúa como un indicador de la situación social de una familia, influenciado por diversos aspectos como su capacidad financiera, grado de educación, posición profesional y vínculo político (Parsons et al., 2001). De manera similar, Nwigwe et al. (2021) definieron el estatus socioeconómico de los padres como la conjunción de sus ingresos, educación y ocupación.

La problemática del bajo rendimiento académico en los primeros semestres de la educación universitaria constituye un desafío significativo, particularmente en las carreras de ingeniería. Este fenómeno se refleja en una elevada tasa de reprobación en las asignaturas de los primeros periodos académicos, incrementando el riesgo de deserción o abandono por parte de los estudiantes. Una de las instituciones educativas del nivel superior universitario que presenta este problema, y el cual fue el lugar donde se realizó este estudio es la Facultad de Ingeniería de la Universidad Peruana Los Andes en Perú. El análisis de los primeros semestres académicos en la de la Escuela Profesional de Ingeniería de Sistemas y Computación de la Facultad de Ingeniería, durante los semestres académicos 2021 I, 2021 II y 2023 I; revela un alto porcentaje de asignaturas aprobadas en los niveles iniciales (100%). Sin embargo, en el tercer nivel, se observó un aumento en la tasa de desaprobación (14%) y en el cuarto nivel se mantiene en un 16%. Por lo tanto, es necesario identificar los factores asociados al bajo rendimiento académico en los primeros semestres.

Este estudio tiene como objetivo de identificar las variables significativas que inciden en el rendimiento académico de los estudiantes en los primeros semestres de estudios de la escuela profesional de ingeniería de sistemas y computación de la Universidad Peruana los Andes en Perú, utilizando regresión logística y el análisis discriminante. En este contexto se formula las siguientes hipótesis de investigación:

**H1:** Las características individuales y socioeconómicas como sexo, edad, nivel de instrucción de los padres y condición socioeconómica del entorno familiar, influyeron en el bajo rendimiento académico de los estudiantes.

**H2:** Las variables de preingreso que contribuyeron en el bajo rendimiento académico fueron, el promedio de calificaciones del nivel secundario, el tipo de institución donde cursó estudios secundarios (público o privado), la localidad donde radica la institución donde estudió la secundaria.

**H3:** Las características que tienen una relación directa con el desempeño académico previo en los primeros semestres son, el número de asignaturas aprobadas y el promedio de calificaciones obtenidas.

**H4:** Las variables significativas que inciden en el rendimiento académico de los estudiantes en los primeros semestres de estudios de la escuela profesional de ingeniería de sistemas y computación de la Universidad Peruana los Andes son las características individuales y socioeconómicas, variables de preingreso y el desempeño académico previo.

### **Method (Método Utilizado)**

Esta investigación presenta un enfoque cuantitativo, no experimental, de tipo proyectiva y predictiva (Hurtado, 2000); nos permite medir la influencia de las características individuales y socioeconómicas, así como las variables de preingreso y el desempeño académico en el rendimiento académico.

### **Population and sample (Población y muestra)**

En el presente estudio, el conjunto de elementos que conforma a la población, son los estudiantes de la Escuela Profesional de Ingeniería de Sistemas y Computación de la “Universidad Peruana Los Andes”, matriculados en el semestre 2022 I, 2021 II y 2023 I (900 estudiantes) correspondientes a la modalidad de Educación Presencial.

El tamaño de la muestra para esta investigación fue de 360 estudiantes universitarios de ingeniería; la muestra tiene las mismas características definidas en la población, y fue calculada con un nivel de confianza del 95% y un margen de error de 5%.

### **Data collection instrument (Instrumento de recolección de datos)**

La estructura del cuestionario instrumento presenta 10 ítems en total; respecto a sexo, edad, procedencia de la Institución Educativa donde cursó la Educación Secundaria (público o privado), nivel educativo más alto del padre y la madre, nivel de ingreso del padre, la madre y el estudiante, nivel de ingreso familiar, promedio de notas alcanzado durante la educación secundaria, densidad poblacional de la localidad donde se ubica la institución educativa donde cursó sus estudios secundarios, número de asignaturas matriculadas en el semestre académico y promedio ponderado de notas obtenidas durante los semestres académico. Para la validación del instrumento se utilizó el formulario de validación y la confiabilidad se determinó a través del coeficiente Alfa de Cronbach (0,954).

### **Elección del modelo predictivo**

Varios modelos predictivos son utilizados por diversas universidades para determinar el éxito académico de los estudiantes, y, por otro, identificar los factores que influyen en él. La regresión logística es un método estadístico consolidado que permite predecir variables dicotómicas.

Al considerar los pesos de las variables independientes, se puede calcular la probabilidad de que ocurra un determinado evento (Weber, 2023). Este enfoque resulta especialmente útil para pronosticar el éxito académico, ya que se centra en dos posibles resultados: "finalización exitosa de los estudios" frente a "no finalización exitosa de los estudios".

Los árboles de decisión son útiles tanto para tareas de clasificación como para pronósticos. Este algoritmo genera un diagrama en forma de árbol donde las variables independientes se encuentran en las ramas, mientras que los posibles valores se representan en dichas ramas. El inicio del proceso se basa en la variable que presenta la mayor desigualdad en su distribución, conocida como entropía condicional, en relación con la variable dependiente (Teng et al., 2025). Las otras variables se encuentran en las siguientes ramas hasta que finalmente se puede leer el resultado de la clasificación al final de las ramas; se utiliza con diferentes tipos de datos de entrada (es decir, nominales, numéricos y de texto) generalmente para predecir el fracaso escolar utilizando combinaciones de variables psicológicas y demográficas (Villarrasa-Sapiña et al., 2024).

Los bosques aleatorios son algoritmos supervisados de "aprendizaje conjunto" (ensemble learning algorithms) en base a una gran cantidad de árboles de decisión, construidos a partir de una selección aleatoria de variables (o atributos) del total de variables independientes del conjunto de datos. Cada árbol del bosque evalúa una muestra aleatoria con reemplazo del conjunto de datos (proceso conocido como bootstrapping) y luego los resultados de todos los árboles de clasificación son tomados en cuenta y usando el principio de "sabiduría de las masas" se consideran las clasificaciones más frecuentes del conjunto de árboles como la solución final (Breiman, 2004).

Máquinas de vectores de soporte (SVM) es un algoritmo dentro del área de machine learning supervisado que nos ayuda a clasificar los datos en grupos similares (categorías), gracias a que no es necesario conocer la data en detalle para utilizarlo y su habilidad de trabajar en altas dimensiones (Nti et al., 2020). Para poder identificar a que categoría pertenece cada dato, el algoritmo crea un hiperplano entre los datos en un espacio multidimensional, que separa los datos de forma óptima. El hiperplano se coloca en el espacio de modo que las distancias entre los puntos de los grupos separados sean lo más grandes posible, intentando identificar la separación entre categoría.

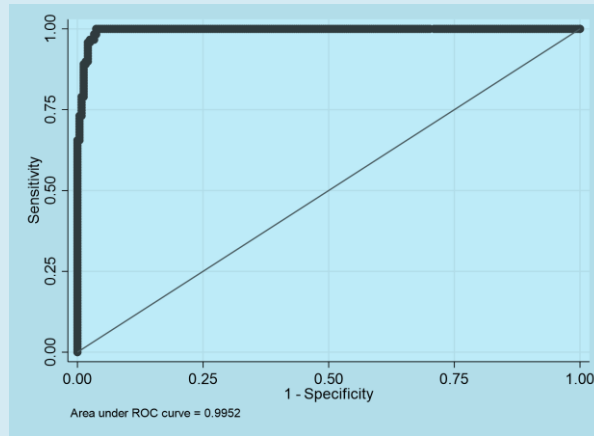
El algoritmo de Redes Neuronales también llamado redes neurales artificiales (Hosseini et al., 2021) se compone de una capa de entrada, capas de salidas y capas ocultas, éstas a su vez están constituidas por neuronas artificiales e interconectadas mediante sinapsis, cada sinapsis con su respectivo peso. Los valores de entrada se multiplican, en primer lugar, por un peso específico. Luego, se suman y se envían a la neurona, que cuenta con una función de activación que puede presentarse de diversas formas. Este procedimiento se repite hasta que los valores finalmente llegan a la capa de salida (Weber, 2023). El resultado final será un valor de 1 o 0, dependiendo de si se supera o no un umbral determinado.

Herrmann and Weigert (2024), investigaron la precisión con la que los algoritmos pueden predecir el éxito o el fracaso académico de los estudiantes (Tabla 1); identificando que la regresión logística funciona mejor, pero clasifica correctamente solo el 73 % de los casos. Los predictores más utilizados, como los bosques aleatorios o las máquinas de vectores de soporte, clasificaron correctamente solo el 70 y el 68 % de los NTS, respectivamente.

**Tabla 1.** Precisión (Accuracy) predictiva del éxito académico después del primer semestre de estudio

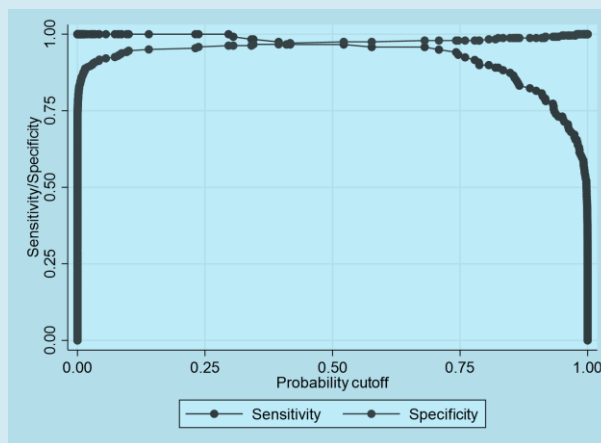
Algorithm	Accuracy all students	Accuracy non-traditional students
logistic regression	0,87	0,73
naive bayes	0,86	0,86
random forest	0,86	0,70
support vector machine	0,87	0,68
neural network	0,87	0,72

En la Figura 1, se observan los valores predictivos del para el análisis del rendimiento académico.



**Figura 1.** Nivel de predictibilidad de la información

Además, como la variable objetivo fue una variable levemente desbalanceada se obtuvo que el cutoff este alrededor de 0,5, mismo valor que se toma como referencia para la clasificación (Figura 2).



**Figura 2.** Nivel de sensibilidad y especificidad de la información

## Análisis por algoritmo

### Logistic regression (Regresión logística)

- ✓ **Accuracy all students (Todos los estudiantes):** Alta precisión (0,87), lo que indica que el modelo es confiable para la población general.
- ✓ **Accuracy non-traditional students (Estudiantes No tradicionales):** Baja precisión relativa (0.73), lo que sugiere que este algoritmo no captura bien las particularidades de este grupo

### Naïve Bayes

- ✓ **Accuracy all students (Todos los estudiantes):** Precisión ligeramente menor (0,86), pero muy cercana a la regresión logística.
- ✓ **Accuracy non-traditional students (Estudiantes No tradicionales):** Precisión igual (0,86), lo que destaca su capacidad para generalizar bien en ambos grupos. Este es el único algoritmo que mantiene la misma precisión en ambos segmentos.

### Random Forest

- ✓ **Accuracy all students (Todos los estudiantes):** Precisión de 0,86, similar a los anteriores.
- ✓ **Accuracy non-traditional students (Estudiantes No tradicionales):** Precisión baja (0,70), lo que indica que el modelo pierde efectividad con estudiantes que tienen trayectorias educativas menos convencionales.

### Support Vector Machine (SVM)

- ✓ **Accuracy all students (Todos los estudiantes):** Precisión alta (0,87).
- ✓ **Accuracy non-traditional students (Estudiantes No tradicionales):** **La más baja de todos** los algoritmos (0,68), lo que sugiere que SVM no es adecuado para este grupo.

### Neural network (Red neuronal)

- ✓ **Accuracy all students (Todos los estudiantes):** Precisión alta (0,87).
- ✓ **Accuracy non-traditional students (Estudiantes No tradicionales):** Moderada (0,72), mejor que SVM y Random Forest, pero inferior a Naïve Bayes.

## Conclusiones

- **Naïve Bayes** es el algoritmo más equilibrado, con buen desempeño en ambos grupos.
- **SVM y Random Forest** muestran limitaciones para estudiantes no tradicionales.
- **Regresión logística y redes neuronales** funcionan bien en general, pero pierden precisión en poblaciones con características más diversas.

## Results

*H1: Las características individuales y socioeconómicas como sexo, edad, nivel de instrucción de los padres y condición socioeconómica del entorno familiar, influyeron en el bajo rendimiento académico de los estudiantes.*

Las características individuales y socioeconómicas si lograron explicar de manera conjunta el desempeño académico, mostrando en la Tabla 2 un valor de probabilidad (p-valor) de 0,001 y un R2 del 11,23 % que se traduce en una mediana significancia estadística y que no son significativos como condicionantes inmediatos a un desempeño académico bajo y tienen poca incidencia para el pronóstico, con ligera excepción de las variables: sexo del estudiantes, nivel de instrucción de los padres e incluso los niveles de ingreso.

**Tabla 2.** Resultados de las características individuales y socioeconómicas

Logistic regression	Number of obs =	358
	LR chi2(24) =	51.11
	Prob > chi2 =	0.001
Log likelihood = -20208	Pseudo R2 =	0.1123

Des_acad_bajo_23_1	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
Edad	0.1516031	0.0262519	5.77	0	0.1001504	0.2030559
SEXO						
Femenino	-0.0621252	0.3224909	-0.19	0.847	-0.6941958	0.5699454
Nivedu_padre						
Primaria	0.9430885	1.067982	0.88	0.377	-1.150118	3.036295
Secundaria	0.8625178	1.025009	0.84	0.4	-1.146464	2.871499
Superior	1.179622	1.051744	1.12	0.262	-0.8817591	3.241003
Posgrado	0.8460958	1.209633	0.7	0.484	-1.524741	3.216933
Nivedu_madre						
Primaria	-0.7939552	0.8960411	-0.89	0.376	-2.550164	0.962253
Secundaria	-0.7135494	0.8839332	-0.81	0.42	-2.446027	1.018928
Superior	-0.6246684	0.9447902	-0.66	0.509	-2.476423	1.227086
Posgrado	-0.6769611	1.128807	-0.6	0.549	-2.889381	1.535459
Ing_padre						
< 0 ; 1,000 >	-0.0174588	0.4321127	-0.04	0.968	-0.8643841	0.8294664
[ 1,000 ; 3,000 >	-0.2301944	0.4152883	-0.55	0.579	-1.044145	0.5837558
[ 3,000 ; 5,000 ]	0.0276914	0.9123539	0.03	0.976	-1.760489	1.815872
< 5,000 ; + >	-2.055971	1.352988	-1.52	0.129	-4.707778	0.5958362
Ing_madre						
< 0 ; 1,000 >	0.5413113	0.3068361	1.76	0.078	-0.0600764	1.142699
[ 1,000 ; 3,000 >	0.2294851	0.3737669	0.61	0.539	-0.5030845	0.9620547
[ 3,000 ; 5,000 ]	-0.5219928	1.197296	-0.44	0.663	-2.86865	1.824664
< 5,000 ; + >	0 (empty)					
Ing_estudiante						
< 0 ; 1,000 >	0.0702836	0.2662146	0.26	0.792	-0.4514875	0.5920547
[ 1,000 ; 3,000 ]	-0.9871915	0.7527467	-1.31	0.19	-2.462548	0.488165
[ 3,000 ; 5,000 ]	0.7800921	1.31669	0.59	0.554	-1.800572	3.360757
Ing_familiar						
[ 1,000 ; 2,000 >	0.1211472	0.3865835	0.31	0.754	-0.6365426	0.8788371
[ 2,000 ; 3,000 >	-0.0505703	0.4688605	-0.11	0.914	-0.96952	0.8683794
[ 3,000 ; 5,000 ]	-0.1063511	0.5782402	-0.18	0.854	-1.239681	1.026979
< 5,000 ; + >	1.282122	1.324991	0.97	0.333	-1.314813	3.879056
_cons	-4.744114	1.275647	-3.72	0	-7.244335	-2.243893

Así mismo se demostró que la sensibilidad de los regresores fue del 35,29 %, la especificidad del 89,96% y la posibilidad de clasificación del 71,79 % (Tabla 3); y su calidad de las características individuales y socioeconómicas (Tabla 4).

**Tabla 3.** Sensibilidad de los regresores

Logistic regression	Number of obs	=	360
	LR chi2(5)	=	4.49
	Prob > chi2	=	0.4807
Log likelihood = -226.19878	Pseudo R2	=	0.0098

Des_acad_bajo_23_1	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]
Inst_proc					
Privado	0.1261307	0.2495728	0.51	0.613	-0.363023 0.6152844
Nota_Sec					
[ 12 ; 14 >	0.2808987	0.7534025	0.37	0.709	-1.195743 1.75754
[ 14 ; 16 >	0.1099243	0.7342868	0.15	0.881	-1.329251 1.5491
[ 16 ; 18 ]	-0.4103476	0.7756309	-0.53	0.597	-1.930556 1.109861
< 18 ; 20 >	-0.5040697	1.074905	-0.47	0.639	-2.610844 1.602705
_cons	-0.7917176	0.7339682	-1.08	0.281	-2.230269 0.6468336

**Tabla 1** Calidad de los regresores de las características individuales y socioeconómicas

Sensitivity	Pr( + D)	35.29%
Specificity	Pr( ~D)	89.96%
Correctly classified		71.79%

Los resultados obtenidos en la Tabla 4 evidencian que los regresores utilizados para analizar las características individuales y socioeconómicas presentan una alta capacidad para identificar correctamente a los estudiantes que no están en riesgo académico (especificidad = 89,96 %), pero una baja capacidad para detectar a aquellos que sí lo están (sensibilidad = 35,29 %). La posibilidad global de clasificación del modelo fue moderada (71,79 %), lo que indica que, si bien el modelo es eficaz para reconocer casos negativos, su utilidad para anticipar situaciones de riesgo académico es limitada. Por tanto, se recomienda complementar estos regresores con otras variables académicas o psicopedagógicas para mejorar la sensibilidad del modelo y fortalecer su capacidad predictiva.

**H2: Las variables de preingreso que contribuyeron en el bajo rendimiento académico fueron, el promedio de calificaciones del nivel secundario, el tipo de institución donde cursó estudios secundarios (público o privado), la localidad donde radica la institución donde estudió la secundaria.**

Los resultados fueron un valor de probabilidad de 0,4807 y un  $R^2$  del 0,98 %, claramente se observó en la Tabla 5 que la significancia global de esta dimensión fue mínima o casi nula, por lo que concluimos que cualquier combinación de estas variables no fueron consistentes y afirmamos que esta dimensión no contribuyó a la mejor explicabilidad del modelo, específicamente sobre la variable objetivo. Y en la Tabla 6, el nivel de predictibilidad del desempeño académico.

La dimensión evaluada no mostró significancia estadística ni capacidad explicativa relevante ( $p = 0,4807$ ;  $R^2 = 0,98 \%$ ), por lo tanto, se concluye que sus variables no aportan de manera consistente al modelo predictivo sobre la variable objetivo.

**Tabla 2** Resultados del desempeño académico previo

Logistic regression	Number of obs	=	360
	LR chi2(4)	=	403.17
	Prob > chi2	=	0
Log likelihood = -26.861978	Pseudo R2	=	0.8824

Des_acad_bajo_23_1	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
PromPond_2022_2	-22.007	4.785136	-4.6	0	-31.3857	-12.62831
PromPond_2022_1	3.619853	1.369353	2.64	0.008	0.9359713	6.303735
1.CantAsign_2023_1	-0.7894064	1.140381	-0.69	0.489	-3.024511	1.445699
1.CantAsign_2022_2	0.7357429	1.304769	0.56	0.573	-1.821557	3.293043
_cons	239.2335	49.36434	4.85	0	142.4812	335.9858

**Tabla 3** Nivel de predictibilidad del desempeño académico previo

Sensitivity	Pr( + D)	98.32%
Specificity	Pr( -~D)	98.34%
Correctly classified		98.33%

**H3: Las características que tienen una relación directa con el desempeño académico previo en los primeros semestres son, el número de asignaturas aprobadas y el promedio de calificaciones obtenidas.**

Sobre las características que tienen una relación directa con el desempeño académico previo en los primeros semestres de estudio, que constituye la tercera dimensión, se tuvo como resultados un valor de probabilidad (p-valor) de 0 y un  $R^2$  del 88,24 %, que tácitamente implica una significancia estadística aceptable (Tabla 13). Así mismo el nivel de predictibilidad de esta dimensión se resumió en una sensibilidad del 98,32 %, especificidad del 98,34 % y el nivel de clasificación correcta del 98,321 %, por lo que se trata de un modelo equilibrado. Concluimos que esta dimensión es la que mejor incide en la variable dicotómica de desempeño académico de los estuantes.

La tercera dimensión, que corresponde a las características directamente relacionadas con el desempeño académico previo en los primeros semestres, es la que mejor incide en la variable dicotómica de desempeño académico. Esto se evidencia por un p-valor de 0, un  $R^2$  de 88,24%, y

una sensibilidad del 98,32%, especificidad del 98,34% y precisión de clasificación del 98,321%, indicando un modelo equilibrado y con significancia estadística aceptable.

**H4: Las variables significativas que inciden en el rendimiento académico de los estudiantes en los primeros semestres de estudios de la escuela profesional de ingeniería de sistemas y computación de la Universidad Peruana los Andes son las características individuales y socioeconómicas, variables de preingreso y el desempeño académico previo.**

En la Tabla 7 se evidencia que todas las variables de las tres dimensiones si inciden en la significancia conjunta y es estadísticamente significativa con un valor de probabilidad (p-valor) de 0 y un R<sup>2</sup> del 89,36 %. Pero, en cuanto a la significancia individual de variables se tuvo que la institución educativa de procedencia, el nivel educativo de los padres, el nivel de ingresos, la cantidad de asignaturas matriculadas y las calificaciones de la educación secundaria no figura como representativas, por tanto, estas variables no deben consignarse en un modelo adecuado con fines predictivos o de pronóstico. El modelo presenta significancia conjunta de las tres dimensiones con p = 0 y un R<sup>2</sup> de 89,36%, lo que indica que explica una alta proporción de la variabilidad y es adecuado para pronósticos a nivel global. Sin embargo, a nivel de significancia individual, las variables institución educativa de procedencia, nivel educativo de los padres, nivel de ingresos, cantidad de asignaturas matriculadas y calificaciones de la educación secundaria no son representativas y no deberían incluirse en un modelo predictivo. En consecuencia, conviene conservar las tres dimensiones en el modelo y descartar esas variables individuales para una especificación más adecuada.

**Tabla 7. Resultados de las características individuales y socioeconómicas, variables de preingreso y el desempeño académico previo**

Logistic regression	Number of obs	=	360
	LR chi2(12)	=	408.28
	Prob > chi2	=	0
Log likelihood = -24.307369	Pseudo R2	=	0.8936

Des_acad_bajo_23_1	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
Edad	-0.1017263	0.0846675	-1.2	0.23	-0.2676715	0.0642189
Inst_proc	-0.1700657	0.8747896	-0.19	0.846	-1.884622	1.54449
Nivedu_padre	0.3206181	0.7492354	0.43	0.669	-1.147856	1.789093
Nivedu_madre	-1.064624	0.7281299	-1.46	0.144	-2.491733	0.3624842
Ing_padre	0.5860351	0.5971934	0.98	0.326	-0.5844424	1.756513
Ing_madre	0.2636197	0.5875559	0.45	0.654	-0.8879688	1.415208
Ing_estudiante	1.237965	0.6588435	1.88	0.06	-0.0533449	2.529274
Ing_familiar	-1.093524	0.5671518	-1.93	0.054	-2.205121	0.0180732
PromPond_2022_2	-19.49009	4.389868	-4.44	0	-28.09407	-10.8861
Nota_Sec	0.5768875	0.6931142	0.83	0.405	-0.7815914	1.935366
Cant_Asig_2022_2	1.136861	1.301625	0.87	0.382	-1.414277	3.687998
Cant_Asig_2023_1	-0.5814071	1.190912	-0.49	0.625	-2.915552	1.752738
_cons	255.5367	57.75493	4.42	0	142.3391	368.7343

## Discusión y conclusiones

Teniendo en cuenta los resultados del análisis de datos estadísticos, podemos concluir que tanto el estatus socioeconómico de los padres como los ingresos de los estudiantes influyen de manera significativa y positiva en el rendimiento académico de estos (Werang et al., 2024). Además, es fundamental implementar de manera exhaustiva las políticas gubernamentales relacionadas con el estatus socioeconómico de las familias y las subvenciones escolares, con el fin de fomentar la equidad educativa. Esto garantizaría que los estudiantes de entornos económicos desfavorecidos tengan igual acceso a una educación de calidad. Reconocer la importancia del estatus socioeconómico de los padres es esencial para abordar las desigualdades en el ámbito educativo y asegurar que todos los estudiantes, independientemente de su trasfondo socioeconómico, dispongan de las mismas oportunidades para alcanzar el éxito académico. Amador Ortíz et al., (2023) proponen ofrecer modalidades de estudio flexibles que se adapten a las necesidades de los estudiantes que trabajan. Mientras que la educación tradicional puede presentar desafíos para aquellos que buscan combinar trabajo y estudios, las tecnologías de la información han posibilitado el desarrollo de modelos educativos flexibles que permiten mayor independencia en el tiempo de estudio. Esta medida no solo beneficiaría a los estudiantes que trabajan, sino que también podría atraer a otros sectores de la población que, debido a sus responsabilidades laborales, encuentran difícil realizar una carrera profesional por falta de tiempo.

Se descubrió una relación negativa entre los ingresos obtenidos por horas trabajadas y el rendimiento académico en estudiantes que tienen una clara orientación hacia el trabajo, donde a mayor horas laborales semanales, aumentan las probabilidades de reprobar un mayor número de asignaturas; identificándose que los estudiantes que se dedican exclusivamente a estudiar presentan un desempeño académico superior, obteniendo promedios más altos, mientras que aquellos que trabajan y estudian tienden a reprobar más materias. Este resultado por Baert et al., (2017) es relevante tanto para académicos como para quienes diseñan políticas educativas, ya que sugiere la necesidad de disuadir a los estudiantes de anteponer su empleo a sus estudios. En tal sentido, se sugiere a los directivos de las instituciones de educación superior diseñar una bolsa de trabajo estudiantil eficiente que facilite la colocación de los estudiantes en empleos compatibles con sus estudios y preferiblemente vinculados a su área de formación.

Del análisis individual de las variables del modelo se tuvo que la institución de procedencia del estudiante, el nivel educativo de los padres, el nivel de ingreso, la cantidad de asignaturas matriculadas en el semestre académico y las notas de la educación secundaria no figuran como representativas, por tanto, su incidencia en la variable rendimiento académico es mínima. El nivel educativo de los padres, empleado en este estudio como un indicador del nivel socioeconómico, no parece estar correlacionado con el rendimiento académico de los estudiantes. Esto se alinea, hasta cierto punto, con los hallazgos relativamente débiles observados en investigaciones anteriores, como las de Marks (2017) y Rodríguez-Hernández et al. (2020), particularmente en lo que respecta a la educación superior. No obstante, es fundamental interpretar estos resultados teniendo en cuenta el notable sesgo persistente en el reclutamiento hacia la educación universitaria. Una posible explicación de esta situación podría ser que los estudiantes enfrentan dificultades para evaluar el nivel educativo de sus padres, o podría deberse a que la muestra de estudiantes no ha sido lo suficientemente representativa (Hort, 2024).

La aplicación de la regresión logística multinomial ha demostrado ser una metodología robusta para explorar la relación cuantitativa entre las variables independientes y el rendimiento académico. La matriz de confusión resultante evidencia una precisión del 73%, lo que, valida la capacidad del modelo para clasificar correctamente todas las instancias en sus respectivas categorías, reflejando su idoneidad para este tipo de análisis. La regresión es útil cuando se desea predecir valores numéricos con un alto grado de precisión, pero puede no ser tan efectiva para identificar patrones complejos en los datos mientras que los árboles de clasificación son ideales para identificar relaciones no lineales y jerárquicas entre variables (Reyes et al., 2025).

La analítica educativa y la minería de datos educativos permiten identificar factores que influyen en el rendimiento académico y predecir resultados para orientar intervenciones. En estudiantes de ingeniería, donde existen variables como horas de estudio, asistencia, antecedentes académicos y engagement, se pueden construir modelos que:

- Predicen la nota final o el rendimiento en una asignatura.
- Detectan a tiempo estudiantes en riesgo de desertar o con bajo rendimiento para aplicar intervenciones oportunas. A continuación, se proponen ejercicios prácticos en Python que ilustran estas ideas: centrado en la predicción continua de una nota final (regresión) y otro en la clasificación de riesgo (detección de rendimiento bajo).

# Capítulo 5

## Implicaciones Educativas

La verdadera utilidad de un modelo predictivo no reside solo en su capacidad para pronosticar un resultado, sino en cómo esa información se traduce en acciones concretas que benefician a la comunidad educativa. Los resultados obtenidos del modelo de regresión logística, como el analizado en el caso de estudio, tienen profundas **implicaciones educativas**. La implementación de modelos predictivos en el ámbito educativo ha cobrado gran relevancia en los últimos años, especialmente en el contexto de la formación de ingenieros. Estos modelos no solo permiten identificar patrones de rendimiento entre los estudiantes, sino que también ofrecen herramientas valiosas para mejorar la calidad educativa y optimizar los procesos de enseñanza-aprendizaje. A continuación, se presentan algunas implicaciones educativas derivadas del uso de modelos predictivos en el rendimiento académico.

**Intervenciones tempranas:** La capacidad de predecir el rendimiento académico a través de modelos estadísticos permite implementar intervenciones tempranas para aquellos estudiantes que muestren signos de bajo rendimiento. Según Gómez et al. (2019), estas intervenciones pueden incluir tutorías personalizadas o asesoramiento académico que ayuden a los estudiantes a superar sus dificultades antes de que se conviertan en problemas graves.

**Mejora en la toma de decisiones curriculares:** Los datos recogidos a través de modelos predictivos proporcionan información valiosa sobre la efectividad de los planes de estudio. Esto puede guiar a las instituciones educativas en la toma de decisiones informadas sobre la revisión curricular y la asignación de recursos, asegurando que se cumplan las necesidades de los estudiantes (Chawla & Davis, 2013).

**Fomento de la investigación educativa:** La integración de modelos predictivos en la educación también promueve una cultura de investigación continua entre los docentes. Al utilizar datos empíricos para evaluar y ajustar sus métodos de enseñanza, los educadores pueden participar activamente en la mejora continua del proceso educativo (Picciano, 2014).

**Desarrollo de competencias analíticas:** El uso de herramientas predictivas en la enseñanza de la ingeniería fomenta el desarrollo de competencias analíticas en los estudiantes. Esta habilidad es fundamental en el campo de la ingeniería, donde la capacidad de interpretar datos y hacer predicciones informadas es crucial para el éxito profesional (Santos & Plaza, 2018).

### Uso del modelo para la intervención: De la predicción a la acción

La aplicación de modelos predictivos en la educación, especialmente en el ámbito de la ingeniería, ha cobrado relevancia en los últimos años. Estos modelos permiten no solo anticipar el rendimiento académico de los estudiantes, sino también intervenir de manera efectiva para mejorar

su aprendizaje y éxito académico. En este sentido, el uso de la predicción como un paso previo a la acción se convierte en un enfoque fundamental para la intervención educativa.

La utilización de predicciones permite que las instituciones educativas respondan proactivamente a las necesidades de los estudiantes. Por ejemplo, una vez que se identifica a un grupo de estudiantes en riesgo de bajo rendimiento, se pueden implementar intervenciones específicas, tales como tutorías personalizadas, asesoramiento académico o programas de capacitación en habilidades de estudio (Nash et al., 2019).

Estas acciones buscan abordar las causas subyacentes del bajo rendimiento, optimizando así los recursos educativos disponibles. La información generada por el modelo predictivo es una poderosa herramienta para el diseño de estrategias de **intervención temprana**. En lugar de esperar a que el fracaso académico ocurra, las instituciones pueden actuar de forma proactiva. Por ejemplo, los estudiantes identificados con una alta probabilidad de bajo rendimiento pueden ser asignados a:

**Programas de tutoría personalizada:** Si el modelo señala que un promedio de bachillerato bajo es un factor de riesgo, se puede ofrecer a estos estudiantes un tutor que los ayude a fortalecer sus bases académicas en las áreas más débiles. Personalización del Aprendizaje en los modelos predictivos pueden analizar datos sobre el desempeño de los estudiantes, ayudando a identificar sus fortalezas y debilidades. Esto permite a los educadores diseñar programas personalizados que aborden las necesidades específicas de cada estudiante, favoreciendo así un aprendizaje más efectivo (Romero Alonso et al., 2020)

**Asesoramiento académico y psicológico:** Los estudiantes en riesgo pueden ser invitados a sesiones de asesoramiento para abordar variables personales no incluidas en el modelo, como la motivación o la gestión del estrés. Apoyos teóricos, aprendizaje analítico y minería de datos educativos destacan que el feedback oportuno y la personalización pedagógica pueden mejorar la retención y el rendimiento (Ifenthaler & Yau, 2019; Papamitsiou & Economou, 2014) Implicación educativa, desarrollo de guías docentes que integren análisis de datos de aprendizaje para adaptar tareas, retroalimentación y ritmos de instrucción sin perder la autonomía del profesorado.

**Planes de estudio personalizados:** Para aquellos que luchan con la carga académica, se puede diseñar un plan de estudios más flexible que les permita distribuir las asignaturas a lo largo de un período más largo, reduciendo la presión. Los modelos predictivos pueden complementar la evaluación tradicional al proporcionar indicadores tempranos de dificultad académica. No deben sustituir la evaluación sumativa, sino enriquecerla con métricas formativas que permitan intervenciones oportunas (p. ej., asesoría académica, tutorías) antes de que se produzca una caída sostenida en el rendimiento.

- ✓ **Desafíos:** la interpretabilidad de los modelos y la necesidad de evitar resultados sesgados que penalicen a determinados grupos. Se recomienda usar enfoques de explicabilidad y validar modelos por cohortes para garantizar equidad (Delen, 2010; Ifenthaler & Yau, 2019; Selwyn, 2019).
- ✓ **Ética, gobernanza de datos y consentimiento;** La recopilación y el análisis de datos de rendimiento deben realizarse bajo principios éticos claros: consentimiento informado, minimización de datos, seguridad de la información, y transparencia

sobre cómo se utilizan los datos y qué intervenciones pueden derivarse. Especificar roles y responsabilidades entre estudiantes, docentes y equipos de análisis de datos. Recomendación operativa es establecer comités de ética institucional que revisen proyectos de analítica educativa, desarrollar políticas de retención de datos y definir umbrales para intervenciones que respeten la autonomía estudiantil. Referencias sobre ética y gobernanza (Selwyn 2019; Ifenthaler & Yau, 2019). El modelo no es un veredicto; es una señal de alerta que permite a la institución ofrecer apoyo dirigido antes de que el problema se agrave.

## Limitaciones y futuras líneas de investigación

Los modelos predictivos utilizados para evaluar el rendimiento académico en estudiantes de ingeniería presentan diversas limitaciones que pueden afectar la validez y la aplicabilidad de sus resultados. En primer lugar, muchos de estos modelos dependen de datos históricos que pueden no reflejar las condiciones actuales del contexto educativo. Según García Rojas et al. (2020), "el uso de datos desactualizados puede llevar a inferencias erróneas sobre el rendimiento académico actual de los estudiantes" (p. 145). Esto es especialmente relevante en disciplinas como la ingeniería, donde los planes de estudio y las metodologías de enseñanza están en constante evolución. Además, la mayoría de los modelos se basan en variables cuantitativas, como calificaciones y asistencia, que pueden no capturar la complejidad del proceso de aprendizaje. De acuerdo con García y Rodríguez (2023), "la exclusión de variables cualitativas, como la motivación y el contexto social, limita la capacidad de los modelos para predecir con precisión el rendimiento". La integración de factores emocionales y sociales podría enriquecer el análisis, ofreciendo una visión más holística del estudiante.

Otra limitación importante es la generalización de los resultados. Muchos estudios previos se han centrado en poblaciones específicas, lo que dificulta la aplicación de sus conclusiones a otros contextos o grupos demográficos. Tal como indica Slade et al (2019), "los modelos que no consideran la diversidad de experiencias y backgrounds de los estudiantes corren el riesgo de perpetuar sesgos y desigualdades" (p. 342). Por lo tanto, es crucial desarrollar modelos que sean inclusivos y representativos de la población estudiantil en su totalidad. Estas limitaciones sugieren futuras líneas de investigación. Por ejemplo, se podrían desarrollar modelos que incorporen variables psicométricas o indicadores de interacción digital (participación en plataformas virtuales, uso de recursos en línea) para mejorar la precisión predictiva. Además, la investigación podría enfocarse en la validación de los modelos en diferentes contextos institucionales y culturales.

Si bien los modelos predictivos son valiosos, es crucial reconocer sus limitaciones. Un modelo, por más sofisticado que sea, nunca capturará la totalidad de la experiencia humana. Algunas de sus limitaciones clave son:

- **Variables no consideradas:** El modelo no incluye factores como la resiliencia del estudiante, el apoyo emocional de su familia, o la calidad de la enseñanza de un profesor específico. Estas variables, que son difíciles de cuantificar, pueden tener un impacto significativo en el rendimiento. Variables Contextuales como las diferencias en los métodos de enseñanza y los entornos académicos pueden ser determinantes en el rendimiento estudiantil. Según Kember (2009), el enfoque pedagógico y la interacción social en el aula contribuyen a la

experiencia de aprendizaje. La consideración de estas variables contextuales permitiría desarrollar modelos más robustos y aplicables en diferentes escenarios educativos.:

- **Factores tecnológicos:** En la era digital actual, el acceso y uso de tecnologías educativas pueden ser relevantes. Maza Guamán et al. (2025) sugieren que la integración de herramientas digitales en el proceso de aprendizaje mejora la interacción y compromiso del estudiante. Estudiar cómo el uso de plataformas digitales afecta el rendimiento podría abrir nuevas avenidas para la investigación. La aplicación de tecnologías avanzadas en el ámbito educativo ha transformado la manera en que se evalúan y predicen los rendimientos académicos, especialmente en disciplinas tan dinámicas como la ingeniería. En este contexto, diversos factores tecnológicos juegan un rol crucial en el desarrollo y la implementación de modelos predictivos que pueden ofrecer insights valiosos sobre el rendimiento estudiantil.

Uno de los factores más influyentes es la disponibilidad de grandes volúmenes de datos (Big Data). Las instituciones educativas recopilan una variedad de información, desde calificaciones y asistencia hasta interacciones en plataformas de aprendizaje en línea. Mediante técnicas de análisis de datos, estos conjuntos de datos masivos pueden ser procesados para extraer patrones y tendencias que ayuden a predecir el rendimiento de los estudiantes. Según Daniel (2019), "el uso de Big Data en la educación permite la creación de modelos predictivos más robustos que pueden identificar a los estudiantes en riesgo de bajo rendimiento" (p. 525).

Los algoritmos de aprendizaje automático son otro factor tecnológico significativo. Estos algoritmos pueden descubrir relaciones complejas entre variables que influyen en el rendimiento académico. Por ejemplo, un estudio realizado por Almarashdeh (2020) encontró que "los modelos de machine learning pueden mejorar sustancialmente la precisión de las predicciones sobre el rendimiento de los estudiantes, superando métodos estadísticos tradicionales" (p. 102). La capacidad de estos modelos para adaptarse y aprender de nuevos datos también les confiere una ventaja significativa en la educación.

Las plataformas de aprendizaje en línea representan una oportunidad invaluable para recoger datos en tiempo real sobre el comportamiento y desempeño de los estudiantes. Estas plataformas permiten a los profesores acceder a métricas sobre el tiempo de dedicación a las tareas, participación en foros y resultados en evaluaciones. Investigaciones recientes indican que "las plataformas online no solo facilitan la enseñanza, sino que también generan datos que son fundamentales para la implementación de modelos predictivos efectivos" (Kahraman et al., 2021, p. 413).

La inteligencia artificial juega un papel fundamental en la personalización del aprendizaje y la evaluación del rendimiento. Los sistemas basados en IA pueden identificar las áreas en las que un estudiante está luchando y ofrecer recursos específicos para mejorar su comprensión. Un estudio de Sinha et al. (2022) destaca que "la inteligencia artificial puede ofrecer recomendaciones personalizadas de

aprendizaje, lo que permite a los educadores intervenir a tiempo y mejorar el rendimiento académico" (p. 78).

- **Diversidad cultural y lingüística:** La diversidad en el contexto cultural y lingüístico de los estudiantes puede influir en su rendimiento. Como se señala en el trabajo de Gay (2010), los estudiantes de diferentes orígenes culturales pueden experimentar desigualdades en el acceso a la educación y en la dinámica del aula. Incluir esta variable puede ayudar a crear modelos predictivos que sean más inclusivos y representativos: Conceptualizar la diversidad cultural como variable contextual, no solo demográfica. el rendimiento de un estudiante no depende sólo de habilidades técnicas, sino también de contextos culturales y prácticas lingüísticas que influyen en el acceso, la persistencia y la interpretación de contenidos de ingeniería.

Los estudiantes pueden utilizar repertorios léxicos y conceptuales que atraviesan lenguas; comprender esto es clave para interpretar medidas de desempeño, lectura de enunciados técnicos y resolución de problemas

Las investigaciones señalan que las barreras socioculturales y lingüísticas pueden afectar la retención y el rendimiento, por lo que los modelos predictivos deben contemplar estos factores para no sesgar decisiones institucionales.

- **Cambios en el entorno:** El modelo se basa en datos históricos. Los cambios en el plan de estudios, las políticas universitarias o incluso la situación socioeconómica general podrían alterar la validez del modelo con el tiempo. Entornos socioeconómicos y de apoyo familiar características del entorno fuera de la universidad, como situación socioeconómica, apoyo familiar y carga de trabajo externo.

Para abordar estas limitaciones, se sugieren varias líneas futuras de investigación que podrían mejorar la efectividad de los modelos predictivos en el ámbito del rendimiento académico en estudiantes de ingeniería. Una de las propuestas más inmediatas es la incorporación de enfoques mixtos, que incluyan tanto datos cuantitativos como cualitativos. La combinación de métodos puede proporcionar una comprensión más profunda y precisa de los factores que influyen en el rendimiento académico.

Asimismo, es fundamental investigar la implementación de técnicas de aprendizaje automático e inteligencia artificial que puedan adaptar los modelos a cambios en el entorno educativo. La personalización de los modelos, podría mejorar la precisión de las predicciones al considerar las características individuales de los estudiantes, permitiendo una intervención más oportuna y efectiva.

Finalmente, se debe prestar especial atención a la evaluación del impacto de variables extrínsecas, como el apoyo familiar y las condiciones socioeconómicas, en el rendimiento académico. La literatura existente muestra que estos factores juegan un papel significativo en el éxito educativo. Según Sirin (2005), ignorar el contexto externo del estudiante limita la capacidad de los modelos para hacer predicciones precisas. Explorar esta área podría resultar en modelos más robustos y equitativos.

## Epílogo: *La Estadística al Servicio de la Educación*

- ✓ Este recorrido por los fundamentos, la metodología y la aplicación de los modelos de regresión nos ha demostrado que la **estadística predictiva** es mucho más que un conjunto de fórmulas y ecuaciones. Es una herramienta poderosa para entender la complejidad del rendimiento académico y para transformar la forma en que las instituciones educativas abordan el fracaso. Desarrollo de capacidades institucionales: construir comunidades de práctica en estadística educativa dentro de facultades de ingeniería, con capacitación continua para docentes en interpretación de modelos, manejo de datos y ética de datos. Esto fortalece la sostenibilidad de las mejoras pedagógicas basadas en Bruzzone & González (2020).
  - ✓ Limitaciones y horizonte de mejora: reconocer que los modelos son aproximaciones y que el rendimiento estudiantil depende de múltiples factores no capturados por datos históricos. La validación continua, la actualización de modelos y la incorporación de nuevas variables (entorno laboral, proyectos de ingeniería, prácticas de laboratorio) son esenciales para mantener su relevancia (Siemens & Long, 2011; Baker & Inventado, 2014).
  - ✓ Futuro disciplinar: la estadística educativa en ingeniería puede evolucionar hacia enfoques de aprendizaje adaptativo, análisis de redes de colaboración entre estudiantes y simulaciones de escenarios de ingeniería que integren incertidumbre, para apoyar decisiones pedagógicas más robustas y personalizadas (Romero & Ventura, 2010).
- Integrar la estadística como columna vertebral de la toma de decisiones pedagógicas: los modelos predictivos del rendimiento deben verse no como fines en sí mismos, sino como herramientas para identificar apoyos oportunos, adaptar itinerarios formativos y mejorar la equidad en el aprendizaje de la ingeniería. Las métricas de desempeño deben iluminar tanto los resultados como las trayectorias de aprendizaje a lo largo del tiempo.
  - Al pasar de una gestión reactiva a una proactiva, las universidades pueden no solo mejorar sus tasas de retención y graduación, sino también ofrecer una experiencia educativa más equitativa y personalizada. La estadística, al servicio de la educación, se convierte en un motor para el éxito estudiantil, asegurando que cada estudiante tenga las mejores oportunidades para alcanzar su máximo potencial, implementación institucional y cultura de datos.
  - La efectividad de modelos predictivos depende de una cultura institucional orientada a la toma de decisiones basada en datos. Esto implica formación para docentes y personal académico en interpretación de métricas, mantenimiento de pipelines de datos, y un marco claro para la evaluación de impacto de las intervenciones.
  - Dinámica de causalidad frente a correlación, aunque los modelos predictivos pueden identificar correlaciones útiles, la acción educativa debe basarse en evidencias sobre efectos causales cuando sea posible. Diseños cuasi-experimentales y pruebas de intervención pueden ayudar a separar efectos de intervención de efectos contextuales (Angrist & Pischke, 2009).

- Implicaciones para la currícula de ingeniería es la estadística que debe informar no solo al rendimiento académico, sino también al diseño de cursos, estrategias de tutoría, distribución de recursos y temporalidad de evaluaciones. El objetivo es una educación en ingeniería más predecible, justa y adaptativa (Kuh,2008; Freeman et al., 2014).

## Diseño de un programa de tutoría basado en el modelo predictivo

### Identificación de estudiantes en riesgo

El primer paso es **implementar el modelo predictivo** en el sistema de gestión académica de la universidad. Al inicio de cada semestre (o incluso antes de la inscripción final), se ingresan los datos de los nuevos estudiantes. El modelo genera una **puntuación de riesgo** o una probabilidad de bajo rendimiento para cada uno.

En ingeniería, el contexto de alta demanda académica y carga de integración de fundamentos teóricos y prácticos exige herramientas de predicción que consideren tanto variables cognitivas como de entorno de aprendizaje (laboratorios, prácticas, proyectos) y características de la malla curricular.

El propósito: reducir tasas de fracaso, deserción y sobrecarga, promoviendo apoyos enfocados (tutorías, mentorías, reforzamientos de habilidades) sin generar estigmatización

En programas rigurosos como Ingeniería, donde las tasas de deserción pueden ser elevadas, esta capacidad predictiva permite transformar el enfoque de la retención, pasando de ser una reacción a un evento a ser una estrategia proactiva y personalizada.

El proceso comienza con la recolección de datos históricos (demográficos, de bachillerato y de comportamiento en plataforma) para entrenar un algoritmo que clasifique a los estudiantes en categorías de riesgo (ej. Bajo, Medio, Alto). Los algoritmos más utilizados para esta clasificación binaria (éxito vs. fracaso) o multinomial (riesgo bajo, medio, alto)

Estos son los algoritmos más utilizados:

✓ **Regresión Logística (RL):** Es la base, ya que proporciona la probabilidad de un resultado binario y, crucialmente, ofrece interpretabilidad a través de sus coeficientes. (Hastie et al., 2017).

✓ **Máquinas de Vectores de Soporte (SVM) y Bosques Aleatorios (Random Forests):** Estos modelos son más complejos y a menudo alcanzan una mayor precisión (accuracy), especialmente cuando las relaciones entre las variables predictoras no son lineales, lo cual es frecuente en los datos educativos.

Los estudiantes que superan un umbral de riesgo predefinido (por ejemplo, aquellos con una probabilidad de bajo rendimiento superior al 60%) son automáticamente identificados.

Existen diferentes Variables Clave para la Predicción: tales como variables de ingreso, variables de comportamiento y Variables de rendimiento intermedio

## Segmentación de estudiantes para intervenciones específicas

Representa la fase más avanzada y estratégica en la aplicación de modelos predictivos en la educación en Ingeniería. No basta con identificar a un estudiante en riesgo; es fundamental comprender porque está en riesgo para aplicar el recurso educativo correcto.

Los modelos predictivos, más allá de clasificar a los estudiantes como "Aprobarán" o "Reprobarán" (clasificación binaria), deben usarse para agruparlos en categorías que reflejen la naturaleza de su desafío. Esta segmentación, a menudo lograda mediante el análisis de componentes principales (PCA) o algoritmos de clustering, permite diseñar intervenciones específicas en lugar de soluciones genéricas

### VARIABLES DE SEGMENTACIÓN

Para lograr una segmentación efectiva en el contexto de la Ingeniería, donde la base matemática y la disciplina son cruciales, debemos centrarnos en las siguientes dimensiones:

Segmento de Riesgo	Variable Dominante (Ejemplo)	Intervención Sugerida
Riesgo por Base Académica	GPA de Bachillerato bajo (< 2,5) y puntajes bajos en pruebas de ingreso de Matemáticas.	<b>Cursos de Nivelación/Tutorías Intensivas</b> en fundamentos, centradas en la brecha de conocimiento previa.
Riesgo por Compromiso/Disciplina	Baja frecuencia de acceso a la plataforma LMS y baja entrega de tareas cortas.	<b>Mentoría Personalizada</b> enfocada en gestión del tiempo y hábitos de estudio (López, 2023).
Riesgo por Dificultad Específica	Alto GPA de Bachillerato, pero baja calificación en un solo examen parcial de cálculo o física.	<b>Refuerzo Temático Hiper-Específico</b> con material adaptativo y sesiones de consulta sobre el módulo fallido.

## Algoritmos de agrupación (Clustering)

Mientras que la Regresión Logística predice el *resultado* (éxito/fracaso), los algoritmos de agrupación se usan para predecir el *tipo* de estudiante de riesgo, lo que impulsa la segmentación:

- ✓ **K-Means:** Agrupa a los estudiantes en un número predefinido ( ) de clústeres basados en la similitud de sus características (ej. alta horas de estudio + bajo GPA van a un clúster diferente que bajo horas de estudio + alto GPA).
- ✓ **Análisis Factorial:** Identifica las dimensiones subyacentes que realmente explican la varianza en el rendimiento, permitiendo segmentar a los estudiantes en función de factores como "habilidad cuantitativa" o "motivación".

Basándonos en los predictores claves identificados por el modelo, podemos segmentar a los estudiantes en riesgo para ofrecerles apoyos más específicos.

## Grupo A: Bajo Promedio de Bachillerato (y riesgo moderado/alto)

✓ **Diagnóstico:** Estos estudiantes pueden necesitar refuerzo en fundamentos matemáticos y físicos, o mejorar sus habilidades de estudio. Los datos utilizados para construir el modelo pueden estar sesgados. Si el conjunto de datos proviene de una única institución, el modelo podría no ser aplicable a otras universidades con perfiles estudiantiles diferentes:

✓ **Intervención y diferenciación educativa:** Las predicciones deben traducirse en intervenciones escalables y sostenibles. Esto puede incluir tutoría focalizada, rutas de estudio guiadas, módulos de revisión de conceptos centrales y recomendaciones de estudio personalizadas. La personalización debe balancearse con la viabilidad operativa y con la valoración de la autonomía del estudiante. Consideración práctica es diseñar intervenciones basadas en umbrales de riesgo transparentes y revisables, acompañadas de evaluaciones de efecto para ajustar las estrategias.

✓ **Tutorías Académicas Específicas:** Sesiones de repaso semanales enfocadas en las materias troncales de primer año (cálculo, física básica, química). Los tutores podrían ser estudiantes avanzados de ingeniería o docentes de apoyo:

**Problema:** es predecir, a partir de datos disponibles al ingreso o durante el primer año, qué estudiantes con bajo promedio de bachillerato presentan mayor riesgo de bajo rendimiento en ingeniería, para activar tutorías específicas y preventivas.

✓ **El Objetivo de los modelos:** es identificar a priori (o en ventanas tempranas) a estudiantes de alto riesgo y, concurrentemente, priorizar a quienes más podrían beneficiarse de tutorías académicas específicas. Esto debe hacerse con una metodología que permita interpretabilidad y acción directa en el diseño de intervenciones.

✓ **Talleres de habilidades de estudio:** Cursos intensivos sobre técnicas de toma de apuntes, gestión del tiempo, preparación de exámenes y lectura crítica. Los talleres de habilidades de estudio, combinados con modelos predictivos, pueden ser una estrategia efectiva para mejorar el rendimiento académico de estudiantes de bajo promedio en bachillerato. Al proporcionar herramientas prácticas y apoyo personalizado, se puede aumentar la probabilidad de éxito académico y preparar a estos estudiantes para los desafíos futuros en sus carreras de ingeniería.

Los talleres de habilidades de estudio son una herramienta fundamental para mejorar el rendimiento académico de estudiantes de bajo promedio, especialmente aquellos que están en riesgo moderado o alto de no completar sus estudios. Estos talleres pueden enfocarse en diversas áreas, como la gestión del tiempo, técnicas de estudio efectivas, y el desarrollo de habilidades críticas de pensamiento y resolución de problemas. Según un estudio de McGuire y McGuire (2015), la enseñanza de habilidades de estudio puede aumentar significativamente la motivación y el rendimiento académico de los estudiantes.

## Ideas para talleres

- ✓ **Taller de gestión del tiempo:** Enseñar a los estudiantes a planificar su tiempo de manera efectiva, utilizando herramientas como calendarios y aplicaciones de gestión de tareas. Esto puede ayudar a reducir la procrastinación y mejorar la productividad (Covington, 2000).
- ✓ **Técnicas de estudio activo:** Introducir métodos como el aprendizaje basado en problemas y el uso de mapas conceptuales. Estas técnicas fomentan un aprendizaje más profundo y significativo (Prince, 2004).
- ✓ **Desarrollo de habilidades de resolución de problemas:** A través de ejercicios prácticos y simulaciones, los estudiantes pueden aprender a abordar problemas complejos, una habilidad crucial en la ingeniería (Jonassen, 2000).
- ✓ **Taller de estrategias de motivación:** Incluir sesiones sobre cómo establecer metas académicas y mantener la motivación personal. La autoeficacia juega un papel importante en el rendimiento académico (Bandura, 1997).
- ✓ **Uso de tecnología en el aprendizaje:** Capacitar a los estudiantes en el uso de plataformas digitales y recursos en línea que pueden complementar su aprendizaje y facilitar el acceso a materiales educativos (Hattie, 2009).

**Mentoría: Asignación de un estudiante de años superiores o un egresado como mentor** para guiar en la adaptación a la vida universitaria. A través de relaciones de mentoría efectivas, los estudiantes pueden recibir orientación personalizada, apoyo emocional y desarrollo de habilidades que son esenciales para su éxito académico y profesional. Según un estudio de Crisp y Cruz (2009), la mentoría puede mejorar la retención y el rendimiento académico al proporcionar a los estudiantes un sentido de pertenencia y apoyo. La mentoría, combinada con modelos predictivos, puede ser una estrategia poderosa para mejorar el rendimiento académico de los estudiantes de ingeniería. Al identificar a los estudiantes en riesgo y proporcionarles un apoyo personalizado, las instituciones educativas pueden aumentar la probabilidad de éxito académico y preparar a los estudiantes para sus futuras carreras.

### Ideas para programas de mentoría

- **Asignación de mentores basada en datos:** Utilizar modelos predictivos para identificar estudiantes en riesgo y asignarles mentores que tengan experiencia en áreas específicas donde los estudiantes necesitan mejorar. Esto puede aumentar la efectividad de la mentoría (Higgins & Kram, 2001).
- **Sesiones de tutoría personalizadas:** Organizar sesiones de tutoría donde los mentores trabajen uno a uno con los estudiantes, abordando sus necesidades académicas específicas y ayudándoles a desarrollar estrategias de estudio efectivas (Baker, 2010).
- **Desarrollo de habilidades blandas:** Incluir en el programa de mentoría talleres sobre habilidades blandas, como la comunicación efectiva y el trabajo en equipo, que son esenciales para el éxito en la ingeniería.
- **Seguimiento y evaluación continua:** Implementar un sistema de seguimiento que permita a los mentores evaluar el progreso de los

estudiantes a lo largo del tiempo. Esto puede incluir reuniones regulares y el uso de herramientas de análisis de datos para ajustar las estrategias de mentoría según sea necesario.

- **Fomento de redes profesionales:** Facilitar la creación de redes profesionales mediante la conexión de estudiantes con profesionales de la industria. Esto no solo proporciona orientación, sino que también puede abrir puertas a oportunidades laborales futuras (Ragins & Scandura, 1999).

### Grupo B: Baja Asistencia a Clases (y riesgo moderado/alto)

✓ **Diagnóstico:** Estos estudiantes podrían estar enfrentando problemas de motivación, responsabilidades externas, o dificultades de acceso a la universidad. La baja asistencia a clases es un fenómeno preocupante que puede afectar significativamente el rendimiento académico de los estudiantes de ingeniería. La asistencia a clases no solo está correlacionada con el éxito académico, sino que también influye en el desarrollo de habilidades interpersonales y en la integración social dentro del entorno educativo (Saldaña Villa & Barriga, 2010). La identificación temprana de estudiantes con riesgo moderado o alto de baja asistencia es crucial para implementar intervenciones efectivas. Los modelos predictivos pueden ser herramientas valiosas para este diagnóstico.

El diagnóstico de la baja asistencia a clases, combinado con modelos predictivos, puede ser una estrategia efectiva para abordar el riesgo moderado o alto de los estudiantes de ingeniería. Al identificar a los estudiantes en riesgo y proporcionar intervenciones adecuadas, las instituciones educativas pueden mejorar la asistencia y, en consecuencia, el rendimiento académico.

Los modelos predictivos utilizan datos históricos y variables relevantes para anticipar comportamientos futuros, como la asistencia a clases. Según un estudio de Romero y Ventura (2007), la minería de datos en educación puede ayudar a identificar patrones que preceden la baja asistencia, permitiendo a las instituciones tomar medidas proactivas. Por ejemplo, factores como el rendimiento académico previo, la participación en actividades extracurriculares y el contexto socioeconómico pueden ser indicadores clave de riesgo (Kotsiantis et al., 2006).

### Estrategias de intervención

- ✓ **Análisis de datos históricos:** Implementar un sistema que recoja y analice datos sobre la asistencia de los estudiantes, correlacionándolos con su rendimiento académico y otros factores. Esto puede ayudar a identificar a los estudiantes en riesgo de baja asistencia (Baker, 2010).
- ✓ **Encuestas de clima académico:** Realizar encuestas para evaluar la percepción de los estudiantes sobre el ambiente de aprendizaje, la calidad de la enseñanza y su motivación. Esta información puede ser crucial para entender las razones detrás de la baja asistencia (Galavis & Álvarez, 2010).
- ✓ **Intervenciones personalizadas:** Una vez identificados los estudiantes en riesgo, se pueden implementar intervenciones personalizadas, como sesiones de orientación,

tutorías, o mentorías, que aborden sus necesidades específicas y fomenten su compromiso con la clase (Crisp & Cruz, 2009).

- ✓ **Programas de incentivos:** Desarrollar programas que incentiven la asistencia, como recompensas por asistencia perfecta o participación activa en clase. Esto puede motivar a los estudiantes a asistir regularmente (Hattie, 2009).
- ✓ **Monitoreo continuo:** Establecer un sistema de seguimiento que permita a los educadores monitorear la asistencia y el rendimiento académico de los estudiantes a lo largo del tiempo. Esto puede incluir reuniones regulares para discutir el progreso y ajustar las estrategias de intervención según sea necesario (Kuh, 2008)

**Seguimiento personalizado:** Un asesor académico se comunica proactivamente con estos estudiantes para entender las razones de su baja asistencia. El seguimiento personalizado para estudiantes con baja asistencia, apoyado por modelos predictivos, puede ser una estrategia efectiva para mejorar el rendimiento académico en ingeniería. Al identificar a los estudiantes en riesgo y proporcionar intervenciones adecuadas, las instituciones educativas pueden fomentar un mayor compromiso y éxito académico. La baja asistencia a clases es un problema significativo que puede afectar el rendimiento académico de los estudiantes de ingeniería. La asistencia regular es fundamental para el aprendizaje efectivo y la integración en el entorno académico. Según Saldaña Villa & Barriga (2010), la falta de asistencia puede ser un indicador de deserción y bajo rendimiento. Por lo tanto, es crucial implementar un seguimiento personalizado para aquellos estudiantes identificados con riesgo moderado o alto de baja asistencia.

Los modelos predictivos son herramientas que permiten a las instituciones educativas anticipar comportamientos de los estudiantes, como la asistencia a clases. Utilizando datos históricos, como el rendimiento académico previo, la participación en actividades extracurriculares y factores socioeconómicos, estos modelos pueden identificar a los estudiantes en riesgo (Romero & Ventura, 2007). Al aplicar un seguimiento personalizado, las instituciones pueden intervenir de manera proactiva para mejorar la asistencia y, en consecuencia, el rendimiento académico.

### **Ideas para implementar un seguimiento personalizado**

- **Monitoreo de asistencia en tiempo real:** Implementar un sistema que registre la asistencia de los estudiantes en tiempo real y genere alertas cuando se detecten patrones de baja asistencia. Esto permite a los educadores actuar rápidamente (Kuh, 2008).
- **Entrevistas individuales:** Realizar entrevistas personalizadas con estudiantes que presentan baja asistencia para comprender las razones detrás de su comportamiento. Esto puede incluir factores personales, académicos o emocionales que afectan su compromiso (Crisp & Cruz, 2009).
- **Planes de acción personalizados:** Desarrollar planes de acción específicos para cada estudiante en riesgo, que incluyan estrategias de apoyo, como tutorías,

sesiones de orientación y recursos adicionales. Estos planes deben ser revisados y ajustados regularmente (Baker, 2010).

- **Involucramiento de padres y tutores:** Fomentar la comunicación con los padres o tutores de los estudiantes para crear un entorno de apoyo en casa. Esto puede incluir reuniones periódicas para discutir el progreso y las estrategias de intervención (Hattie, 2009).
- **Uso de tecnología para el seguimiento:** Utilizar plataformas digitales que permitan a los estudiantes y educadores hacer un seguimiento de la asistencia y el rendimiento académico. Herramientas como aplicaciones de gestión del aprendizaje pueden facilitar la comunicación y el monitoreo (Kotsiantis et al., 2006).

**Apoyo psicológico o de bienestar:** Si se identifican problemas personales (estrés, ansiedad, problemas familiares), se les canaliza hacia los servicios de apoyo psicológico de la universidad. La baja asistencia a clases es un fenómeno que puede estar relacionado con diversos factores, incluyendo problemas emocionales y psicológicos. Para los estudiantes de ingeniería, donde la carga académica puede ser intensa, el bienestar psicológico es crucial para el rendimiento académico. Según Saldaña Villa & Barriga (2010), la falta de asistencia puede ser un indicador de problemas subyacentes que afectan la motivación y el compromiso del estudiante. Por lo tanto, es fundamental implementar un seguimiento personalizado que incluya apoyo psicológico para aquellos identificados con riesgo moderado o alto de baja asistencia.

## Ideas para implementar apoyo psicológico

- **Evaluaciones psicológicas iniciales:** Realizar evaluaciones psicológicas para identificar factores que puedan estar contribuyendo a la baja asistencia, como ansiedad, depresión o estrés. Esto puede ayudar a personalizar el apoyo ofrecido (Kuh, 2008).
- **Sesiones de consejería individual:** Proporcionar sesiones de consejería personalizadas donde los estudiantes puedan hablar sobre sus preocupaciones y recibir orientación sobre cómo manejar el estrés académico y emocional (Crisp & Cruz, 2009).
- **Talleres de manejo del estrés:** Organizar talleres que enseñen técnicas de manejo del estrés, como la meditación, la atención plena (mindfulness) y la gestión del tiempo. Estas habilidades pueden ayudar a los estudiantes a enfrentar mejor la presión académica (Kabat-Zinn, 2003).
- **Grupos de apoyo:** Establecer grupos de apoyo donde los estudiantes puedan compartir sus experiencias y estrategias para superar desafíos académicos y emocionales. La interacción con compañeros puede fomentar un sentido de comunidad y pertenencia (Hattie, 2009).
- **Seguimiento continuo:** Implementar un sistema de seguimiento que permita a los consejeros y educadores monitorear el progreso de los estudiantes y ajustar

el apoyo según sea necesario. Esto puede incluir reuniones regulares y el uso de herramientas de análisis de datos para evaluar la efectividad de las intervenciones (Baker, 2010).

**Flexibilización (si aplica):** En algunos casos, se podría explorar la posibilidad de ofrecer acceso a materiales de clase grabados o formatos híbridos, siempre que esto no comprometa los objetivos de aprendizaje. La flexibilidad en la asistencia a clases, apoyada por modelos predictivos, puede ser una estrategia eficaz para abordar la baja asistencia en estudiantes de ingeniería. Al identificar a los estudiantes en riesgo y ofrecer opciones adaptadas a sus necesidades, las instituciones educativas pueden mejorar la retención y el rendimiento académico, preparando a los estudiantes para un futuro exitoso en sus carreras.

Los modelos predictivos son herramientas que permiten a las instituciones educativas identificar a los estudiantes que podrían beneficiarse de una mayor flexibilidad en la asistencia. Al analizar datos como el rendimiento académico, la participación en actividades extracurriculares y factores socioeconómicos, estos modelos pueden prever quiénes son más propensos a tener problemas de asistencia (Romero & Ventura, 2007). Esto permite a los educadores implementar políticas de flexibilidad dirigidas a aquellos que más lo necesitan.

### **Ideas para implementar flexibilización**

- **Clases Híbridas o en Línea:** Ofrecer opciones de clases híbridas o completamente en línea puede proporcionar a los estudiantes la flexibilidad necesaria para asistir a las clases según su disponibilidad. Esto es especialmente útil para aquellos que trabajan o tienen responsabilidades familiares (Garrison & Vaughan, 2008).
- **Grabaciones de clases:** Proporcionar grabaciones de las clases para que los estudiantes puedan revisarlas en su propio tiempo. Esto no solo ayuda a aquellos que no pueden asistir a la clase en vivo, sino que también permite a todos los estudiantes repasar el material (Baker, 2010).
- **Políticas de asistencia flexible:** Implementar políticas que permitan a los estudiantes justificar su ausencia por razones personales o académicas, sin penalizaciones severas. Esto puede incluir la posibilidad de realizar trabajos adicionales o participar en actividades alternativas para compensar la falta de asistencia (Kuh, 2008).
- **Sesiones de recuperación:** Ofrecer sesiones de recuperación o tutorías para aquellos estudiantes que han faltado a clases, donde puedan ponerse al día con el contenido y recibir apoyo adicional (Crisp & Cruz, 2009).
- **Monitoreo y evaluación:** Utilizar modelos predictivos para monitorear la asistencia y el rendimiento académico de los estudiantes de manera continua.

Esto permitirá ajustar las políticas de flexibilidad según sea necesario y proporcionar intervenciones personalizadas (Kotsiantis et al., 2006).

**Reuniones motivacionales:** Sesiones grupales o individuales para reforzar la importancia de la presencia en clase y la participación activa. Las reuniones motivacionales, apoyadas por modelos predictivos, pueden ser una estrategia efectiva para abordar la baja asistencia a clases en estudiantes de ingeniería.

### Ideas para reuniones motivacionales

- **Charlas inspiradoras:** Invitar a exalumnos o profesionales de la industria que compartan sus experiencias y cómo superaron desafíos similares. Esto puede motivar a los estudiantes a asistir a clases y a comprometerse con su educación (Hattie, 2009).
- **Talleres de establecimiento de metas:** Realizar talleres donde los estudiantes aprendan a establecer metas académicas y personales. Este enfoque puede ayudarles a visualizar su futuro y a entender la importancia de la asistencia regular (Locke & Latham, 2002).
- **Actividades de Team Building:** Organizar actividades que fomenten la cohesión del grupo y el trabajo en equipo. Estas actividades pueden ayudar a los estudiantes a sentirse más conectados con sus compañeros y con la institución, lo que puede aumentar su motivación para asistir a clases (Galavis & Álvarez, 2010).
- **Sesiones de Retroalimentación:** Proporcionar un espacio donde los estudiantes puedan expresar sus preocupaciones y sugerencias sobre el curso y la enseñanza. La retroalimentación puede hacer que se sientan valorados y escuchados, lo que puede aumentar su compromiso (Kuh, 2008).
- **Reconocimiento y Recompensas:** Implementar un sistema de reconocimiento para estudiantes que mejoren su asistencia. Esto puede incluir certificados, menciones honoríficas o pequeñas recompensas, lo que puede incentivar a otros a asistir más regularmente (Baker, 2010).

### Grupo C: Combinación de Bajo Promedio y Baja Asistencia (Alto Riesgo)

**Diagnóstico:** Estos estudiantes presentan el mayor riesgo y probablemente requieren un **paquete de intervención integral**. La combinación de bajo promedio académico y baja asistencia a clases representa un alto riesgo para el éxito de los estudiantes en carreras de ingeniería. Esta situación puede ser indicativa de problemas subyacentes que afectan tanto el rendimiento académico como la motivación del estudiante. Según Saldaña Villa & Barriga (2010), la integración social y académica es crucial para la retención de estudiantes, y la falta de asistencia puede ser un precursor de

la deserción. Por lo tanto, es fundamental implementar un diagnóstico efectivo que identifique a estos estudiantes y permita la intervención oportuna.

Se les ofrece una combinación de las estrategias descritas para los Grupos A y B, con un seguimiento más intensivo por parte de un equipo multidisciplinario (asesor académico, tutor, consejero).

### ***ESTRATEGIAS DE INTERVENCIÓN***

- ✓ **Análisis de datos combinados:** Implementar un sistema que combine datos de rendimiento académico y asistencia para identificar a los estudiantes que presentan tanto bajo promedio como baja asistencia. Esto puede incluir el uso de herramientas de análisis de datos para visualizar patrones y tendencias (Baker, 2010).
- ✓ **Encuestas diagnósticas:** Realizar encuestas para evaluar las percepciones de los estudiantes sobre su entorno académico, sus motivaciones y las barreras que enfrentan. Esta información puede ser crucial para entender las razones detrás de su bajo rendimiento y asistencia (Galavis & Álvarez, 2010).
- ✓ **Entrevistas individuales:** Llevar a cabo entrevistas con los estudiantes identificados en riesgo para explorar sus experiencias y desafíos personales. Esto puede ayudar a personalizar las intervenciones y proporcionar el apoyo necesario (Crisp & Cruz, 2009).
- ✓ **Programas de apoyo académico:** Desarrollar programas de tutoría y apoyo académico dirigidos a estudiantes con bajo promedio y baja asistencia. Estos programas pueden incluir sesiones de estudio, mentoría y talleres de habilidades de estudio (McGuire & McGuire, 2015).
- ✓ **Monitoreo y evaluación continua:** Establecer un sistema de seguimiento que permita a los educadores monitorear el progreso de los estudiantes a lo largo del tiempo. Esto puede incluir reuniones regulares y el uso de herramientas de análisis de datos para ajustar las estrategias de intervención según sea necesario (Kuh, 2008).

### **Seguimiento y Evaluación Continua**

El seguimiento y la evaluación continua se refiere al proceso sistemático de recolectar, analizar y utilizar información sobre el desempeño de los estudiantes a lo largo del tiempo. Según Stiggins (2005), este enfoque es fundamental para informar prácticas pedagógicas y mejorar los resultados educativos. En el contexto de modelos predictivos, implica la recolección regular de datos académicos y conductuales que alimentan los algoritmos de predicción. El programa de tutoría no termina con la intervención inicial. Es crucial realizar un **seguimiento continuo** de los estudiantes que participan en él. El seguimiento y la evaluación continua son componentes clave que permiten la identificación temprana de estudiantes con riesgo de bajo rendimiento y facilitan la intervención oportuna. En este documento, se presentan ideas claras sobre cómo estas prácticas pueden ser aplicadas específicamente en el desarrollo y la aplicación de modelos predictivos.

**Monitoreo de indicadores:** Se siguen de cerca sus calificaciones, su asistencia a tutorías y a clases, y su participación en talleres. El monitoreo de indicadores es el proceso continuo de recolección, análisis y visualización de métricas relevantes para evaluar y calibrar modelos predictivos del rendimiento académico en estudiantes de ingeniería, con el objetivo de mejorar intervenciones pedagógicas, asignación de recursos y soporte individualizado.

En ingeniería, donde las trayectorias de aprendizaje pueden ser complejas y heterogéneas, el monitoreo debe ser iterativo: medir, interpretar, ajustar el modelo y comunicar resultados a docentes y responsables educativos.

#### Indicadores clave a monitorear

- ✓ **Rendimiento académico:** puntuaciones en asignaturas críticas (cálculo, física, termodinámica), promedio ponderado, tendencias de calificaciones por semestre.
- ✓ **Retención y progresión:** tasas de avance PAC (prematuras abandonos), tiempos de graduación, retrabajos requeridos.
- ✓ **Comportamiento de estudio:** horas de estudio registradas, uso de plataformas de aprendizaje, participación en actividades prácticas, asistencia a laboratorios.
- ✓ **Indicadores de riesgo:** predicción de bajo rendimiento con umbrales de alerta, sesgos de desigualdad por género, origen socioeconómico o grupo de ingeniería.
- ✓ **Calidad de intervención:** efectividad de tutorías, intervención temprana, cambios en diseño instruccional tras la detección de prioridades.
- ✓ **Confiablez y sesgo del modelo:** calibración, sesgo poblacional, estabilidad temporal (tanto de los predictores como de las predicciones).

#### Arquitectura de monitoreo de indicadores

- ✓ **Fuente de datos:** sistemas de gestión académica (SIA), plataformas LMS, sistemas de control de asistencia, registros de laboratorio, encuestas de satisfacción y clima.
- ✓ **Modelo predictivo:** modelos supervisados para predicción de rendimiento (regresión, clasificación), modelos de series temporales para tendencias, enfoques de aprendizaje automático para detección de patrones complejos.
- ✓ **Medición de desempeño del modelo:** exactitud, AUC-ROC, MAE, RMSE; calibration plots; validación cruzada; evaluación de sesgo y fairness.
- ✓ **Visualización y tablero (dashboard):** indicadores clave en tiempo real o casi real, alertas automatizadas, gráficos de tendencias y cohortes.
- ✓ **Ciclo de mejora continua:** reentrenamiento periódico, revisión de características, ajuste de umbrales de alerta, y validación con datos recientes.

#### Proceso recomendado de monitoreo (pasos prácticos)

- **Paso 1:** Definir objetivos y métricas alineadas con el programa de ingeniería (p. ej., reducción de tasa de fallo en cursos clave).
- **Paso 2:** Identificar indicadores principales y secundarios; mapear fuentes de datos y responsables.

- **Paso 3:** Construir una línea base del modelo y de las métricas de rendimiento, estableciendo límites de alerta razonables.
- **Paso 4:** Implementar un tablero de mando con vistas por cohorte (carrera, año, grupo de ingreso) y por indicador.
- **Paso 5:** Establecer umbrales de alerta y protocolos de respuesta (intervenciones docentes, asesorías, ajustes curriculares).
- **Paso 6:** Reentrenar y recalibrar el modelo periódicamente (p. ej., semestral o anual) con datos recientes.
- **Paso 7:** Evaluar efectos de las intervenciones a través de métricas de resultado y de proceso (control pre/post, análisis de cohortes).
- **Paso 8:** Garantizar ética y transparencia: sesgos, privacidad de datos, consentimiento, y comunicación responsable a estudiantes y docentes.

**Retroalimentación:** Se recopila feedback tanto de los estudiantes como de los tutores para ajustar y mejorar el programa. Para que los modelos predictivos sean efectivos, es esencial contar con datos precisos y actualizados. Esto incluye no solo las calificaciones de los estudiantes, sino también variables como asistencia, participación en clase y actividades extracurriculares (Baker & Inventado, 2014). La combinación de estos datos permite la creación de un perfil integral del estudiante, lo cual es crucial para la predicción de su rendimiento futuro. Una vez implementados los modelos, el seguimiento continuo permite identificar estudiantes que podrían beneficiarse de intervenciones específicas. Por ejemplo, si un modelo predictivo señala que un estudiante tiene altas probabilidades de fracaso en una materia particular, se puede implementar un programa de tutoría personalizado (Siemens, 2013). La retroalimentación inmediata se convierte así en un mecanismo crítico para mejorar el rendimiento.

**Re-evaluación del riesgo:** Periódicamente, se puede re-ejecutar el modelo predictivo con los datos actualizados de los estudiantes para ver si su nivel de riesgo ha disminuido.

**Medición del impacto:** Se compara el rendimiento y la tasa de retención de los estudiantes que participaron en el programa con un grupo de control similar (si es éticamente posible y metodológicamente viable) para evaluar la efectividad del programa. Finalmente, es esencial evaluar la efectividad de los modelos predictivos y su impacto en el rendimiento estudiantil. La evaluación continua del sistema debe incluir métricas de éxito, tales como tasas de retención y mejora en las calificaciones promedio (Hwang et al., 2017). Estos resultados no solo validan el modelo, sino que también ofrecen oportunidades de mejora para los algoritmos y las estrategias pedagógicas empleadas.

Este enfoque iterativo, donde el modelo predictivo informa la intervención y la evaluación del programa retroalimenta al modelo, permite una **mejora continua** y asegura que los recursos de la universidad se utilicen de la manera más efectiva para apoyar el éxito estudiantil.

En conclusión, la adopción de modelos predictivos en el ámbito educativo tiene importantes implicaciones que pueden transformar la experiencia de aprendizaje de los estudiantes de ingeniería. Estos modelos ofrecen una perspectiva valiosa que puede optimizar los procesos educativos, fomentar la personalización del aprendizaje e informar la toma de decisiones a nivel curricular. La integración de estas prácticas representa un avance significativo hacia una educación más efectiva y centrada en el estudiante.



# Consideraciones *f*inales

## La inversión estratégica en el talento ingenieril

La adopción de modelos predictivos en la educación en ingeniería representa una de las inversiones estratégicas más lucrativas y éticas que una institución puede realizar. Esto va más allá de un simple ejercicio estadístico; es la aplicación del rigor ingenieril para optimizar el recurso más valioso: el talento estudiantil.

## Retorno de la Inversión (ROI) de la analítica predictiva

El verdadero **valor de la predicción** se manifiesta en tres esferas de retorno cuantificable:

- **Optimización de Recursos (Eficiencia):** Al pasar de la reacción tardía a la intervención temprana y focalizada, evitamos el costoso "efecto dominó" del fracaso académico. En lugar de financiar soluciones genéricas, dirigimos recursos limitados (tutorías, asesoramiento, nivelación) exactamente a los estudiantes que los necesitan y por la razón precisa (ej., baja base académica vs. falta de disciplina), maximizando el impacto por cada dólar invertido.
- **Retención y Excelencia (Rentabilidad Académica):** Los modelos, especialmente la **Regresión Logística**, no solo identifican el riesgo de deserción, sino que nos dicen qué factores son más poderosos para el éxito. Al actuar sobre estos predictores claves (como el promedio de bachillerato o la asistencia), aumentamos directamente las tasas de retención y graduación, garantizando que el talento formado en nuestras aulas llegue al sector productivo.
- **Innovación Curricular y Gobernanza Ética (Valor Sostenible):** Los resultados del modelo se convierten en un mecanismo de *feedback* para la institución. Informan la revisión curricular con evidencia (*¿Funciona el pre-requisito?*) y obligan a establecer marcos de gobernanza de datos y ética. Esto no solo mejora la calidad de la enseñanza, sino que posiciona a la institución como líder en el uso responsable y transparente de la tecnología educativa.

En última instancia, integrar la **Estadística Predictiva** transforma el desafío del bajo rendimiento en una oportunidad de diseño. Armados con la capacidad de prever el futuro, nuestras facultades garantizan la máxima eficiencia, equidad y excelencia, formando a los ingenieros que la economía digital demanda, con el menor índice de falla.

# Referencias

- Agresti, A. (2018). *An introduction to categorical data analysis* (3.<sup>a</sup> ed.). John Wiley & Sons.
- Almarashdeh, I. (2020). Predicting student performance using data mining techniques: A case study of Saudi universities. *International Journal of Advanced Computer Science and Applications*, 11(5), 372–378. <https://doi.org/10.14569/IJACSA.2020.0110549>
- Álvarez-Yaulema, M., Ponce-Valle, M., Alban-Alcivar, J., & Zambrano-Salazar, L. (2024). Revisión de modelos estadísticos para pronosticar el desempeño académico en estudiantes universitarios. *MQRInvestigar*, 8(2), 3806–3823. <https://doi.org/10.56048/MQR20225.8.2.2024.3806-3823>
- Alyahyan, E., & Dustegor, D. (2020). Predicting academic success in higher education: Literature review and best practices. *International Journal of Educational Technology in Higher Education*, 17(3), 1–21. <https://doi.org/10.1186/s41239-020-0177-7>
- Amador Ortíz, C. M., Velarde Peña, L., Torres Isirdia, M. L., Rodríguez Palomera, J., & Sánchez Navarrete, Á. (2023). Empleo en estudiantes universitarios de un instituto tecnológico y su efecto en el desempeño académico. *RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 14(27).
- Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press. <https://doi.org/10.1515/9781400829828>
- Arias Ospina, S., Arce Guerrero, S., & González Jiménez, D. A. (2024). Transformando el aprendizaje en primer año universitario de ingeniería una estrategia innovadora. *Encuentro Internacional de Educación en Ingeniería*, 1–10. <https://doi.org/10.26507/paper.4048>
- Aricop, C., Rico Páez, A., & Gaytán Ramírez, N. D. (2022). Modelos predictivos del rendimiento académico a partir de características de estudiantes de ingeniería. *IE Revista de Investigación Educativa de la REDIECH*, 13, e1426. [https://doi.org/10.33010/ie\\_rie\\_rediech.v13i0.1426](https://doi.org/10.33010/ie_rie_rediech.v13i0.1426)
- Atencia Oliva, D. J., Plaza Gómez, M. T., & Hernández Riaño, H. E. (2020). Resiliencia, burnout y fracaso académico en estudiantes de Ingeniería de la Universidad de Córdoba, Colombia. *Revista Espacios*, 41(11), 23.
- Baert, S., Marx, I., Neyt, B., Van Belle, E., & Van Casteren, J. (2017). Student employment and academic performance: An empirical exploration of the primary orientation theory. *Applied Economics Letters*, 25(8), 547–552.

- Baker, R. S. J. D., & Inventado, P. S. (2014). Educational data mining and learning analytics. En J. A. Larusson & B. White (Eds.), *Learning analytics* (pp. 61–75). Springer. [https://doi.org/10.1007/978-1-4614-3305-7\\_6](https://doi.org/10.1007/978-1-4614-3305-7_6)
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman and Company.
- Blanco Torres, Y., Fragozo Alvarez, L. D. J., & Gómez Lesport, M. E. (2024). Inteligencia Artificial: Posibilidades, límites y desafíos en la educación superior. *Revista de Ciencias Sociales*, 30(4), 178–187. <https://doi.org/10.31876/rcs.v30i4>
- Breiman, L. (2004). *Consistency for a simple model of random forests*. Department of Statistics, University of California at Berkeley.
- Bruzzone, S., & González, M. (2020). Comunidades de práctica y formación docente en educación superior: Una estrategia para el desarrollo profesional. *Revista Iberoamericana de Educación Superior*, 11(31), 45–60. <https://doi.org/10.22201/iisue.24486167e.2020.31.582>
- Camones Gonzales, F. C., Sihuay Fernandez, M. T., Nolberto Sifuentes, V. A., & Padilla Caballero, J. E. A. (2024). Minería de datos: Un enfoque perspectivo desde el contexto educativo. *Revista Tribunal*, 4(9), 138–158. <https://doi.org/10.59659/revistatribunal.v4i9.70>
- Carlin, J. B., & Moreno-Betancur, M. (2025). On the uses and abuses of regression models: A call for reform of statistical practice and teaching. *Statistics in Medicine*, 44(13-14). <https://doi.org/10.1002/sim.10244>
- Castrillón, O. D., Sarache, W., & Ruiz-Herrera, S. (2020). Predicción del rendimiento académico por medio de técnicas de inteligencia artificial. *Formación Universitaria*, 13(1), 93–102. <http://dx.doi.org/10.4067/S0718-50062020000100093>
- Castro-Méndez, N. P., & Suárez-Cretton, X. A. (2024). La resiliencia en estudiantes universitarios no tradicionales. *Revista Colombiana de Educación*, (91), 33–55. <https://doi.org/10.17227/rce.num91-16430>
- Chacha, N., Mtebe, J. S., & Kalinga, E. (2023). Predicting students' academic performance using machine learning techniques: A systematic literature review. *Education and Information Technologies*, 28, 1123–1147. <https://doi.org/10.1007/s10639-022-11239-6>
- Chaves, E., Castillo, M., & Gamboa, R. (2008). Correlación entre el examen de admisión y el rendimiento en el primer año de la carrera Enseñanza de la Matemática en la UNA. *Revista Electrónica Educare*, 12(2), 65–80. <https://doi.org/10.15359/ree.12-2.7>
- Chawla, D., & Davis, L. (2013). Mining educational data to improve students' performance: A step towards quality education. En *Advanced computing, networking and informatics* (pp. 13–22). Springer. [https://doi.org/10.1007/978-3-642-32137-1\\_13](https://doi.org/10.1007/978-3-642-32137-1_13)
- Chinnakum, W., Intapan, C., Singvejsakul, J., Wongsirikajorn, M., Thongkaw, B., Eakkapun, P., & Chaiboonsri, C. (2024). The socio-economic impact of

- university in Thailand: Evidence from Chiang Mai University. *Economies*, 12(12), 339. <https://doi.org/10.3390/economies12120339>
- Chinnakum, W., Phumchusri, N., & Chinnakum, P. (2024). Education and sustainable development: A global perspective. *Journal of Educational Planning and Administration*, 38(1), 12–29.
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8.<sup>a</sup> ed.). Routledge. <https://doi.org/10.4324/9781315456539>
- Contreras, L. E., Fuentes, H. J., & Rodríguez, J. I. (2020). Predicción del rendimiento académico como indicador de éxito/fracaso de los estudiantes de ingeniería, mediante aprendizaje automático. *Formación Universitaria*, 13(5), 89–98. <https://doi.org/10.4067/S0718-50062020000500233>
- Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology*, 51(1), 171–200. <https://doi.org/10.1146/annurev.psych.51.1.171>
- Crisp, G., & Cruz, I. (2009). Mentoring college students: A critical review of the literature between 1990 and 2007. *Research in Higher Education*, 50(6), 525–545. <https://doi.org/10.1007/s11162-009-9130-2>
- Daniel, B. (2019). Big data and learning analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, 50(5), 2035–2048. <https://doi.org/10.1111/bjet.12750>
- Del Carpio-Mendoza, J. (2024). Predicción del rendimiento académico utilizando modelos de aprendizaje automático. *593 Digital Publisher*, 9(3), 444–456. <https://doi.org/10.33386/593dp.2024.3.2797>
- Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, 49(4), 498–506. <https://doi.org/10.1016/j.dss.2010.06.003>
- Díaz, C., Almeralla, R., García, M., & Hernández, P. (2021). Rendimiento académico de estudiantes en educación superior. *Dialnet*.
- Escribano Hervis, E. (2018). El desempeño del docente como factor asociado a la calidad educativa en América Latina. *Revista Educación*, 42(2), 1–25. <https://doi.org/10.15517/revedu.v42i2.27033>
- Félix Tipian, L. E., Muñoz Félix, A. P., Flores Arriola, A. L., & Cárdenas Valverde, J. C. (2025). Monitoreo y acompañamiento académico de estudiantes de educación superior utilizando tecnologías emergentes. *e-Revista Multidisciplinaria del Saber*, 3(1), 216–234. <https://doi.org/10.61286/e-rms.v3i.194>
- Field, A. P. (2018). *Descubriendo la estadística con IBM SPSS Statistics* (5.<sup>a</sup> ed.). Sage.
- Figueroa Guerra, D. A., García Narváez, P. A., & Carranco Madrid, S. P. (2025). Relación entre el nivel socioeconómico, entorno familiar, apoyo social y el rendimiento académico de estudiantes universitarios. *Revista InveCom*, 5(2). <https://doi.org/10.5281/zenodo.13737678>

- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410–8415. <https://doi.org/10.1073/pnas.1319030111>
- Galavis, S., & Álvarez, G. (2010). La encuesta de opinión estudiantil: un sistema de información para la evaluación por competencias de la actividad docente. *Revista de la Facultad de Ingeniería Universidad Central de Venezuela*, 25(3), 47-56. [http://ve.scielo.org/scielo.php?script=sci\\_arttext&pid=S0798-40652010000300006&lng=es&tlng=esc](http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0798-40652010000300006&lng=es&tlng=esc)
- García Rojas, M., López, J., & Hernández, P. (2020). Actualización de modelos predictivos en entornos educativos dinámicos. *Revista Iberoamericana de Tecnología Educativa*, 15(2), 135–150.
- García, A. (2023). Modelos de regresión en estadística predictiva. *Revista de Estadística Aplicada*, 28(3), 210–225.
- García, J. A., & Rodríguez, M. E. (2023). Relación entre diseño instruccional y rendimiento académico en un curso de Cálculo y Álgebra Lineal. *Revista Electrónica de Investigación Educativa*, 25(1), 1–20. <https://doi.org/10.32870/reie.v25i1.5513>
- Garrison, D. R., & Vaughan, N. D. (2008). *Blended Learning in Higher Education: Framework, Principles, and Guidelines*. Jossey-Bass.
- Gay, G. (2010). *Culturally responsive teaching: Theory, research, and practice*. Teachers College Press.
- Gómez, J., Martínez, A., & Rodríguez, M. (2019). Predicción del rendimiento académico mediante modelos estadísticos en estudiantes universitarios. *Revista Electrónica de Investigación Educativa*, 21, e01. <https://doi.org/10.24320/redie.2019.21.e01.1652>
- González, G., & Rodríguez, M. (2023). Factores que inciden en el rendimiento académico de estudiantes de ingeniería: Un enfoque multivariado. *Revista Educación y Desarrollo*, 17(2), 45–62.
- Gutiérrez-Monsalve, J. A., Garzón, J., & Segura-Cardona, A. M. (2021). Factores asociados al rendimiento académico en estudiantes universitarios. *Formación Universitaria*, 14(1), 13–24. <http://dx.doi.org/10.4067/S0718-50062021000100013>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8.ª ed.). Cengage Learning.
- Harrell, F. E., Jr. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis* (2.ª ed.). Springer.
- Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The elements of statistical learning: Data mining, inference, and prediction* (2.ª ed.). Springer. <https://doi.org/10.1007/978-0-387-84858-7>

- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Herrmann, L., & Weigert, J. (2024). AI-based prediction of academic success: Support for many, disadvantage for some? *Computers and Education: Artificial Intelligence*, 7, 100303. <https://doi.org/10.1016/j.caeai.2024.100303>
- Hidalgo Zambrano, J. A., & Márquez Tejena, D. M. (2025). Los factores socioeconómicos en la deserción escolar de los estudiantes de la Unidad Educativa Eloy Velázquez Cevallos. *Ciencia Latina Revista Científica Multidisciplinar*, 9(1), 9242–9254. [https://doi.org/10.37811/cl\\_rcm.v9i1.16548](https://doi.org/10.37811/cl_rcm.v9i1.16548)
- Higgins, M. C., & Kram, K. E. (2001). Reconceptualizing mentoring at work: A developmental network perspective. *Academy of Management Review*, 26(2), 264–288. <https://doi.org/10.5465/amr.2001.4378023>
- Hort, S. (2024). Studieförframgång inom högre utbildning: En utforskande undersökning av sambanden mellan högskolestudenter studieresultat och ålder, kön, språkbakgrund samt socioekonomisk bakgrund. *Högre Utbildning*, 14(3), 12–27.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3.<sup>a</sup> ed.). Wiley. <https://doi.org/10.1002/9781118548387>
- Hosseini, M.-P., Hosseini, A., & Ahi, K. (2021). A review on machine learning for EEG signal processing in bioengineering. *IEEE Reviews in Biomedical Engineering*, 14, 204–218. <https://doi.org/10.1109/RBME.2020.2969915>
- Hurtado, J. (2000). *Metodología de la investigación holística* (3.<sup>a</sup> ed.). Fundación Sygal.
- Hwang, G.-J., Tsai, C.-H., & Yang, S.-Y. (2017). An analysis of learning behavior and performance in a cloud-based learning environment: A learning analytics approach. *Computers & Education*, 111, 1–10. <https://doi.org/10.1016/j.compedu.2017.04.001>
- Ifenthaler, D., & Yau, J. (2019). Utilizing learning analytics for study success: Reflections on current empirical findings. En D. Ifenthaler (Ed.), *Utilizing learning analytics to support study success* (pp. 1–17). Springer. [https://doi.org/10.1007/978-3-030-35328-4\\_1](https://doi.org/10.1007/978-3-030-35328-4_1)
- Incio-Flores, F. A., & Capuñay-Sánchez, D. L. (2023). Factores endógenos y exógenos para el modelado y predicción del rendimiento académico de estudiantes universitarios. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 26(2), 233–247.
- James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications in Python*. Springer.
- Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, 48(4), 63–85. <https://doi.org/10.1007/BF02300500>
- Jones, L., Barnett, A., & Vagenas, D. (2025). Linear regression reporting practices for health researchers, a cross-sectional meta-research study. *PLOS ONE*, 20(3), Artículo e0305150. <https://doi.org/10.1371/journal.pone.0305150>

- Kahraman, H., & Karataş, S. (2021). Learning analytics in online education: A systematic review. *Education and Information Technologies*, 26(6), 7331–7358. <https://doi.org/10.1007/s10639-021-10579-2>
- Kember, D. (2009). *Understanding and teaching the Chinese learner: Resolving the paradox of high achievement and high stress*. Springer.
- Kleinbaum, D. G., Kupper, L. L., Nizam, A., & Rosenberg, E. S. (2013). *Applied regression analysis and other multivariable methods* (5.<sup>a</sup> ed.). Cengage Learning.
- Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. En C. H. Papadimitriou (Ed.), *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)* (pp. 43:1–43:23). Schloss Dagstuhl. <https://doi.org/10.4230/LIPIcs.ITCS.2017.43>
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (pp. 1137–1145). Morgan Kaufmann.
- Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, 160, 3–24.
- Kuh, G. D. (2008). *High-impact educational practices: What they are, who has access to them, and why they matter*. Association of American Colleges and Universities.
- Kuhn, M., & Johnson, K. (2019). *Feature engineering and selection: A practical approach for predictive models*. CRC Press. <https://doi.org/10.1201/9781315108230>
- Kumar, R., Singh, A., & Sharma, P. (2021). Academic performance evaluation in higher education: A competency-based approach. *International Journal of Educational Research*, 105, 101712. <https://doi.org/10.1016/j.ijer.2021.101712>
- Kumar, S., Agarwal, M., & Agarwal, N. (2021). Defining and measuring academic performance of HEI students: A critical review. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(6), 3091–3105.
- Lledó, G. L., Lledó, A. L., Carreres, A. L., & Cabrera, E. A. (2024). Utilidad percibida de la robótica en el currículum de Educación Primaria para el alumnado con Necesidades Específicas de Apoyo Educativo. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 27(2), 111–122.
- Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. *The American Psychologist*, 57(9), 705–717. <https://doi.org/10.1037//0003-066x.57.9.705>
- López, J. A. F. (2021). La motivación docente para obtener calidad educativa en instituciones de educación superior. *Revista Virtual Universidad Católica del Norte*, (64), 151–179. <https://doi.org/10.35575/rvucn.n64a7>
- Marks, G. N. (2017). Is SES really that important for educational outcomes in Australia? A review and some recent evidence. *Australian Educational Researcher*, 44, 191–211. <https://doi.org/10.1007/s13384-016-0219-2>

- Martínez Pérez, J. R., Pérez Leyva, E. H., Ferrás Fernández, Y., & Bermúdez Cordoví, L. L. (2021). Análisis predictivo de la deserción estudiantil en la carrera de Medicina mediante regresión logística. *EDUMECENTRO*, 13(3), 217–230.
- Maza Guamán, M. P., Pizarro Duran, T. J., Piedra Tito, P. F., Llivisaca Llivicura, C. R., Guachizaca Uyaguari, J. M., & Camacho Castillo, B. R. (2025). Impacto de las tecnologías digitales en el rendimiento académico. *Revista InveCom*, 5(2). <https://revistainvecom.org/index.php/invecom/article/view/2841>
- McGuire, S. Y., & McGuire, S. (2015). *Teach students how to learn: Strategies you can incorporate into any course to improve student metacognition, study skills, and motivation*. Stylus Publishing.
- Medina, J., Pinzón, K., & Salazar-Méndez, Y. (2021a). Determinantes del rendimiento académico de los estudiantes de una universidad pública ecuatoriana. *Revista Politécnica*, 47(2), 53–62.
- Medina, J., Pinzón, K., & Salazar-Méndez, Y. (2021b). La educación superior como inversión pública: Implicancias para el desarrollo económico. *Revista Latinoamericana de Políticas Públicas*, 13(3), 88–104.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5.ª ed.). Wiley.
- Moreno, J. E., & Chiecher, A. C. (2018). Trayectorias estudiantiles en ingeniería: Factores asociados con el logro y el retraso en la cursada. En M. E. Panaia (Ed.), *Profesión e innovación en un contexto flexible* (pp. 49–80). Miño y Dávila.
- Nash, S., Baggott, L., & Winter, R. (2019). *Data and learning analytics: Applications in higher education*. Routledge. <https://doi.org/10.4324/9780429244646>
- Navarro Roldán, C. P., & Zamudio Sisa, L. E. (2021). Cuestionario de riesgo de deserción universitaria (CDUe) basado en el modelo ecológico. *Tesis Psicológica*, 16(1), 1–20. <https://doi.org/10.37511/tesis.v16n1a12>
- Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear statistical models* (4.ª ed.). WCB McGraw-Hill.
- Nti, E. K., Wongnaa, C. A., Edusah, N. S. E., & Bakang, J. E. A. (2020). Assessment of the sustainability of community-managed water supply services in Ghana. *Environment, Development and Sustainability*, 22(7), 7097–7120. <https://doi.org/10.1007/s10668-019-00506-1>
- Nti, I. K., Quarcoo, J. A., Aning, J., & Fosu, G. K. (2022). A mini-review of machine learning in big data analytics: Applications, challenges, and prospects. *Big Data Mining and Analytics*, 5(2), 81–97. <https://doi.org/10.26599/BDMA.2021.9020028>
- Nwigwe, J. C., Ogbuanya, P. C., Ezemba, E. U., & Anyakaorah, C. C. (2021). Influence of parents' socioeconomic status on their children academic performance in secondary schools in Izzi Local Government area of Ebonyi State. *International Research Journal of Human Resource and Social Science*, 8(7), 18–27.

- Osei-Owusu, B., Twum Ampofo, E., Oteng Akyina, K., Ampomah, R., & Osei-Owusu, E. (2018). Socio-economic status of parents and its effects on the academic performance of students of Yamfo Anglican Senior High School in the Brong Ahafo region of Ghana. *Advances in Social Sciences Research Journal*, 5(4). <https://doi.org/10.14738/assrj.54.4355>
- Papamitsiou, Z. K., & Economides, A. A. (2013). Towards the alignment of computer-based assessment outcome with learning goals: The LAERS architecture. *2013 IEEE Conference on e-Learning, e-Management and e-Services*.
- Parsons, R. D., Hinson, S. L., & Sardo-Brown, D. (2001). *Educational psychology: A practitioner-researcher model of teaching*. Wadsworth/Thomson Learning.
- Pedraza Rejas, L., Acosta García, K., & Rodríguez Cisterna, C. (2025). Acciones pedagógicas que promueven el pensamiento crítico en el futuro profesorado de ciencias. *Revista Eureka Sobre Enseñanza y Divulgación de las Ciencias*, 22(3). [https://doi.org/10.25267/rev\\_eureka\\_ensen\\_divulg\\_cienc.2025.v22.i3.3604](https://doi.org/10.25267/rev_eureka_ensen_divulg_cienc.2025.v22.i3.3604)
- Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96(1), 3–14. <https://doi.org/10.1080/00220670209598786>
- Pérez, J. L. (2023). Modelos predictivos en educación superior: Variables influyentes y contexto socioeconómico para anticipar el desempeño estudiantil. *Revista de Ingeniería y Educación*, 12(1), 45–62.
- Pérez, J. M., & Salamanca, S. R. (2013). Influencia de las estrategias pedagógicas en los procesos de aprendizaje de los estudiantes de una institución de básica primaria de la ciudad de Bucaramanga. *Puente Revista Científica*, 7(2).
- Picciano, A. G. (2014). *Big data and learning analytics in higher education: Current theory and practice*. Routledge. <https://doi.org/10.4324/9781315889802>
- Prince, M. J. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, 93(3), 223–231. <https://doi.org/10.1002/j.2168-9830.2004.tb00809.x>
- Ragins, B. R., & Scandura, T. A. (1999). Burden or blessing? Expected costs and benefits of mentoring in the workplace. *Journal of Organizational Behavior*, 20(4), 493–509.
- Reyes, J. E. A., Peñafiel, E. F. M., Barrionuevo, T. P. M., & Castelo, G. V. (2025). Estudio del rendimiento académico mediante la comparación de modelos de regresión y árboles de clasificación. *Telos: Revista de Estudios Interdisciplinarios en Ciencias Sociales*, 27(1), 94–115.
- Reyes-González, N., Meneses-Báez, A. L., & Díaz-Mujica, A. (2022). Planificación y gestión del tiempo académico de estudiantes universitarios. *Formación Universitaria*, 15(1), 57–66. <https://doi.org/10.4067/S0718-50062022000100057>
- Rico Páez, A. (2023). Modelos de predicción del rendimiento académico con técnica de apilamiento. En *Función y sentido de la investigación en las instituciones de educación superior*. Instituto Politécnico Nacional. <https://orcid.org/0000-0002-6450-318X>

- Rodríguez-Hernández, C. F., Cascallar, E., & Kyndt, E. (2020). Socio-economic status and academic performance in higher education: A systematic review. *Educational Research Review*, 29, 100305. <https://doi.org/10.1016/j.edurev.2019.100305>
- Romero Alonso, R., Araya Carvajal, K., & Reyes Acevedo, N. (2025). Rol de la Inteligencia Artificial en la personalización de la educación a distancia: Una revisión sistemática. *RIED - Revista Iberoamericana de Educación a Distancia*, 28(1). <https://doi.org/10.5944/ried.28.1.41538>
- Romero, C., & Ventura, S. (2010). *Handbook of educational data mining*. CRC Press.
- Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1), 368–384. <https://doi.org/10.1016/j.compedu.2007.05.007>
- Saldaña Villa, M., & Barriga, O. A. (2010). Adaptación del modelo de deserción universitaria de Tinto a la Universidad Católica de la Santísima Concepción, Chile. *Revista de Ciencias Sociales (Ve)*, 16(4), 616–628.
- Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.
- Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5), 30–40.
- Sindhu Priya, K. (2021). Linear regression algorithm in machine learning through MATLAB. *International Journal for Research in Applied Science and Engineering Technology*, 9(12), 989–995. <https://doi.org/10.22214/ijraset.2021.39410>
- Sinha, T., & Cassell, J. (2022). Personalized learning through AI: Opportunities and challenges in educational contexts. *Journal of Educational Technology & Society*, 25(1), 65–80.
- Siqueiros-Quintana, M. G., & Vera-Noriega, J. A. (2022). Cultura organizacional en instituciones de educación superior: Conceptualización, medidas y variables asociadas. *Revista Iberoamericana de Educación Superior*, 13(36), 181–198. <https://doi.org/10.22201/iissue.20072872e.2022.36.1190>
- Slade, S., Prinsloo, P., & Khalil, M. (2019). Learning analytics: Ethical issues and dilemmas. En C. Lang, G. Siemens, & A. Wise (Eds.), *Handbook of learning analytics* (pp. 17–28). Society for Learning Analytics Research. <https://doi.org/10.18608/hla17.017>
- Smith, J., Doe, A., & Roe, B. (2020). Impacto de la retroalimentación pedagógica en línea en el rendimiento académico: un estudio comparativo. *Journal of Educational Psychology*, 15, 123–136.
- Solis Ventura, J. C., Castro Alay, M. G., & Fienco Calderón, N. A. (2025). Factores socioeconómicos que limitan el acceso educativo en zonas rurales de América Latina. *RECIMUNDO*, 9(2), 959–969. [https://doi.org/10.26820/recimundo/9.\(2\).abril.2025.959-969](https://doi.org/10.26820/recimundo/9.(2).abril.2025.959-969)

- Steyerberg, E. W. (2019). *Clinical prediction models: A practical approach to development, validation, and updating* (2.<sup>a</sup> ed.). Springer.
- Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, 87, 324–328.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6.<sup>a</sup> ed.). Pearson Education.
- Teng, Y., Zhang, H., Zhan, Y., Liu, X., Chang, Z., Chen, Y., & Liu, H. (2025). Dynamic correlation analysis of sectoral electricity consumption and urban carbon concentration using machine learning models. *Scientific Reports*, 15(1), 9939. <https://doi.org/10.1038/s41598-025-94201-z>
- Thai Jo. (2024). *Aprovechando los datos para predicciones más inteligentes*. Simple Science.
- Torres Cardona, L. (2025, 23 de septiembre). *Rendimiento académico y su relación con los problemas emocionales y conductuales en el aula*. ISEP.
- Torres Lopez, C. A., Pachas Huaytan, J. V., & Ortiz Fernandez, J. H. (2025). Análítica predictiva mediante regresión logística para la intervención temprana en ingeniería. *e-Revista Multidisciplinaria del Saber*, 3, e-RMS10112025. <https://doi.org/10.61286/e-rms.v3i.299>
- Valero, A., & Van Reenen, J. (2019). The economic impact of universities: Evidence from across the globe. *Economics of Education Review*, 68, 53–67. <https://doi.org/10.1016/j.econedurev.2018.09.005>
- Vázquez, A., Alba, B., Lois, D., García-Ael, C., & Bustillos, A. (2024a). The role of subjective socioeconomic status in predicting academic performance: Exploring the mediating influence of sense of belonging among students in a distance learning university. *The Spanish Journal of Psychology*, 27, e9. <https://doi.org/10.1017/SJP.2024.9>
- Vázquez, M., Torres, L., & Ríos, C. (2024b). Factores psicosociales que inciden en el rendimiento académico universitario. *Revista Iberoamericana de Psicología Educativa*, 17(1), 101–120.
- Verdugo-Guamán, M. E., Cabrera-Ortiz, F. P., Cabrera-Tenecela, H. P., & Escudero-Durán, M. L. (2023). Factores que inciden en el rendimiento académico en los primeros años de los estudiantes de la Universidad de Cuenca, Ecuador. *Revista Andina de Educación*, 6(2), 1–20. <https://doi.org/10.32719/26312816.2022.6.2.10>
- Villarrasa-Sapiña, I., García-Massó, X., Liébana, E., & Monfort Torres, G. (2024). Predicción del rendimiento académico en educación secundaria mediante el análisis de árboles de decisión. *Educación XX1*, 27(1), 253–279. <https://doi.org/10.5944/educxx1.33351>
- Villegas, B. R., & Núñez Lira, L. A. (2024). Factores asociados a la deserción estudiantil en el ámbito universitario: Una revisión sistemática 2018–2023. *RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 14(28), e1923. <https://doi.org/10.23913/ride.v14i28.1923>

- Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2012). *Probabilidad y estadística para ingeniería y ciencias* (9.<sup>a</sup> ed.). Pearson Educación.
- Weber, F. (2023). *Artificial intelligence for business analytics: Algorithms, platforms and application scenarios*. Springer Nature.
- Werang, B. R., Agung, A. A. G., Sri, A. A. P., Leba, S. M. R., & Jim, E. L. (2024). Parental socioeconomic status, school physical facilities availability, and students' academic performance. *Edelweiss Applied Science and Technology*, 8(5), 1–15.
- Zapata, M., Rodríguez, J., & Torres, L. (2021). Resiliencia y autoeficacia académica como predictores del rendimiento en estudiantes de ingeniería. *Revista de Psicología Educativa*, 27(1), 45–60. <https://doi.org/10.1016/j.psicoeu.2020.10.004>
- Zhang, Y., Yun, Y., An, R., Cui, J., Dai, H., & Shang, X. (2021). Educational data mining techniques for student performance prediction: Method review and comparison analysis. *Frontiers in Psychology*, 12, 698490. <https://doi.org/10.3389/fpsyg.2021.698490>

# Glosario

- **Estadística Predictiva:** Rama de la estadística que utiliza modelos para pronosticar resultados futuros basándose en datos históricos.
- **Modelo de Regresión Logística:** Un modelo estadístico utilizado para predecir la probabilidad de un resultado binario (por ejemplo, éxito/fracaso, deserción/permanencia) a partir de una o más variables predictoras.
- **Variable Dependiente:** También conocida como variable de resultado. Es la variable que se intenta predecir o explicar en un modelo.
- **Variable Independiente:** También conocida como variable predictora o explicativa. Es la variable que se utiliza para predecir la variable dependiente.
- **Odds Ratio (OR):** Medida de la asociación entre una exposición y un resultado. En la regresión logística, un OR mayor que 1 indica que la exposición aumenta las *odds* del resultado.
- **Significancia Estadística (valor  $p$ ):** La probabilidad de obtener un resultado tan extremo o más que el observado, asumiendo que la hipótesis nula (no hay relación) es verdadera. Un valor  $p$  bajo (generalmente  $< 0.05$ ) sugiere que el resultado no es por azar.
- **Hosmer-Lemeshow:** Una prueba de bondad de ajuste utilizada para evaluar si las probabilidades predichas por un modelo de regresión logística coinciden con las probabilidades observadas.
- **Análisis Bivariado:** El estudio de la relación entre dos variables.
- **Análisis Multivariado:** El estudio de la relación entre tres o más variables.
- **Deserción Estudiantil:** El abandono de los estudios universitarios antes de obtener la titulación.
- **GPA:** Promedio de notas de bachillerato

# Conjunto de Datos de Muestra

Para facilitar la replicación de los análisis presentados en este capítulo, se proporciona un conjunto de datos ficticio en formato CSV. Los lectores pueden descargar este archivo y seguir los pasos detallados en la sección de "Aplicación Práctica".

**Nombre del Archivo:** datos\_rendimiento\_ingenieria.csv

## Variables incluidas:

- ID\_Estudiante: Identificador único del estudiante.
- BajoRendimiento: Variable binaria (1 = sí, 0 = no).
- Promedio\_Bachillerato: Promedio de calificaciones del estudiante en la escuela secundaria.
- Genero: Variable categórica (0 = Femenino, 1 = Masculino).
- Horas\_Estudio\_Semanal: Horas de estudio reportadas.
- Zona\_Origen: Variable categórica (0 = Urbana, 1 = Rural).
- Asistencia\_Clases\_Porcentaje: Porcentaje de asistencia.

# Código de Software

Se ofrecen los siguientes scripts detallados para realizar los análisis en los entornos de programación estadística más comunes.

## *Código en R*

R

```
# Instalar y cargar librerías necesarias
```

```
install.packages("dplyr")
```

```
install.packages("readr")
```

```
install.packages("ggplot2")
```

```
library(dplyr)
```

```
library(readr)
```

```
library(ggplot2)
```

```
# Cargar el conjunto de datos
```

```
datos <- read_csv("datos_rendimiento_ingenieria.csv")
```

```
# Realizar análisis bivariado (ejemplo: promedio de bachillerato)
```

```
ggplot(datos, aes(x = Promedio_Bachillerato, fill = factor(BajoRendimiento))) +
```

```
  geom_density(alpha = 0.5) +
```

```
  labs(title = "Densidad del Promedio de Bachillerato por Rendimiento",
```

```
        x = "Promedio de Bachillerato", y = "Densidad")
```

```
# Construir el modelo de regresión logística
```

```
modelo_logistico <- glm(BajoRendimiento ~ Promedio_Bachillerato + Genero +
```

```
Horas_Estudio_Semanal + Zona_Origen + Asistencia_Clases_Porcentaje,
```

```
  data = datos, family = "binomial")
```

```
# Resumen del modelo
```

```
summary(modelo_logistico)
```

```
# Calcular Odds Ratios
```

```
exp(coef(modelo_logistico))
```

## Código en Python

Python

```
import pandas as pd
import statsmodels.api as sm
import matplotlib.pyplot as plt
import seaborn as sns

# Cargar el conjunto de datos
datos = pd.read_csv("datos_rendimiento_ingenieria.csv")

# Realizar análisis bivariado (ejemplo: asistencia a clases)
sns.boxplot(x='BajoRendimiento', y='Asistencia_Clases_Porcentaje', data=datos)
plt.title('Asistencia por Rendimiento')
plt.show()

# Definir variables dependiente e independientes
y = datos['BajoRendimiento']
X = datos[['Promedio_Bachillerato', 'Genero', 'Horas_Estudio_Semanal', 'Zona_Origen',
'Asistencia_Clases_Porcentaje']]
X = sm.add_constant(X) # Agregar intercepto

# Construir el modelo de regresión logística
modelo_logistico = sm.Logit(y, X).fit()

# Resumen del modelo
print(modelo_logistico.summary())

# Calcular Odds Ratios
odds_ratios = pd.DataFrame(
    {
        'OR': modelo_logistico.params,
        'P_Valor': modelo_logistico.pvalues
    }
)
print(odds_ratios)
```



Sello  
Editorial CITSA

ISBN: 978-980-8050-19-6



9 789808 050196