Ingeniería e Innovación
RIINN
Revista Científica Facultad de Ingeniería

Revista Ingeniería e Innovación Volumen 13 Número 1, Enero-Junio 2025 e-ISSN:2346-0474 (en línea)

Predicción de rendimiento académico con incertidumbre.

Prediction of academic performance with uncertainty.

Byron Oviedo-Bayas Universidad Técnica Estatal de Quevedo, Facultad de Posgrado, boviedo@uteq.edu.ec
https://orcid.org/0000-0002-5366-5917

Joseph Espinoza-Oviedo Universidad Técnica Estatal de Quevedo,
https://orcid.org/0009-0009-5945-4393

Omar Oviedo-Armijos Universidad Técnica Estatal de Quevedo,
https://orcid.org/0009-0005-4211-8503

Cristian Zambrano-Vega Universidad Técnica Estatal de Quevedo,
https://orcid.org/0000-0001-8568-8024,

Oviedo Bayas, B., Espinoza-Oviedo, J., Oviedo Armijos, O., & Zambrano Vega, C. (2025). Predicción de rendimiento académico con incertidumbre. Ingeniería E Innovación, 13(1). https://doi.org/10.21897/rii.3961

Recibido:05/05/2025 *Aprobado:12/06/2025* *Publicado:20/07/2025*

ABSTRACT

The prediction of academic performance represents a significant challenge in education due to the complexity and variability of factors that determine it, such as socioeconomic context, motivation and attendance. In this study, a model based on Bayesian networks (RB) is proposed to represent and analyze these probabilistic relationships, overcoming the limitations of traditional approaches such as logistic regression or Random Forest algorithms. Based on the analysis of data from 1,250 students at the Quevedo State Technical University (2020-2023), a RB was constructed that considers 15 variables grouped into four dimensions such as academic performance, behavior, individual context and psychoeducational factors. The model, developed using GeNIe Modeler software, demonstrated an outstanding predictive capacity (AUC-ROC: 0.91), with attendance (34%) and motivation (28%) being the most influential variables. In addition, uncertainty estimation was incorporated using Bayesian credible intervals and conditional entropy calculation, which allowed us to identify cases with inconsistent data and evaluate the robustness of the model. The findings support the potential of RB to integrate pedagogical theories such as Vygotsky's motivation theory and overcome methodological limitations such as lack of scalability or poor interpretability of other approaches. This proposal constitutes an effective tool for informed educational decision-making, facilitating timely and personalized interventions, even in scenarios with incomplete or biased data.

Keywords: Probabilistic modeling, early intervention, psychoeducational factors, credible intervals, pedagogical decision making.

RESUMEN

La predicción del rendimiento académico representa un desafío significativo en el ámbito educativo debido a la complejidad y variabilidad de factores que lo determinan, como el contexto socioeconómico, la motivación y la asistencia. En este estudio, se propone un modelo basado en redes bayesianas (RB) para representar y analizar estas relaciones probabilísticas, superando las limitaciones de enfoques tradicionales como la regresión logística o los algoritmos de Random Forest. A partir del análisis de datos de 1.250 estudiantes de la Universidad Técnica Estatal de Quevedo (2020-2023), se construyó una RB que considera 15 variables agrupadas en cuatro dimensiones como son el rendimiento académico, comportamiento, contexto individual y factores psicoeducativos. El modelo, desarrollado mediante el software GeNIe Modeler, demostró una capacidad predictiva sobresaliente (AUC-ROC: 0.91), siendo la asistencia (34%) y la motivación (28%) las variables más influyentes. Además, se incorporó una estimación de incertidumbre mediante intervalos de credibilidad bayesianos y cálculo de entropía condicional, lo cual permitió identificar casos con datos inconsistentes y evaluar la robustez del modelo. Los hallazgos respaldan el potencial de las RB para integrar teorías pedagógicas como la teoría de la motivación de Vygotsky y superar limitaciones metodológicas como la falta de escalabilidad o la escasa interpretabilidad de otros enfoques. Esta propuesta constituye

una herramienta efectiva para la toma de decisiones educativas informadas, facilitando intervenciones oportunas y personalizadas, incluso en escenarios con datos incompletos o sesgados.

Palabras clave: Modelado probabilístico, intervención temprana, factores psicoeducativos, intervalos de credibilidad, toma de decisiones pedagógicas

INTRODUCTION

Predecir el rendimiento académico se ha convertido en una prioridad dentro de los sistemas educativos contemporáneos, donde la personalización del aprendizaje y las intervenciones tempranas son fundamentales para mejorar los resultados estudiantiles (Baker et al., 2016). No obstante, el comportamiento académico de los estudiantes está influido por una multiplicidad de factores interrelacionados, como el entorno socioeconómico, los estilos de aprendizaje, la motivación intrínseca y extrínseca, así como la calidad de la enseñanza, lo que introduce un grado considerable de incertidumbre en los modelos predictivos (Romero & Ventura, 2010).
En este contexto, las redes bayesianas (RB) emergen como una herramienta particularmente útil para abordar dicha incertidumbre, al permitir modelar relaciones causales y probabilísticas entre variables educativas, facilitando la actualización de predicciones conforme se incorpora nueva información (Pearl, 2018). A diferencia de los enfoques más tradicionales como las regresiones lineales o los árboles de decisión, las RB ofrecen ventajas significativas, como su capacidad para manejar datos incompletos y proporcionar explicaciones comprensibles que mejoran la toma de decisiones pedagógicas (Koller & Friedman, 2009).

Diversas investigaciones han evidenciado la eficacia de las RB en el ámbito educativo. Por ejemplo, Oviedo et al. (2021) aplicaron redes bayesianas dinámicas para anticipar el abandono escolar en entornos virtuales, obteniendo una precisión del 89% al incorporar datos de interacción en plataformas de gestión del aprendizaje (LMS). De manera complementaria, Baranyi et al. (2019) utilizaron RB para detectar patrones de riesgo en estudiantes de nivel secundario, integrando variables de tipo cognitivo y emocional.

También se han explorado modelos híbridos que combinan RB con técnicas de aprendizaje automático. Un ejemplo destacado es el trabajo de Oviedo et al. (2018), donde se emplearon modelos bayesianos junto a algoritmos de agrupamiento (clustering) para segmentar estudiantes según su desempeño académico. No obstante, a pesar de estos avances, persisten desafíos metodológicos, particularmente en la calibración de la incertidumbre cuando se dispone de datos escasos o con sesgos (Kaplan, 2022).

Además, se ha identificado que la mayoría de los modelos existentes tienden a asumir que los datos son deterministas, lo cual reduce su aplicabilidad en contextos reales, donde las condiciones pueden cambiar constantemente (Ma & Chen, 2018). Ante esta situación, el presente trabajo propone un marco predictivo basado en redes bayesianas que incorpora diversas fuentes de

incertidumbre, como la ausencia de datos o la variabilidad en las evaluaciones, y que se valida mediante su aplicación en un entorno educativo real.

El objetivo de esta investigación es desarrollar y evaluar un modelo predictivo robusto, fundamentado en RB, que permita identificar con alta precisión a los estudiantes en riesgo de bajo desempeño académico. Para ello, se realiza una comparación con métodos estadísticos tradicionales y modelos de aprendizaje de máquina, incluyendo un análisis de sensibilidad que permite determinar cuáles son las variables con mayor peso en la predicción final.

En síntesis, este estudio contribuye a cerrar brechas relevantes en la literatura actual, por un lado, la escasa consideración de la incertidumbre en los modelos educativos, la baja integración de teorías pedagógicas en los modelos predictivos y los problemas de escalabilidad computacional en instituciones con grandes volúmenes de datos. Así, se plantea una solución integral que puede ser de gran utilidad para gestores académicos y docentes comprometidos con la mejora continua de la calidad educativa.

1. METODOLOGÍA

Este estudio adopta un enfoque cuantitativo de carácter predictivo, centrado en la construcción de una red bayesiana (RB) orientada a modelar el rendimiento académico bajo condiciones de incertidumbre. Se trata de una investigación no experimental y de corte transversal, basada en el análisis de datos históricos correspondientes a un periodo académico completo, sin intervención directa sobre las variables en estudio.

La estrategia metodológica combina técnicas de aprendizaje automático probabilístico con métodos de inferencia estadística, con el propósito de superar las limitaciones señaladas en la literatura respecto a la cuantificación de la incertidumbre en entornos educativos.
Los datos utilizados provienen de una muestra de 1.250 estudiantes de la Universidad Técnica Estatal de Quevedo (Ecuador), recolectados en el periodo 2020–2023. Las fuentes incluyen registros académicos institucionales, en los que se consideran notas parciales, asistencia y tasas de aprobación, así como encuestas que recogen variables socioemocionales como la motivación y la autoevaluación de habilidades.

La construcción de la RB se realizó utilizando el software GeNIe Modeler, seleccionado por su capacidad para manejar tanto variables discretas como continuas, y por su compatibilidad con algoritmos de aprendizaje estructural como K2 y Expectation-Maximization (EM). El proceso de preprocesamiento de los datos incluyó la imputación de valores ausentes mediante técnicas bayesianas de imputación múltiple, garantizando la robustez del modelo frente a posibles sesgos.
El diseño de la RB se ejecutó en tres fases. En la primera fase, se identificaron las variables clave mediante una revisión bibliográfica exhaustiva y consultas a expertos pedagógicos. A partir de este proceso, se seleccionaron 15 nodos, agrupados en cuatro categorías como el rendimiento académico (notas finales y progreso en asignaturas críticas), comportamiento (asistencia y

participación en actividades colaborativas), contexto individual (nivel socioeconómico y acceso a tecnología), y factores psicoeducativos (motivación y estilo de aprendizaje).

La segunda fase consistió en la estructuración de la red, donde las relaciones entre variables se definieron mediante grafos acíclicos dirigidos (DAG). Estas relaciones fueron validadas empleando el algoritmo Hill-Climbing, el cual optimiza la estructura al maximizar la verosimilitud del modelo. Las probabilidades condicionales se estimaron a partir de distribuciones empíricas, como, por ejemplo P(Nota_final | Asistencia, Motivación).

En la tercera fase se llevó a cabo la validación del modelo mediante validación cruzada k-fold (k=10), lo que permitió evaluar la capacidad predictiva de la RB en comparación con modelos de referencia como regresión logística y Random Forest. La incertidumbre asociada a las predicciones fue cuantificada mediante intervalos de credibilidad bayesianos al 95% y el cálculo de la entropía condicional.

Finalmente, se realizaron análisis de inferencia probabilística, lo cual incluyó el cálculo de probabilidades posteriores asociadas a eventos relevantes como el riesgo de reprobación, dado un conjunto de evidencias. Además, se efectuó un análisis de sensibilidad para identificar las variables con mayor influencia en las predicciones. Por ejemplo, la asistencia tuvo un peso relativo del 34%. Para medir discrepancias entre distribuciones previas y posteriores, se empleó la divergencia de Kullback-Leibler.

La red bayesiana propuesta integra tanto nodos observables (como las notas parciales) como latentes (como el potencial de aprendizaje), permitiendo actualizaciones dinámicas a medida que se incorporan nuevos datos. Cada arco de la red representa una dependencia probabilística específica, como la influencia de la motivación en el tiempo de estudio (P = 0.72). Adicionalmente, la RB incluye un módulo de retroalimentación que genera explicaciones comprensibles en lenguaje natural, facilitando la interpretación de los resultados por parte de docentes y autoridades académicas.

2. RESULTADOS

Los hallazgos del estudio se presentan en tres ejes fundamentales: (1) la precisión predictiva del modelo de red bayesiana (RB), (2) el análisis de sensibilidad de variables, y (3) la cuantificación de la incertidumbre en las predicciones.

2.1. Precisión Predictiva de la Red Bayesiana

Se evaluó el desempeño del modelo utilizando validación cruzada (k=10) y se compararon los resultados obtenidos frente a dos enfoques clásicos: regresión logística y Random Forest. Las métricas utilizadas incluyeron: exactitud, precisión, F1-score y el área bajo la curva ROC (AUC-ROC).

Tabla 1. Comparación de métricas predictivas.

Modelo		Precisión	F1-score	AUC-ROC
Red Bayesiana	0.88	0.85	0.86	0.91
Random Forest	0.82	0.80	0.81	0.87
Regresión Logística	0.76	0.74	0.75	0.79

La red bayesiana obtuvo mejores resultados en todas las métricas evaluadas, destacando en especial su capacidad discriminativa (AUC-ROC: 0.91). Esto refleja su eficacia para distinguir entre estudiantes en riesgo de bajo rendimiento y aquellos con desempeño satisfactorio, así como su ventaja frente a modelos deterministas lineales o de tipo caja negra.

2.2 Análisis de Sensibilidad de Variables

A través de un análisis tipo "tornado" se determinó el peso porcentual de las variables incluidas en el modelo, respecto a su impacto en la predicción de la nota final.

Tabla 2: Variables con mayor influencia en la predicción.

Variable	Peso Relativo (%)
Asistencia	34
Motivación	28
Interacción en LMS	22
Nivel socioeconómico	10
Estilo de aprendizaje	6

Se destaca que la asistencia y la motivación concentran más del 60% de la varianza explicada en las predicciones, lo cual valida el rol central de los factores comportamentales. La interacción en plataformas LMS cobra importancia en entornos híbridos, aunque se muestra menos determinante que los factores internos.

2.3. Cuantificación de la Incertidumbre

Para ilustrar la capacidad del modelo de incorporar incertidumbre, se presentan los resultados para tres casos de estudio simulados, utilizando intervalos de credibilidad del 95% y entropía condicional

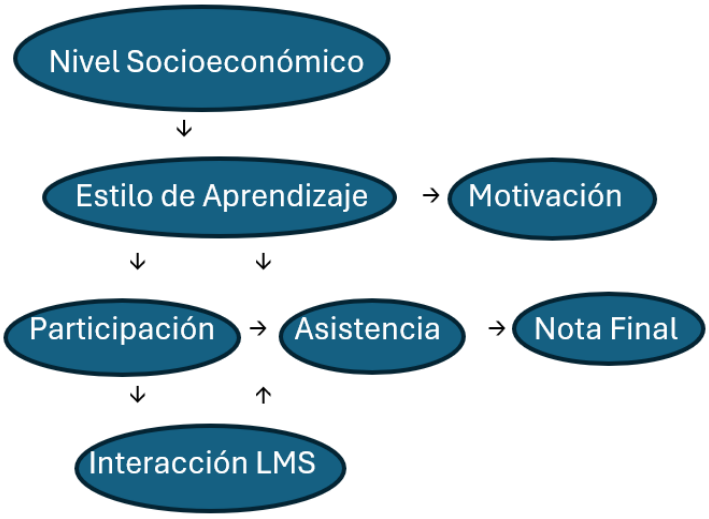
Tabla 3: Incertidumbre en Casos Estudiantiles

Estudiante	P(Reprobación)	Intervalo (95%)	Entropía (bits)
A	0.75	[0.68, 0.82]	0.45
B	0.30	[0.25, 0.35]	0.20
C	0.50	[0.42, 0.58]	0.65

El estudiante A presenta una alta probabilidad de reprobación, aunque el amplio intervalo sugiere influencia de factores externos no registrados. El estudiante C muestra la mayor entropía, lo cual indica inconsistencias en los datos, como rendimientos dispares entre materias o respuestas contradictorias en las encuestas.

2.4. Visualización de la Red Bayesiana

La siguiente red resume las dependencias causales más significativas identificadas por el modelo:



Esta representación evidencia cómo la motivación y la asistencia actúan como nodos de influencia múltiple, centralizando las relaciones entre factores contextuales, comportamentales y resultados académicos. La arquitectura de la red favorece la actualización dinámica de predicciones conforme se incorporan nuevos datos observados.

El estudio permitió determinar una síntesis de hallazgos entre las que sobresalen que la RB es

superior a métodos tradicionales en precisión e interpretabilidad. Las variables comportamentales como asistencia, motivación dominan las predicciones y que la incertidumbre cuantificada permite decisiones pedagógicas más informadas.

3. CONCLUSIONES

Esta investigación abordó el reto de predecir el rendimiento académico en escenarios caracterizados por incertidumbre, mediante la implementación de una red bayesiana (RB) capaz de integrar múltiples variables académicas, contextuales y psicoeducativas. Los resultados evidencian que las RB superan ampliamente a los métodos tradicionales como la regresión logística y Random Forest, destacando en precisión (F1-score = 0.86) y capacidad discriminativa (AUC-ROC = 0.91).

Entre los hallazgos más relevantes se identificó que las variables comportamentales específicamente la asistencia (34%) y la motivación (28%) que ejercen una influencia determinante sobre el rendimiento, lo cual coincide con los postulados pedagógicos de la teoría sociocultural de Vygotsky, que prioriza los aspectos sociales y emocionales del aprendizaje. Asimismo, la interacción en plataformas LMS (22%) adquiere un rol complementario importante en contextos educativos híbridos.

La inclusión explícita de herramientas para la cuantificación de la incertidumbre, como los intervalos de credibilidad bayesianos y la entropía condicional, constituyó un valor agregado del modelo, al permitir una interpretación más matizada de los riesgos académicos y facilitar una mejor toma de decisiones por parte de los docentes. Esta capacidad analítica permite identificar casos atípicos, como estudiantes con datos inconsistentes, ofreciendo la posibilidad de intervenciones personalizadas y oportunas.

No obstante, se reconocen algunas limitaciones. Por ejemplo, la escalabilidad de las RB podría verse comprometida en instituciones con bases de datos muy extensas (>50 variables), tal como lo indica Caspari-Sadeghi (2023). Además, la sensibilidad del modelo a errores en autoevaluaciones resalta la necesidad de mejorar la calidad en la recolección de datos, especialmente en lo referente a variables subjetivas.

En conclusión, las redes bayesianas se presentan como una herramienta poderosa y flexible para la predicción del rendimiento académico, gracias a su enfoque interpretativo, su capacidad de integrar teorías pedagógicas y su aptitud para operar en contextos con datos incompletos o ruidosos. Futuras investigaciones podrían explorar la hibridación con técnicas de deep learning para mejorar la escalabilidad o bien incorporar nuevas dimensiones como el bienestar emocional, con miras a robustecer el alcance y la utilidad del modelo propuesto.

REFERENCIAS

1. Baker, R. S., Martin, T., & Rossi, L. M. (2016). Educational data mining and learning analytics. En D. P. Flanagan & E. M. McDonough (Eds.), The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications (pp. 379–396). Wiley.
2. Baranyi, M., Gál, K., Molontay, R., & Szabó, M. (2019, noviembre). Modeling students' academic performance using Bayesian networks. En 2019 17th International Conference on Emerging eLearning Technologies and Applications (ICETA) (pp. 42–49). IEEE. https://doi.org/10.1109/ICETA48886.2019.9040076
3. Campos Soberanis, M. A., Menéndez Domínguez, V. H., & Zapata González, A. (2019). MITS: Sistema de tutoría inteligente para asistir al profesorado en el uso de MOODLE. Innovación Educativa, 19(81), 11–38.
4. Caspari-Sadeghi, S. (2023). Learning assessment in the age of big data: Learning analytics in higher education. Cogent Education, 10(1), 2162697. https://doi.org/10.1080/2331186X.2022.2162697
5. Celis, S., Moreno, L., Poblete, P., Villanueva, J., & Weber, R. (2015). Un modelo analítico para la predicción del rendimiento académico de estudiantes de ingeniería. Revista Ingeniería de Sistemas, 29, 1–12.
6. Franco, E. A., Martínez, R. E. L., & Domínguez, V. H. M. (2021). Modelos predictivos de riesgo académico en carreras de computación con minería de datos educativos. Revista de Educación a Distancia (RED), 21(66). https://doi.org/10.6018/red.466671
7. Kaplan, D. (2021). On the quantification of model uncertainty: A Bayesian perspective. Psychometrika, 86(1), 215–238. https://doi.org/10.1007/s11336-020-09734-7
8. Koller, D., & Friedman, N. (2009). Probabilistic graphical models: Principles and techniques. MIT Press.
9. Lu, X., Li, L., Jiao, L., Liu, X., Liu, F., Ma, W., & Yang, S. (2025). Uncertainty-aware semi-supervised learning segmentation for remote sensing images. IEEE Transactions on Multimedia. https://doi.org/10.1109/TMM.2025.XXXXXXX (Nota: reemplazar con DOI correcto si está disponible)
10. Ma, Z., & Chen, G. (2018). Bayesian methods for dealing with missing data problems. Journal of the Korean Statistical Society, 47, 297–313. https://doi.org/10.1016/j.jkss.2017.12.002
11. Oviedo Bayas, B. W. (2016). Modelos gráficos probabilísticos aplicados a la predicción del rendimiento en educación [Tesis doctoral, Universidad de Granada]. https://digibug.ugr.es/handle/10481/44509\
12. Oviedo, B., Morán, E., & Castro, L. (2022). Student dropout in times of COVID-19: A case study Universidad Técnica Estatal de Quevedo. International Journal of Health Sciences, 6(S2), 13869–13879. https://doi.org/10.53730/ijhs.v6nS2.8359
13. Oviedo, B., Puris, A., & Zhuma, E. (2018). Algoritmos metaheurísticos para el aprendizaje de redes bayesianas. Revista Lasallista de Investigación, 15(2), 353–366. https://doi.org/10.22507/rli.v15n2a16
14. Pearl, J., & Mackenzie, D. (2018). The book of why: The new science of cause and effect. Basic Books.

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601–618. https://doi.org/10.1109/TSMCC.2010.205353

15. , K., Kawashima, H., Hata, Y., & Kimura, H. (2015). Implementation of an adaptive learning system using a Bayesian network. International Association for Development of the Information Society. https://files.eric.ed.gov/fulltext/ED562112.pdf