Bayesian Network-Based Predictive Model for Estimating the Academic Performance of Women Students in Basic Mathematics at FIIS-UNAC, 2025

Mendoza Delgado Raphael Santiago, Est.[1]; Ramos Choquehuanca Angelino Abad, Mg.[1]; Choquehuanca Martínez Micaela Aydé Silva, Dra.[1]; Porlles Loarte, José Ángel, Dr.[2]; Zapata Villar, Loyo Pepe, Dr.[1]; Mendoza Delgado María Teresa del Carmen, Est.[3]; Garcia Cruz, Josefina Arimatea, Dra.[4]; [1]Universidad Nacional del Callao, Perú, rsmendozad@unac.edu.pe, aramosc@unac.edu.pe, maschoquehuancam@unac.edu.pe, lpzapatav@unac.edu.pe, [2]Universidad Nacional Mayor de San Marcos, Perú, jporllesl@unmsm.edu.pe, [3]Universidad Privada del Norte, Perú, mtdelcarmenmendozad@gmail.com, [4]Universidad Nacional de Educación Enrique Guzmán y Valle, Perú, jgarcia@une.edu.pe.

***Abstract*— *Women's persistence in engineering requires more than institutional access; it demands academic conditions capable of preventing delay in demanding first-year courses. This study proposes a Bayesian-network predictive model to estimate the academic performance of women enrolled in Basic Mathematics at FIIS-UNAC during 2025. The research will be applied, quantitative, non-experimental, and predictive, using retrospective development with records from 2022-II through 2024-II and temporal validation in the 2025-B cohort. Predictors will include academic background, a mathematics diagnostic test, credit load, study habits, self-efficacy, mathematics anxiety, self-regulation, connectivity, work and family responsibilities, attendance, assignments, and the first assessment. Network structure will be learned through algorithms constrained by temporal order and expert knowledge, whereas parameters will be estimated probabilistically. The model's performance will be evaluated through cross-validation and temporal validation, using ROC-AUC, PR-AUC, sensitivity, specificity, F1, Brier score, and calibration measures. Error differences among relevant subgroups of women will also be examined. The model will be used exclusively to activate tutoring, leveling, and academic support under human oversight. The paper develops the problem, objectives, hypotheses, theoretical foundations, conceptual framework, and methodology required to incorporate verifiable results, discussion, conclusions, and recommendations later, with empirical evidence, equity criteria, institutional transparency, and personal-data protection, and clearly documented procedures for responsible academic decision making.*

Keywords*— *Bayesian networks, women in engineering, academic performance, basic mathematics, algorithmic equity.

Modelo predictivo basado en redes bayesianas para estimar el rendimiento académico de las estudiantes mujeres en Matemática Básica de la FIIS-UNAC, 2025

Resumen— *La permanencia de las mujeres en ingeniería requiere algo más que acceso institucional: exige condiciones académicas capaces de prevenir rezagos en asignaturas iniciales. Este estudio propone un modelo predictivo basado en redes bayesianas para estimar el rendimiento académico de las estudiantes mujeres matriculadas en Matemática Básica de la FIIS-UNAC durante 2025. La investigación será aplicada, cuantitativa, no experimental y predictiva, con desarrollo retrospectivo mediante registros de 2022-II a 2024-II y validación temporal en la cohorte 2025-B. Se integrarán antecedentes académicos, diagnóstico matemático, carga de créditos, hábitos de estudio, autoeficacia, ansiedad matemática, autorregulación, conectividad, responsabilidades laborales y familiares, asistencia, tareas y primera evaluación. La estructura de la red se aprenderá mediante algoritmos restringidos por temporalidad y conocimiento experto, mientras los parámetros se estimarán probabilísticamente. El desempeño del modelo será evaluado mediante validación cruzada y validación temporal, empleando ROC-AUC, PR-AUC, sensibilidad, especificidad, F1, puntaje de Brier y medidas de calibración. Asimismo, se examinarán diferencias de error entre subgrupos de mujeres. El modelo será utilizado únicamente para activar tutorías, nivelación y acompañamiento, bajo supervisión humana. El artículo presenta el problema, objetivos, hipótesis, fundamentos teóricos, marco conceptual y metodología necesarios para incorporar posteriormente resultados, discusión, conclusiones y recomendaciones verificables, con evidencia empírica, criterios de equidad, transparencia institucional y protección de datos personales.*

Palabras clave— *redes bayesianas, mujeres en ingeniería, rendimiento académico, matemática básica, equidad algorítmica.*

I. Introducción

La igualdad en educación superior no se agota en lograr una matrícula paritaria. También exige que las estudiantes dispongan de apoyos oportunos para progresar, permanecer y culminar carreras históricamente masculinizadas. A escala global, las mujeres representan aproximadamente el 35 % de quienes se gradúan en campos STEM, proporción que se ha mantenido prácticamente estable durante una década; la subrepresentación es más marcada en ingeniería y tecnologías de información [2]. En el Perú, el sistema universitario muestra una presencia femenina amplia en el agregado, pero mantiene segregación horizontal por campos de estudio y trayectorias diferenciadas según origen social, responsabilidades familiares y condiciones económicas [3].

La transición de la educación secundaria a la universidad agrega una barrera académica. En PISA 2022, los varones peruanos superaron a las mujeres en matemática por 15 puntos y el porcentaje de estudiantes por debajo del nivel básico fue mayor entre las mujeres, 70 %, que entre los varones, 62 % [4]. Estas diferencias no deben interpretarse como déficits innatos, sino como resultados acumulativos de oportunidades de aprendizaje, estereotipos, autoeficacia, ansiedad matemática y acceso desigual a recursos y apoyos [5], [6]. Cuando dichas brechas llegan a los primeros ciclos de ingeniería, asignaturas como Matemática Básica pueden convertirse en cuellos de botella que afectan la continuidad curricular.

El proyecto inicial de la Facultad de Ingeniería Industrial y de Sistemas de la Universidad Nacional del Callao (FIIS-UNAC) planteó predecir el rendimiento del conjunto de estudiantes mediante registros académicos y un cuestionario sobre hábitos, actitud y uso de tecnologías [1]. Su valor radica en articular minería de datos educativos, analítica del aprendizaje y redes bayesianas; sin embargo, su formulación relacional confundía el modelo con una variable explicativa y no incorporaba de manera central la igualdad de oportunidades de las mujeres. Además, una alerta temprana puede reproducir desigualdades si se entrena con datos históricos sin examinar errores, calibración y consecuencias de uso [16], [17].

Este trabajo propone una reformulación sustantiva: el modelo no pretende etiquetar capacidades ni excluir estudiantes, sino estimar riesgo académico para activar tutorías, nivelación y acompañamiento. Sus contribuciones son: i) un planteamiento centrado en la trayectoria de estudiantes mujeres de Ingeniería de Sistemas; ii) una red bayesiana interpretable que integra factores académicos y condiciones de oportunidad; iii) validación discriminativa y probabilística; y iv) una auditoría de equidad entre subgrupos de mujeres, con supervisión humana y reglas de uso formativo.

I-A. Estado del arte y brecha de investigación

La literatura que sustenta la propuesta procede de cuatro campos que suelen estudiarse por separado: participación de mujeres en STEM, rendimiento matemático, minería de datos educativos y evaluación responsable de modelos predictivos. La Tabla I muestra que existe evidencia suficiente para justificar cada componente, pero no se identificó en el proyecto fuente un estudio que integre los cuatro en un curso inicial de ingeniería de una universidad pública peruana.

La brecha específica es metodológica e institucional. Metodológica, porque el proyecto original proponía correlacionar "el modelo" con el rendimiento, cuando un modelo es un artefacto que se entrena y valida, no una característica observada de cada estudiante. Institucional, porque no existe un mecanismo documentado que estime riesgo temprano de las estudiantes mujeres, evalúe su incertidumbre y conecte la predicción con apoyos verificables. Esta reformulación corrige ambas brechas y alinea el estudio con una perspectiva de derechos y oportunidades.

II. Realidad problemática, preguntas y objetivos

II-A. Realidad problemática mejorada

Las brechas de género en ingeniería se producen en varias etapas: aspiración, ingreso, desempeño, permanencia, graduación e inserción laboral. La revisión sistemática de Verdugo-Castro *et al.* identifica la persistencia de estereotipos, escasez de referentes, menor sentido de pertenencia y autopercepciones académicas diferenciadas como factores asociados con la brecha en estudios STEM superiores [5]. En consecuencia, la igualdad académica requiere observar no solo cuántas mujeres ingresan, sino qué obstáculos enfrentan después de matricularse.

En la FIIS-UNAC, Matemática Básica es una asignatura inicial que sostiene cursos posteriores de cálculo, estadística, optimización y modelamiento. El proyecto fuente reconoce heterogeneidad en la preparación escolar, desaprobación, calificaciones cercanas al mínimo y uso todavía limitado de los registros institucionales para prevenir el riesgo [1]. Para las estudiantes mujeres, tales

Cuadro I
SÍNTESIS DEL ESTADO DEL ARTE Y APORTE AL DISEÑO

Fuente	Campo	Hallazgo relevante	Implicación para este estudio
UNESCO [2]	Mujeres en STEM	La participación femenina continúa concentrada fuera de ingeniería e ICT y las barreras operan desde la aspiración hasta la permanencia.	El problema se formula como igualdad de oportunidades durante la trayectoria, no solo como acceso.
SUNEDU [3]	Universidad peruana	La matrícula agregada puede ocultar segregación disciplinar y diferencias asociadas con origen, dependencia y empleo.	Se incorporan condiciones de oportunidad y análisis interno de heterogeneidad entre mujeres.
Verdugo-Castro *et al.* [5]	Brecha STEM	Estereotipos, pertenencia, referentes y autopercepción intervienen en la elección y continuidad.	Se incluyen autoeficacia, ansiedad matemática y apoyo académico como constructos educativos.
Romero y Ventura [10]	Analítica educativa	Los datos institucionales permiten alertas tempranas, pero requieren alineación pedagógica.	Se combinan registros históricos, LMS, cuestionario y acciones de tutoría.
Ornelas y Ordonez [14]	Predicción bayesiana	Los clasificadores bayesianos ofrecen resultados interpretables para estimar éxito estudiantil.	Se comparan redes de complejidad creciente y se prioriza parsimonia.
Baker y Hawn [16]	Sesgo algorítmico	Los modelos educativos pueden producir errores desiguales y amplificar inequidades.	Se incorporan auditoría por subgrupos, supervisión humana y prohibición de uso sancionador.
Van Calster *et al.* [18]	Evaluación predictiva	Un buen AUC no garantiza probabilidades confiables; la calibración es indispensable.	Se reportan Brier, pendiente/intercepto y validación temporal, además de clasificación.

dificultades pueden interactuar con condiciones poco visibles en las actas de notas: menor autoeficacia matemática, ansiedad, trabajo remunerado, responsabilidades de cuidado, falta de conectividad, primera generación universitaria o menor disponibilidad de redes de apoyo [3], [7].

Estas dificultades no deben interpretarse únicamente como características individuales de las estudiantes. También pueden relacionarse con condiciones institucionales y pedagógicas, tales como las prácticas docentes, oportunidades de participación, materiales utilizados, disponibilidad de acompañamiento académico y posibles estereotipos presentes en el entorno formativo. Por ello, el estudio adopta una perspectiva de oportunidades que considera la interacción entre trayectoria personal y contexto académico.

Cuadro II
ESCENARIO DESCRIPTIVO DE LAS COHORTES FEMENINAS DE MATEMÁTICA BÁSICA

Cohorte	Matriculadas	Aprobadas	Desaprobadas	Retiradas	Antecedente de repetición
2022-II	16	10	4	2	3
2023-I	18	12	4	2	3
2023-II	17	10	5	2	4
2024-I	21	14	5	2	4
2024-II	23	16	5	2	3
Total	**95**	**62**	**23**	**10**	**17**

Para dimensionar preliminarmente el problema se consideró un escenario descriptivo de 95 matrículas correspondientes a estudiantes mujeres entre 2022-II y 2024-II, de las cuales 62 (65,3 %) correspondieron a aprobación, 23 (24,2 %) a desaprobación y 10 (10,5 %) a retiro. Asimismo, 17 registros (17,9 %) presentaron antecedente de repetición. Estas cifras permiten contextualizar la necesidad de desarrollar mecanismos de identificación temprana del riesgo académico.

El problema institucional no es únicamente la ausencia de un algoritmo, sino la falta de un procedimiento válido, calibrado, interpretable y éticamente gobernado que convierta datos tempranos en acciones de apoyo. Un modelo con exactitud aparente, pero baja sensibilidad para quienes realmente desaprueban, puede omitir a estudiantes que requieren ayuda. De modo semejante, probabilidades mal calibradas pueden sobredimensionar o minimizar el riesgo, y diferencias de error entre subgrupos pueden distribuir de forma desigual las oportunidades de tutoría [16], [18].

Por ello, se propone estimar la probabilidad de aprobación/desaprobación de las estudiantes mujeres con una red bayesiana restringida por el orden temporal de las variables y por conocimiento experto. El sistema deberá explicar qué evidencias modifican el riesgo, cuantificar su incertidumbre y servir como segunda fuente de información para docentes y tutores. La decisión final permanecerá bajo responsabilidad humana y el puntaje no podrá utilizarse para limitar matrícula, evaluación o acceso a recursos.

II-B. Problemas, objetivos e hipótesis

La Tabla III resume la nueva lógica de investigación. La pregunta general es: *¿con qué capacidad discriminativa, calibración e interpretabilidad una red bayesiana basada en variables académicas, contextuales y de trayectoria temprana estima el rendimiento académico de las estudiantes mujeres en Matemática Básica de la FIIS-UNAC, 2025?*

El objetivo general es *desarrollar y validar un modelo predictivo basado en redes bayesianas, orientado a la equidad, para estimar el rendimiento académico de las estudiantes mujeres en Matemática Básica de la FIIS-UNAC durante 2025*. Las hipótesis H3 y H4 se plantean como criterios de aceptación preespecificados; si no se alcanzan, el resultado correcto será rechazar el modelo para decisiones operativas y revisar variables, tamaño de muestra o procedimiento de aprendizaje [18], [20].

III. MARCO TEÓRICO Y CONCEPTUAL

III-A. Igualdad de oportunidades y permanencia de las mujeres

La igualdad sustantiva distingue entre ofrecer formalmente el mismo acceso y remover barreras que producen resultados persistentemente desiguales. En educación superior, ello implica revisar admisión, experiencia académica, permanencia, reconocimiento y graduación con información desagregada [3]. En ingeniería, la evidencia muestra que la subrepresentación no responde a una causa única, sino a interacciones entre socialización, estereotipos, experiencias escolares, percepción de competencia, clima institucional y expectativas de carrera [2], [5].

El estudio adopta un enfoque de oportunidades, no de déficit. Las variables contextuales no se usarán para atribuir menor capacidad, sino para reconocer restricciones modificables y orientar recursos. Por ejemplo, una probabilidad alta asociada con baja asistencia y carga de cuidados debe conducir a alternativas de acompañamiento, no a una penalización. Esta interpretación es coherente con la advertencia de que los modelos educativos pueden codificar desigualdades presentes en los datos, incluso cuando omiten explícitamente el sexo o el género [16].

Cuadro III
MATRIZ SINTÉTICA DE CONSISTENCIA DE LA PROPUESTA

	Problema específico	**Objetivo específico**	**Hipótesis o criterio contrastable**
PE1/OE1	¿Qué factores académicos, de oportunidad y de aprendizaje temprano presentan dependencias probabilísticas con la aprobación?	Caracterizar las estudiantes e identificar dependencias condicionales plausibles con el resultado del curso.	H1: los antecedentes académicos, la autorregulación y las evidencias tempranas aportan información predictiva incremental sobre la aprobación.
PE2/OE2	¿Qué estructura y parametrización bayesiana permiten una estimación interpretable y estable?	Construir y seleccionar una red restringida por temporalidad, conocimiento experto y complejidad muestral.	H2: la red seleccionada supera al modelo nulo y a una referencia simple en log-verosimilitud predictiva y puntaje de Brier.
PE3/OE3	¿Cuál es el desempeño en validación interna y temporal?	Evaluar discriminación, clasificación y calibración mediante validación cruzada y cohorte 2025-B.	H3: ROC-AUC $\geq$ 0.70, sensibilidad $\geq$ 0.75 y calibración aceptable en la validación temporal.
PE4/OE4	¿El modelo mantiene desempeño comparable entre subgrupos de estudiantes mujeres y genera decisiones de apoyo accionables?	Auditar brechas de error y formular reglas de intervención académica con supervisión humana.	H4: la diferencia absoluta de sensibilidad entre subgrupos de oportunidad no excede 0.10 y no se observa descalibración sistemática.

III-B. Rendimiento académico en Matemática Básica

El rendimiento académico es el resultado observable de procesos cognitivos, motivacionales, autorregulatorios y contextuales. Las calificaciones y la condición de aprobación son indicadores institucionales útiles, pero no agotan el aprendizaje [8]. En el presente estudio, el desenlace primario será aprobación/desaprobación y la nota final se analizará como resultado secundario. Las evaluaciones parciales, asistencia, entrega de tareas e interacción con el entorno virtual se consideran evidencias temporales que actualizan el riesgo durante el semestre.

La autoeficacia influye en la elección de actividades, persistencia y respuesta ante dificultades [6]. La autorregulación comprende planificación, monitoreo y ajuste de estrategias [7]. En matemática, estas dimensiones pueden coexistir con ansiedad, creencias estereotipadas y experiencias previas. Su inclusión permite pasar de una predicción basada solo en notas históricas a una interpretación educativa del riesgo, siempre que los instrumentos sean válidos y que los datos no se confundan con explicaciones causales.

La ansiedad matemática comprende respuestas de tensión, preocupación y evitación ante tareas numéricas. Los metaanálisis encuentran una asociación negativa consistente con el desempeño y muestran que la magnitud puede variar por etapa educativa y contexto [23], [24]. Para evitar una lectura esencialista, el instrumento distinguirá autoeficacia, valor de la tarea y ansiedad; una estudiante puede valorar la ingeniería y, al mismo tiempo, experimentar ansiedad por experiencias previas. Estas escalas se usarán como señales modificables para apoyo, no como rasgos permanentes.

III-C. Minería de datos educativos y redes bayesianas

La minería de datos educativos y la analítica del aprendizaje extraen patrones de registros académicos y digitales para comprender y mejorar procesos formativos [9]–[11]. Entre sus aplicaciones se encuentran la predicción de rendimiento, identificación de deserción, modelamiento del conocimiento y recomendación de recursos. Las redes bayesianas son especialmente pertinentes cuando se necesita combinar predicción, incertidumbre e interpretabilidad [14].

Una red bayesiana $\mathcal{B} = (G, \Theta)$ contiene un grafo dirigido acíclico G y parámetros Θ. Si $X_1, \ldots, X_p$ son las variables y $\mathrm{Pa}(X_i)$ sus padres, la distribución conjunta se factoriza como

$$P(X_1, \ldots, X_p) = \prod_{i=1}^{p} P\big(X_i \mid \mathrm{Pa}(X_i)\big). \tag{1}$$

Al observar evidencias $\mathbf{e}$, la probabilidad de riesgo se actualiza mediante $P(\mathsf{Riesgo} \mid \mathbf{e})$. Esta propiedad permite formular escenarios

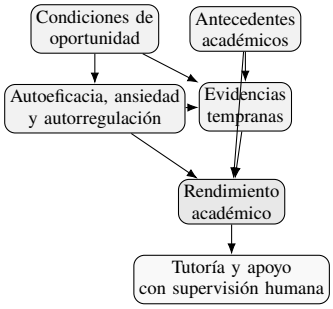


Figura 1. Modelo conceptual orientado a oportunidades académicas.

"qué pasaría si", por ejemplo, cuánto cambia el riesgo cuando mejora la asistencia o se completa la primera evaluación [12], [13].

El aprendizaje de estructura puede realizarse con algoritmos de búsqueda y puntuación, como hill-climbing con BIC o BDe, disponibles en `bnlearn` [15]. Sin embargo, en muestras pequeñas una búsqueda irrestricta produce redes inestables y sobreajustadas. Por ello, se limitará el número de padres por nodo, se bloquearán arcos temporalmente imposibles y se comparará la red aprendida con modelos parsimoniosos, como Naive Bayes o Tree-Augmented Naive Bayes.

III-D. Equidad algorítmica, calibración e interpretabilidad

Un modelo orientado a derechos debe evaluarse más allá de la exactitud global. La discriminación mide la capacidad de ordenar casos de mayor y menor riesgo; la calibración verifica si, entre casos con riesgo estimado de 0.30, aproximadamente 30 % presenta el evento [18]. En desenlaces desbalanceados, la curva precisión-recall es más informativa que la ROC por sí sola [19]. La auditoría por subgrupos examina si sensibilidad, falsos positivos y calibración cambian según condiciones de oportunidad, sin convertirlas automáticamente en criterios de exclusión [16], [17].

La Fig. 1 presenta el modelo conceptual. El sexo/género delimita la población de interés, pero no funciona como predictor dentro de una base compuesta exclusivamente por estudiantes mujeres. Las condiciones de oportunidad y los antecedentes pueden influir en hábitos, autoeficacia, acceso digital y trayectoria temprana; estas evidencias alimentan la estimación del resultado. La intervención académica ocurre después de la predicción y debe documentarse para evitar que futuras bases mezclen el efecto del apoyo con el riesgo inicial.

III-E. Marco conceptual operativo

Los conceptos centrales se delimitan en la Tabla IV. La unidad lógica del estudio es una predicción individual producida en un momento previamente definido. La red estima una probabilidad, mientras que el rendimiento real se observa al finalizar el curso. Por tanto, el modelo predictivo no se operacionaliza mediante una encuesta ni se trata como variable independiente; se evalúa como sistema mediante calidad estructural, desempeño, calibración, estabilidad y utilidad educativa.

IV. Metodología

IV-A. Diseño, población y estrategia muestral

La investigación será aplicada, cuantitativa, no experimental y predictiva. Tendrá una fase retrospectiva para desarrollo y una fase prospectiva/temporal para validación. La base de desarrollo incluirá todos los registros elegibles de estudiantes mujeres matriculadas en Matemática Básica entre 2022-II y 2024-II; la cohorte 2025-B se reservará para validación temporal. Se empleará censo de casos válidos, no muestreo probabilístico. El tamaño final se reportará después de depuración y desagregación por sexo/género; no se fijará a priori el total de 45 consignado en una matriz preliminar del proyecto, porque dicho número corresponde al grupo general y no garantiza suficientes eventos para un modelo exclusivo de mujeres [1].

Los criterios de inclusión serán: i) registro de matrícula en el curso y periodo definido; ii) registro administrativo de sexo femenino o autodeclaración como mujer, según la fuente autorizada disponible, manteniendo conceptualmente diferenciadas ambas categorías y sin inferir identidad de género a partir del nombre u otros identificadores indirectos; y iv) disponibilidad de predictores definidos antes del desenlace. Se excluirán duplicados, registros con resultado inconsistente y casos sin información mínima. Se describirá el flujo de inclusión y las razones de exclusión. Si el número de desaprobaciones es reducido, se priorizará un modelo parsimonioso y se evitará afirmar estabilidad que los datos no sostengan.

La suficiencia de información se evaluará por el número de eventos de desaprobación, la cantidad efectiva de parámetros y la estabilidad en remuestreo, no únicamente por el total de matrículas. Antes de entrenar la red se elaborará una tabla por cohorte con número de mujeres, eventos, prevalencia y completitud. Si la base histórica no permite estimar tablas condicionales confiables, las opciones válidas serán reducir nodos, fusionar categorías con fundamento, usar priors informativos documentados o presentar el estudio como piloto; aumentar complejidad para "mejorar" el ajuste aparente no será aceptable.

IV-B. Variables e instrumentos

La Tabla V operacionaliza los dominios. La variable objetivo primaria será la condición final aprobado/desaprobado conforme a la normativa institucional; la nota final de 0 a 20 será secundaria. Se diferenciarán predictores disponibles antes del inicio y evidencias de las primeras cuatro semanas para construir dos modelos: uno basal y otro de actualización temprana.

El cuestionario base se reorganizará en cuatro escalas: hábitos de estudio, autoeficacia y ansiedad matemática, acceso/uso académico de recursos digitales y autorregulación. Los ítems se redactarán sin atribuir capacidades a estereotipos de género. La validez de contenido se estimará con juicio de al menos tres especialistas y coeficiente V de Aiken; la consistencia interna se reportará mediante alfa ordinal u omega, con intervalos de confianza. Cuando una escala no alcance evidencia suficiente, sus ítems no se sumarán automáticamente y se analizarán individualmente o se excluirán.

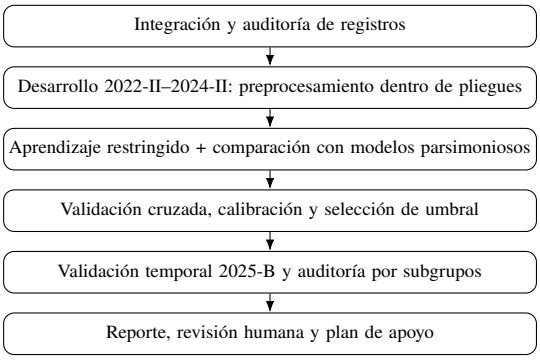


Figura 2. Flujo analítico y de gobernanza del modelo.

El contenido será evaluado en claridad, pertinencia y representatividad. Se conservarán las decisiones de los jueces y la versión final del instrumento. La V de Aiken se interpretará con intervalos y no como un umbral mecánico [25]. Para escalas ordinales se preferirá omega o alfa basado en correlaciones policóricas; la confiabilidad no sustituirá el examen de dimensionalidad [26]. Un piloto verificará comprensión, tiempo de respuesta y presencia de ítems sensibles o redundantes.

IV-C. Preparación de datos y prevención de fuga de información

Se elaborará un diccionario de datos, se verificará rango, unidad, temporalidad y procedencia de cada variable, y se conservará un registro reproducible de transformaciones. Los valores faltantes se describirán por variable y cohorte. La imputación, si corresponde, se realizará dentro de cada pliegue de validación para impedir fuga de información; se comparará con un análisis de casos completos. Las categorías poco frecuentes se agruparán solo con justificación sustantiva.

No se utilizarán como predictores la nota final, evaluaciones posteriores al punto de alerta ni variables que revelen retrospectivamente el resultado. Tampoco se incluirán nombres, códigos personales o atributos sin finalidad educativa. La cohorte 2025-B permanecerá cerrada hasta fijar el modelo y el umbral de decisión.

IV-D. Construcción de la red bayesiana

El proceso seguirá una adaptación de CRISP-DM: comprensión del problema, comprensión y preparación de datos, modelamiento, evaluación y despliegue controlado [22]. La Fig. 2 resume el procedimiento. Las variables continuas se mantendrán cuando el estimador lo permita o se discretizarán con puntos definidos en el conjunto de entrenamiento; ninguna frontera se seleccionará usando la cohorte temporal.

Para el aprendizaje de estructura se compararán: i) Naive Bayes; ii) Tree-Augmented Naive Bayes; iii) hill-climbing o tabu search con puntuación BIC/BDe y lista negra de arcos imposibles. Se limitará a dos el número máximo de padres por nodo, salvo justificación experta. La estabilidad de los arcos se estimará mediante remuestreo *bootstrap*; solo las relaciones con frecuencia suficiente y coherencia temporal se interpretarán. Los parámetros discretos se estimarán con suavizado Dirichlet para evitar probabilidades cero [13], [15].

La especificación final se establecerá antes de abrir la cohorte de validación temporal. Se compararán Naive Bayes, Tree-Augmented Naive Bayes y redes obtenidas mediante hill-climbing o tabu search; se limitará inicialmente a dos el número máximo de padres por nodo y se impedirán mediante lista negra las relaciones temporalmente imposibles. La discretización y los hiperparámetros se definirán únicamente con los datos de entrenamiento.

Cuadro IV
DEFINICIONES CONCEPTUALES Y FUNCIÓN ANALÍTICA

Concepto	Definición adoptada	Función en el estudio
Igualdad de oportunidades académicas	Condiciones institucionales y materiales que permiten acceder a recursos, participar, progresar y recibir apoyo sin discriminación directa o indirecta [2], [3].	Marco normativo e interpretativo; orienta selección de variables y acciones posteriores.
Rendimiento académico	Nivel de logro formal en Matemática Básica, medido primariamente por aprobación/desaprobación y secundariamente por nota final y trayectoria evaluativa [8].	Variable objetivo observada al cierre del curso.
Riesgo académico temprano	Probabilidad estimada, antes del resultado final, de no alcanzar la condición de aprobación bajo las condiciones observadas.	Salida probabilística; no equivale a diagnóstico ni sentencia.
Red bayesiana	Grafo dirigido acíclico y conjunto de distribuciones condicionales que representan dependencias probabilísticas [12], [13].	Motor de inferencia y explicación de escenarios.
Discriminación	Capacidad de asignar mayor riesgo a casos que presentan el evento que a casos que no lo presentan [20].	ROC-AUC, PR-AUC y métricas al umbral.
Calibración	Concordancia entre probabilidades estimadas y frecuencias observadas [18].	Brier, curva, intercepto y pendiente.
Equidad algorítmica contextual	Ausencia de disparidades injustificadas en errores y calidad probabilística entre subgrupos relevantes, considerando propósito y consecuencias [16], [17].	Auditoría y criterio para autorizar, ajustar o suspender el uso.
Alerta formativa	Comunicación revisable que activa oferta de tutoría, nivelación o recursos, sin alterar derechos ni calificaciones.	Producto institucional sujeto a supervisión humana.

Cuadro V
DOMINIOS DE VARIABLES Y FUENTES

Dominio	Indicadores principales	Fuente
Antecedentes	promedio previo, diagnóstico de matemática, créditos, intentos previos	registros
Oportunidades	primera generación, colegio, trabajo, cuidados, conectividad	ficha breve
Psicoeducativo	hábitos, autoeficacia/ansiedad matemática, autorregulación	Likert
Trayectoria temprana	asistencia, tareas, actividad virtual, primera evaluación	registros/LMS
Resultado	nota final y aprobación/desaprobación	acta oficial

IV-E. Validación y contraste de hipótesis

En el conjunto de desarrollo se aplicará validación cruzada estratificada de cinco pliegues, repetida cuando el número de casos lo permita. La evaluación temporal se efectuará una sola vez en 2025-B. Se reportarán ROC-AUC, PR-AUC, sensibilidad, especificidad, valor predictivo positivo y negativo, F1 y matriz de confusión con intervalos *bootstrap*. El umbral se seleccionará en desarrollo priorizando sensibilidad, porque el costo principal es no ofrecer apoyo a una estudiante que finalmente desaprueba [19], [20].

Entre los umbrales que alcancen una sensibilidad mínima de 0,75 se seleccionará aquel con mayor especificidad, y dicho umbral quedará fijado antes de evaluar la cohorte 2025-B.

La calidad probabilística se evaluará con puntaje de Brier

$$\mathrm{BS} = \frac{1}{n}\sum_{i=1}^{n}(\hat{p}_i - y_i)^2, \tag{2}$$

intercepto y pendiente de calibración, y gráfico observado-predicho [18], [21]. H1 se examinará mediante información mutua, estabilidad de arcos y comparación del modelo basal con el temprano. H2 se contrastará con BIC/log-score y Brier frente al modelo nulo y la referencia simple. H3 se juzgará con los criterios preespecificados de la Tabla III; no se reducirá a una prueba de p.

IV-F. Plan de análisis y selección del modelo

La selección se efectuará con reglas definidas antes de abrir la cohorte temporal. Primero se excluirán modelos con fallas de datos, arcos no plausibles o calibración interna deficiente. Entre los restantes se elegirá la alternativa más simple cuyo rendimiento sea comparable dentro de la incertidumbre. La Tabla VII separa evaluación estadística y decisión institucional.

Se realizará análisis de sensibilidad para: definición de retiro, tratamiento de faltantes, discretización, inclusión de la primera evaluación y umbral de alerta. También se calculará la utilidad neta de reglas simples de apoyo cuando sea posible; una mejora estadística sin acción pedagógica viable no justificará el despliegue. El código, diccionario de datos y semillas se conservarán en un repositorio institucional con acceso acorde con la confidencialidad.

IV-G. Auditoría de equidad y ética

Aunque la población está compuesta por mujeres, existe heterogeneidad interna. Se compararán sensibilidad, tasa de falsos positivos, Brier y calibración por primera generación universitaria, procedencia escolar, carga de trabajo/cuidados y acceso digital, siempre que haya tamaño suficiente. H4 se considerará apoyada si las brechas son pequeñas y sus intervalos no sugieren una disparidad relevante; de lo contrario, se recalibrará, se modificará el umbral o se descartará el uso operativo [16], [17].

La recopilación y análisis de información se realizará únicamente después de contar con las autorizaciones institucionales y éticas correspondientes. La información obtenida directamente de las estudiantes requerirá consentimiento informado. Asimismo, no se publicarán resultados desagregados de subgrupos cuyo reducido número permita razonablemente la reidentificación de las participantes.

Los registros serán seudonimizados y el cuestionario requerirá consentimiento informado. Se aplicarán minimización de datos, control de acceso y conservación limitada. El resultado individual será visible solo para el equipo autorizado y se comunicará como probabilidad revisable, no como diagnóstico. Ninguna estudiante perderá derechos, matrícula o calificación por la predicción. La alerta solo habilitará apoyos voluntarios y la revisión humana prevalecerá sobre la salida algorítmica [27].

V. PROTOCOLO DE ANÁLISIS Y REPORTE DE RESULTADOS

La presente sección preespecifica los análisis que serán ejecutados en la fase empírica y no constituye todavía un reporte de desempeño predictivo del modelo.

V-A. Secuencia obligatoria de resultados

El proyecto fuente no contiene resultados empíricos; por integridad científica, este manuscrito no inventa métricas ni asociaciones. La versión posterior deberá iniciar con el diagrama de casos, describir la muestra desagregada por cohorte y documentar faltantes, distribución del resultado y confiabilidad de escalas. Después presentará la red final con frecuencia *bootstrap* de arcos y tablas de probabilidad condicional relevantes.

La Tabla VIII fija el reporte mínimo. Los valores entre corchetes son campos que deben completarse únicamente después de ejecutar el

Cuadro VI
OPERACIONALIZACIÓN ANALÍTICA PROPUESTA

Componente	Definición operacional	Indicadores/codificación inicial	Momento	Criterio de calidad
Resultado primario	Estado formal al cierre	desaprobada=1; aprobada=0; documentar retiros por separado	final	concordancia con acta oficial
Resultado secundario	Calificación acumulada	nota final 0–20; parciales y variación temporal	durante/final	reglas institucionales consistentes
Antecedentes	Información previa al curso	promedio previo, diagnóstico, créditos, intentos, cursos cuantitativos	antes de semana 1	sin variables posteriores al punto de alerta
Oportunidades	Restricciones y recursos	primera generación, empleo, horas de cuidado, conectividad, procedencia escolar	semana 1	categorías con finalidad explícita y consentimiento
Hábitos/autorregulación	Conductas de planificación y estudio	puntajes o ítems Likert; tiempo, metas, revisión, búsqueda de ayuda	semana 1	V de Aiken; omega/alfa ordinal
Autoeficacia/ansiedad	Creencias de capacidad y respuesta afectiva	confianza para resolver tareas, preocupación, evitación, valor de la matemática	semana 1	estructura interna y consistencia
Trayectoria temprana	Evidencias generadas antes de la alerta	asistencia, tareas, actividad LMS y primera evaluación	semanas 1–4	sello temporal, cobertura y trazabilidad
Modelo	Sistema de inferencia probabilística	DAG, tablas condicionales, estabilidad, AUC, Brier, calibración	desarrollo/validación	validación interna y temporal
Equidad	Diferencias de calidad por subgrupo	brechas de sensibilidad/FPR, Brier y calibración	validación	intervalos, tamaño y relevancia práctica

Cuadro VII
CRITERIOS DE EVALUACIÓN Y RESPUESTA

Hallazgo	Respuesta prevista
AUC baja o IC amplio	no desplegar; ampliar datos o simplificar
Buena AUC, mala calibración	recalibrar y repetir validación temporal
Sensibilidad insuficiente	revisar umbral; no usar para alerta
Brecha relevante por subgrupo	analizar medición, recalibrar o suspender
Modelo complejo sin ganancia	elegir red parsimoniosa
Buen desempeño y estabilidad	piloto supervisado con auditoría continua

Cuadro VIII
PLANTILLA PREESPECIFICADA PARA EL REPORTE POSTERIOR DE RESULTADOS PREDICTIVOS

Indicador	CV interna	2025-B
ROC-AUC (IC95 %)	[ ]	[ ]
PR-AUC (IC95 %)	[ ]	[ ]
Sensibilidad / especificidad	[ ]/[ ]	[ ]/[ ]
F1 / valor predictivo positivo	[ ]/[ ]	[ ]/[ ]
Brier / pendiente de calibración	[ ]/[ ]	[ ]/[ ]
Mayor brecha de sensibilidad	[ ]	[ ]

análisis. También se comparará el modelo basal con el temprano para determinar cuánto aporta la información de las primeras semanas y si la mejora justifica la recolección adicional.

V-B. Guía para la discusión

La discusión deberá contrastar los factores observados con la teoría de autoeficacia, autorregulación y oportunidades académicas, y no convertir dependencias probabilísticas en causalidad. Si las notas previas dominan la predicción, se discutirá si el modelo aporta algo más que reproducir desigualdades de ingreso. Si variables modificables como asistencia, tareas o autoeficacia muestran información incremental, se justificarán intervenciones concretas. En todos los casos se examinarán transporte temporal, incertidumbre, tamaño de muestra y posibles sesgos de medición [6], [7], [16].

V-C. Limitaciones y condiciones de interpretación

Las limitaciones previsibles son el tamaño reducido de la población femenina en una sola escuela, cambios de sílabo o evaluación entre cohortes, errores en registros administrativos, autoselección en la encuesta y ausencia de variables sobre clima de aula. La red identifica dependencias predictivas, no efectos causales. Asimismo, una intervención posterior a la alerta modifica el resultado y debe registrarse: si una estudiante recibe tutoría y aprueba, el modelo no "falló" necesariamente, pues la finalidad de la alerta era cambiar la trayectoria.

Cuadro IX
ESQUEMA PRELIMINAR DE ACCIONES FORMATIVAS

Nivel	Oferta de apoyo	Seguimiento
Bajo	recursos universales y retroalimentación	revisión en siguiente hito
Medio	tutoría grupal, banco de problemas, plan de estudio	contacto voluntario y nueva estimación
Alto	entrevista individual, nivelación y derivación a servicios pertinentes	revisión humana semanal, sin sanción
Incierto	no etiquetar; solicitar datos o evaluación docente	mantener acceso a apoyos universales

La transportabilidad se limita inicialmente a Matemática Básica de la FIIS-UNAC. La aplicación en otra carrera, curso o periodo requerirá nueva calibración y validación. Las comparaciones entre subgrupos con pocos casos se presentarán como exploratorias, con intervalos y sin rankings. Cualquier dato de género deberá tratarse con respeto a la identidad y a la finalidad declarada.

V-D. Recomendaciones de implementación

Las conclusiones empíricas solo podrán redactarse después de la validación. De manera provisional, el diseño permite concluir que una alerta orientada a igualdad académica exige definición temporal del riesgo, calibración, auditoría por subgrupos, interpretabilidad y gobernanza. La Tabla IX traduce posibles niveles de riesgo en una oferta escalonada que deberá consensuarse con docentes, tutoría y representantes estudiantiles.

Se recomienda: i) consolidar registros longitudinales desagregados; ii) aplicar diagnóstico matemático y encuesta breve al inicio; iii) vincular cada alerta con apoyos disponibles y medibles; iv) documentar aceptación y efecto de la intervención; v) informar a las estudiantes sobre propósito, límites y revisión; y vi) suspender el modelo si su desempeño o equidad no se sostienen en nuevas cohortes.

VI. GOBERNANZA Y TRANSFERENCIA INSTITUCIONAL

El valor del modelo depende de un circuito de decisión explícito. La unidad académica deberá designar responsables de custodiar

Cuadro X
RESPONSABILIDADES PARA UN PILOTO INSTITUCIONAL

Actor	Responsabilidad principal
Escuela/Departamento	definir finalidad, recursos de apoyo y límites de uso
Unidad de datos	calidad, seguridad, versionado y ejecución reproducible
Docentes/tutores	contextualizar la alerta y acordar acciones con la estudiante
Estudiantes	conocer el propósito, decidir participación y solicitar revisión
Comité ético/calidad	auditar desempeño, equidad, incidentes y continuidad
Equipo investigador	documentar modelo, incertidumbre, cambios y evaluación

datos, ejecutar el algoritmo, revisar alertas, ofrecer apoyos y evaluar consecuencias. Una salida predictiva sin capacidad real de tutoría puede generar preocupación sin beneficio; por ello, el despliegue solo se justificará cuando exista una oferta de intervención, horarios compatibles y mecanismos para que las estudiantes acepten, rechacen o soliciten revisión de la alerta. La comunicación evitará expresiones deterministas como "fracasará" y empleará formulaciones probabilísticas y accionables.

La gobernanza debe diferenciar el uso para investigación del uso operativo. Durante la investigación, los puntajes no influirán en decisiones académicas y se analizarán de manera seudonimizada. Un eventual piloto requerirá aprobación institucional, capacitación de tutores y protocolo de incidencias. Cada consulta del sistema deberá registrar fecha, versión del modelo, variables disponibles, probabilidad generada, decisión humana y apoyo ofrecido. Esta trazabilidad permitirá investigar errores, cambios de criterio y efectos no previstos [16], [27].

Se establecerá una ficha de modelo con población, periodos de entrenamiento, variables, exclusiones, métricas, umbral, usos permitidos, usos prohibidos y fecha de próxima revisión. La versión se congelará durante cada cohorte; cualquier cambio en estructura, parámetros o umbral exigirá nueva validación. El monitoreo incluirá prevalencia de desaprobación, datos faltantes, desplazamiento de distribuciones, calibración y brechas de error. Una degradación relevante activará revisión y, de ser necesario, suspensión temporal.

La evaluación institucional no debe limitarse a verificar que el algoritmo predice. También debe determinar si la intervención mejora oportunidades: aceptación de tutorías, asistencia a nivelación, cambio de desempeño, satisfacción, percepción de respeto y ausencia de estigmatización. En una fase posterior podrá emplearse un diseño cuasiexperimental o escalonado para comparar estrategias de apoyo, separando la calidad predictiva del efecto causal de la intervención. El éxito no será producir más alertas, sino brindar ayuda oportuna sin restringir derechos.

Los productos transferibles serán: i) diccionario de datos y reglas de calidad; ii) instrumento revisado y manual de aplicación; iii) código reproducible; iv) red bayesiana documentada; v) tablero agregado para seguimiento, sin exposición pública de casos; vi) guía de tutoría asociada con factores modificables; y vii) informe anual de desempeño y equidad. Esta arquitectura permite que la investigación contribuya a la mejora académica aun si el primer modelo no alcanza los criterios de despliegue, pues dejará capacidades de gestión de datos, evaluación y acompañamiento.

VII. CONCLUSIÓN

La propuesta establece un procedimiento metodológico para desarrollar y validar un modelo predictivo basado en redes bayesianas orientado a la identificación temprana del riesgo académico de las estudiantes mujeres en Matemática Básica de la FIIS-UNAC. El diseño integra variables académicas, condiciones de oportunidad, trayectoria temprana, calibración, auditoría de equidad, supervisión humana y protección de datos. En esta fase no se formulan conclusiones sobre la capacidad predictiva del modelo, puesto que estas deberán sustentarse mediante la posterior validación empírica y temporal.

REFERENCIAS

[1] L. O. Quispe de la Cruz, S. G. Quispe, A. A. Vicente Galvez, and R. D. Mendoza Arenas, "Modelo predictivo basado en redes bayesianas para estimar el rendimiento académico en el curso de Matemática Básica en la FIIS-UNAC, 2025," proyecto de investigación no publicado, Universidad Nacional del Callao, Callao, Perú, 2025.

[2] UNESCO, *Changing the Equation: Securing STEM Futures for Women*. París, Francia: UNESCO, 2024.

[3] C. Chávez Irigoyen and SUNEDU, *Género y educación superior: una mirada al sistema universitario peruano desde las trayectorias estudiantiles*, Documento de Trabajo no. 1. Lima, Perú: SUNEDU, 2023.

[4] OECD, *PISA 2022 Results (Volumes I and II): Country Note–Peru*. París, Francia: OECD Publishing, 2023.

[5] S. Verdugo-Castro, A. García-Holgado, and M. C. Sánchez-Gómez, "The gender gap in higher STEM studies: A systematic literature review," *Heliyon*, vol. 8, no. 8, Art. e10300, 2022, doi: 10.1016/j.heliyon.2022.e10300.

[6] A. Bandura, *Self-Efficacy: The Exercise of Control*. New York, NY, USA: W. H. Freeman, 1997.

[7] B. J. Zimmerman, "Becoming a self-regulated learner: An overview," *Theory Into Practice*, vol. 41, no. 2, pp. 64–70, 2002.

[8] P. H. Winne and J. C. Nesbit, "The psychology of academic achievement," *Annual Review of Psychology*, vol. 61, pp. 653–678, 2010.

[9] A. Peña-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1432–1462, 2014.

[10] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *WIREs Data Mining and Knowledge Discovery*, vol. 10, no. 3, Art. e1355, 2020.

[11] G. Siemens, "Learning analytics: The emergence of a discipline," *American Behavioral Scientist*, vol. 57, no. 10, pp. 1380–1400, 2013.

[12] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. San Francisco, CA, USA: Morgan Kaufmann, 1988.

[13] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, USA: MIT Press, 2009.

[14] F. Ornelas and C. Ordonez, "Predicting student success: A Naïve Bayesian application to community college data," *Technology, Knowledge and Learning*, vol. 22, no. 3, pp. 299–315, 2017.

[15] M. Scutari, "Learning Bayesian networks with the bnlearn R package," *Journal of Statistical Software*, vol. 35, no. 3, pp. 1–22, 2010.

[16] R. S. Baker and A. Hawn, "Algorithmic bias in education," *International Journal of Artificial Intelligence in Education*, vol. 32, pp. 1052–1092, 2022.

[17] R. F. Kizilcec and H. Lee, "Algorithmic fairness in education," in *The Ethics of Artificial Intelligence in Education*, W. Holmes and K. Porayska-Pomsta, Eds. New York, NY, USA: Routledge, 2022, pp. 174–202.

[18] B. Van Calster *et al.*, "Calibration: The Achilles heel of predictive analytics," *BMC Medicine*, vol. 17, Art. 230, 2019.

[19] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLOS ONE*, vol. 10, no. 3, Art. e0118432, 2015.

[20] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[21] G. W. Brier, "Verification of forecasts expressed in terms of probability," *Monthly Weather Review*, vol. 78, no. 1, pp. 1–3, 1950.

[22] P. Chapman *et al.*, *CRISP-DM 1.0: Step-by-Step Data Mining Guide*. SPSS, 2000.

[23] N. M. Else-Quest, J. S. Hyde, and M. C. Linn, "Cross-national patterns of gender differences in mathematics: A meta-analysis," *Psychological Bulletin*, vol. 136, no. 1, pp. 103–127, 2010.

[24] C. Barroso *et al.*, "A meta-analysis of the relation between math anxiety and math achievement," *Psychological Bulletin*, vol. 147, no. 2, pp. 134–168, 2021.

[25] L. R. Aiken, "Three coefficients for analyzing the reliability and validity of ratings," *Educational and Psychological Measurement*, vol. 45, no. 1, pp. 131–142, 1985.

[26] R. P. McDonald, *Test Theory: A Unified Treatment*. Mahwah, NJ, USA: Lawrence Erlbaum, 1999.

[27] W. Holmes *et al.*, "Ethics of AI in education: Towards a community-wide framework," *International Journal of Artificial Intelligence in Education*, vol. 32, pp. 504–526, 2022.