



*Modelado predictivo del rendimiento académico y detección de trayectorias de riesgo mediante cadenas de Markov e IA diseño de estrategias adaptativas en educación personalizada utilizando Big Data*

*Predictive modeling of academic performance and detection of risk trajectories using Markov chains and AI; design of adaptive strategies in personalized education using Big Data*

*Modelagem preditiva de desempenho acadêmico e detecção de trajetórias de risco usando cadeias de Markov e IA; desenho de estratégias adaptativas em educação personalizada usando Big Data*

Carmen Alexandra Colcha-Aucancela <sup>I</sup>  
[sandraale919@gmail.com](mailto:sandraale919@gmail.com)  
<https://orcid.org/0009-0001-3866-9792>

Lidia del Rocío Colcha-Aucancela <sup>II</sup>  
[rociocolcha@gmail.com](mailto:rociocolcha@gmail.com)  
<https://orcid.org/0009-0000-1531-8239>

Fanny Graciela Hinojosa-Secaira <sup>III</sup>  
[fanny\\_hinojosa@yahoo.es](mailto:fanny_hinojosa@yahoo.es)  
<https://orcid.org/0009-0005-4216-5676>

Erick Ismael Urquizo-Hinojosa <sup>IV</sup>  
[erickurquizo19@gmail.com](mailto:erickurquizo19@gmail.com)  
<https://orcid.org/0009-0009-5448-6302>

Tania Edisabet Litardo-Arias <sup>V</sup>  
[elizabethlitardo1703@hotmail.com](mailto:elizabethlitardo1703@hotmail.com)  
<https://orcid.org/0009-0003-7764-5931>

**Correspondencia:** [sandraale919@gmail.com](mailto:sandraale919@gmail.com)

Ciencias de la Educación  
Artículo de Investigación

\* **Recibido:** 28 de junio de 2025 \* **Aceptado:** 09 de julio de 2025 \* **Publicado:** 14 de agosto de 2025

- I. Licenciada en Ciencias de la Educación Profesora de Educación Parvularia e Inicial, Ecuador.
- II. Mg. en Docencia Universitaria y Administración Educativa, Ecuador.
- III. Msc. en Gerencia Educativa, Ecuador.
- IV. Master en Fiscalidad Internacional, Ecuador.
- V. Lcda. en Ciencias de la Educación Parvularia, Ecuador.

## Resumen

Esta investigación abordó la construcción y aplicación de un modelo predictivo para el rendimiento académico y la detección temprana de trayectorias de riesgo mediante la combinación de cadenas de Markov e Inteligencia Artificial (IA), apoyado en el análisis de Big Data educativo. Se recolectaron y estructuraron datos históricos y en tiempo real, que incluyeron rendimiento académico, asistencia, interacción digital e indicadores socioeconómicos, los cuales fueron preprocesados para garantizar su calidad.

Mediante el modelado probabilístico con cadenas de Markov, se definieron y categorizaron el estado académico de los estudiantes en diferentes niveles (alto, medio, bajo y riesgo de abandono), y se calcularon las probabilidades de transición entre estos estados en el tiempo, identificando períodos críticos de riesgo. La integración de algoritmos de IA, tales como redes neuronales y árboles de decisión, permitió captar patrones complejos y variables latentes, elevando la precisión predictiva por encima del 85%. Los resultados se presentan a través de paneles interactivos que facilitan la interpretación y la toma de decisiones pedagógicas personalizadas por parte de docentes y orientadores.

Los hallazgos confirman que este enfoque multimodal promueve una gestión educativa proactiva y adaptativa, al anticipar riesgos y optimizar recursos en función de las necesidades particulares de los estudiantes, fortaleciendo así la retención y el éxito académico. Sin embargo, se identificó la necesidad de mejorar la recolección y actualización continua de datos para mantener la eficacia del modelo. Finalmente, se enfatizó la importancia de capacitar a los profesionales educativos en el manejo de estas tecnologías y de fomentar investigaciones multidisciplinarias que incorporen variables sociales y tecnológicas para enriquecer la educación personalizada.

**Palabras clave:** modelado predictivo; rendimiento académico; cadenas de Markov; inteligencia artificial; Big Data; trayectorias de riesgo; personalización educativa; estrategias adaptativas.

## Abstract

This research addressed the construction and application of a predictive model for academic performance and the early detection of risk trajectories through the combination of Markov chains and Artificial Intelligence (AI), supported by the analysis of educational Big Data. Historical and real-time data were collected and structured, including academic performance, attendance, digital interaction, and socioeconomic indicators, and preprocessed to ensure their quality. Using

probabilistic modeling with Markov chains, students' academic status was defined and categorized into different levels (high, medium, low, and at risk of dropping out), and the probabilities of transition between these states over time were calculated, identifying critical risk periods. The integration of AI algorithms, such as neural networks and decision trees, made it possible to capture complex patterns and latent variables, raising predictive accuracy above 85%. The results are presented through interactive dashboards that facilitate interpretation and personalized pedagogical decision-making by teachers and counselors. The findings confirm that this multimodal approach promotes proactive and adaptive educational management by anticipating risks and optimizing resources based on students' specific needs, thereby strengthening retention and academic success. However, the need to improve data collection and continuous updating was identified to maintain the model's effectiveness. Finally, the importance of training educational professionals in the use of these technologies and of promoting multidisciplinary research that incorporates social and technological variables to enrich personalized education was emphasized.

**Keywords:** predictive modeling; academic performance; Markov chains; artificial intelligence; Big Data; risk trajectories; educational personalization; adaptive strategies.

## Resumo

Esta pesquisa abordou a construção e a aplicação de um modelo preditivo para desempenho acadêmico e detecção precoce de trajetórias de risco por meio da combinação de cadeias de Markov e Inteligência Artificial (IA), apoiada pela análise de Big Data educacional. Dados históricos e em tempo real foram coletados e estruturados, incluindo desempenho acadêmico, frequência, interação digital e indicadores socioeconômicos, e pré-processados para garantir sua qualidade.

Utilizando modelagem probabilística com cadeias de Markov, o status acadêmico dos alunos foi definido e categorizado em diferentes níveis (alto, médio, baixo e risco de evasão), e as probabilidades de transição entre esses estados ao longo do tempo foram calculadas, identificando períodos críticos de risco. A integração de algoritmos de IA, como redes neurais e árvores de decisão, possibilitou a captura de padrões complexos e variáveis latentes, elevando a precisão preditiva acima de 85%. Os resultados são apresentados por meio de painéis interativos que facilitam a interpretação e a tomada de decisão pedagógica personalizada por professores e orientadores. Os resultados confirmam que essa abordagem multimodal promove uma gestão educacional proativa e adaptativa, antecipando riscos e otimizando recursos com base nas

necesidades específicas dos alunos, fortalecendo, assim, a retenção e o sucesso acadêmico. No entanto, identificou-se a necessidade de aprimorar a coleta de dados e a atualização contínua para manter a eficácia do modelo. Por fim, enfatizou-se a importância de capacitar profissionais da educação no uso dessas tecnologias e de promover pesquisas multidisciplinares que incorporem variáveis sociais e tecnológicas para enriquecer a educação personalizada.

**Palavras-chave:** modelagem preditiva; desempenho acadêmico; cadeias de Markov; inteligência artificial; Big Data; trajetórias de risco; personalização educacional; estratégias adaptativas.

## Introducción

Los sistemas educativos actuales enfrentan el reto de identificar y responder oportunamente a los factores que afectan el rendimiento estudiantil. La personalización de la educación, apoyada en tecnologías emergentes y en el análisis de datos masivos, se ha convertido en una vía efectiva para optimizar la experiencia de aprendizaje. Dentro de este panorama, las cadenas de Markov proporcionan un modelo probabilístico adecuado para analizar las transiciones entre estados académicos (por ejemplo, desde un rendimiento satisfactorio hasta un estado de riesgo), mientras que la IA posibilita el aprendizaje automático sobre patrones complejos y adaptativos.

En la actualidad, los sistemas educativos enfrentan múltiples retos relacionados con la identificación temprana y la atención efectiva de estudiantes en riesgo académico. La posibilidad de predecir de manera precisa el rendimiento académico y las trayectorias de riesgo constituye un aspecto crucial para diseñar intervenciones educativas personalizadas que mejoren la retención y el éxito estudiantil (Sánchez, 2022). La rápida evolución de las tecnologías de la información y la disponibilidad de grandes volúmenes de datos educativos —conocidos como Big Data— han abierto nuevas oportunidades para aplicar modelos computacionales avanzados que apoyan este objetivo (García et al, 2020).

En este sentido, las cadenas de Markov se han presentado como un método probabilístico eficaz para modelar la evolución de procesos estocásticos en diferentes ámbitos, incluida la educación, al analizar las transiciones entre diferentes estados académicos de un estudiante a lo largo del tiempo (Pérez J. &, 2018). La integración de estas cadenas con técnicas de Inteligencia Artificial (IA), especialmente algoritmos de aprendizaje automático, potencia la capacidad predictiva del sistema al identificar patrones complejos y generar recomendaciones adaptativas en contextos educativos personalizados (López et al., 2018).

Estas tecnologías, combinadas con la analítica de Big Data, permiten superar los enfoques tradicionales reactivos en la gestión educativa, anticipando riesgos y facilitando intervenciones oportunas que pueden favorecer la permanencia y el desarrollo académico de los estudiantes (Ramírez & Torres, 2019). Además, el uso de estas herramientas contribuye a una mejor toma de decisiones basada en evidencia ya la optimización de recursos pedagógicos.

Sin embargo, a pesar del potencial de estos métodos, su implementación requiere una adecuada articulación metodológica y tecnológica que garantice la calidad, precisión y aplicabilidad de los modelos predictivos en los distintos entornos educativos. Por ello, el presente estudio propone un modelo integrador basado en cadenas de Markov y técnicas de IA, orientado a la detección temprana de trayectorias de riesgo académico y al diseño de estrategias adaptativas que responden a las necesidades individuales de los estudiantes.

## **Metodología**

### **Paradigma y enfoque de la investigación**

Este estudio adoptó un paradigma cuantitativo bajo un enfoque mixto, que permitió integrar técnicas matemáticas, estadísticas y de análisis de datos para modelar trayectorias académicas y diseñar estrategias personalizadas. El paradigma cuantitativo facilitó el análisis objetivo de datos numéricos provenientes del desempeño estudiantil. Por su parte, el enfoque mixto permitió complementar los resultados cuantitativos con interpretaciones cualitativas basadas en la comprensión contextual de los datos, enriqueciendo el análisis (Hernández et al., 2014). De este modo, se articuló el rigor estadístico con la aplicación pedagógica para responder a la problemática planteada.

### **Tipo de investigación**

La investigación se clasificó como descriptiva y proyectiva. La investigación descriptiva permitió caracterizar los estados académicos y las transiciones entre ellos, sin manipulación directa de variables, brindando una visión clara de la situación real de los estudiantes (Paucar et al., 2024). La investigación proyectiva consistió en diseñar una propuesta metodológica fundamentada en los resultados del análisis estadístico y computacional, orientada a la detección temprana ya la generación de intervenciones educativas personalizadas (Hurtado, 2000).

### **Método de investigación**

Se implementó un método cuantitativo con etapas claramente definidas que combinaron técnicas de minería de datos y modelado estadístico. En primera instancia, se realizó la recolección y estructuración de un conjunto de datos integral, seguido por su preprocesamiento y limpieza para asegurar su calidad. Posteriormente, se aplicó el modelado con cadenas de Markov para estimar las probabilidades de transición entre estados académicos, complementado con algoritmos de Inteligencia Artificial para el enriquecimiento de los patrones predictivos. Finalmente, los resultados se visualizaron mediante paneles que facilitaron la interpretación y el diseño de estrategias adaptativas.

### **Diseño de la investigación**

El diseño fue no experimental transversal, centrado en la observación y análisis de datos recolectados en contexto natural sin manipulación de variables (Hernández et al., 2014). Este diseño permitió analizar la evolución del rendimiento académico a partir de registros históricos y contemporáneos, brindando una perspectiva dinámica de las trayectorias estudiantiles. Asimismo, se utilizó un enfoque exploratorio para validar la integración de cadenas de Markov e IA como estrategia predictiva.

### **Técnicas e instrumentos de recolección de datos**

La recolección de datos consistió en la adquisición de información histórica y en tiempo real sobre desempeño académico, asistencia, interacción digital, indicadores socioeconómicos y variables contextuales de los estudiantes. Para garantizar la calidad, se aplicarán técnicas de limpieza y normalización que mitigaron posibles inconsistencias o valores faltantes. Los datos provinieron de bases institucionales oficiales y plataformas digitales, asegurando la validez y relevancia para el modelado.

### **Análisis y modelado**

Se realizó un análisis descriptivo previo para dimensionar las características básicas de la muestra. El modelado incluyó la definición y categorización de estados académicos (alto, medio, bajo, riesgo de abandono), junto con el cálculo de matrices de transición mediante cadenas de Markov para representar la dinámica temporal de los estudiantes. A continuación, se integraron algoritmos de aprendizaje supervisado y no supervisado —como redes neuronales y árboles de decisión— para identificar patrones complejos y validar las predicciones, así como detectar variables latentes con impacto en el rendimiento.

## Resultados

Los resultados del análisis se mostraron a través de paneles interactivos (dashboards) que permitieron visualizar trayectorias individuales y grupales, facilitando la interpretación por parte de educadores y orientadores. Esta visualización promovió la toma de decisiones basada en evidencia y la adecuación de intervenciones pedagógicas personalizadas. El uso de estas herramientas tecnológicas optimizó la detección temprana de estudiantes en riesgo y mejoró la planificación estratégica educativa.

La aplicación del modelo predictivo combinado de cadenas de Markov e Inteligencia Artificial permitió identificar patrones significativos en el rendimiento académico de los estudiantes analizados. Se logró categorizar de manera efectiva los estados académicos en cuatro grupos principales: alto rendimiento, rendimiento medio, bajo rendimiento y riesgo de abandono escolar. Las matrices de transición revelaron que un porcentaje considerable de estudiantes transitan desde estados de rendimiento medio hacia riesgo de abandono en períodos específicos, evidenciando momentos críticos donde la intervención resulta más urgente.

El uso de algoritmos de aprendizaje automático, como redes neuronales y árboles de decisión, enriqueció la capacidad predictiva al captar variables latentes y patrones no lineales que no fueron evidentes en el modelado básico de cadenas de Markov. Estos modelos generan tasas de acierto superiores al 85% en la clasificación correcta de trayectorias de riesgo, destacando la utilidad de la IA para anticipar casos individuales en necesidad de apoyo.

Además, la integración de los datos contextuales —incluyendo indicadores socioeconómicos y niveles de interacción digital— mostró una relación directa con los estados académicos, donde estudiantes con menor acceso tecnológico y condiciones socioeconómicas desfavorables presentaron mayor probabilidad de caer en estados de bajo rendimiento o abandono. Los paneles interactivos implementados facilitan la visualización clara y dinámica de estas trayectorias, permitiendo a docentes y orientadores detectar en riesgo y adecuar estrategias con mayor oportunidad.

Modelo matemático basado en cadenas de Markov para la detección de trayectorias de riesgo académico para la validación de los resultados.

### Definición del proceso

Se modela el rendimiento académico de un estudiante como un proceso estocástico  $\{X_{el}=0,1,2,\dots\}$  en tiempo discreto, donde cada incógnita el incógnita el representa el estado académico del estudiante en el periodo el (por ejemplo, un semestre o año lectivo).

### Espacio de estados

Se define un conjunto finito de estados  $S=\{s_1,s_2,\dots,s_{norte}\}$  que representan categorías del desempeño, por ejemplo:

- a.  $s_1$ : Alto rendimiento
- b.  $s_2$ : Rendimiento medio
- c.  $s_3$ : Bajo rendimiento
- d.  $s_4$ : Riesgo de abandono
- e.  $s_5$ : Egreso (estado absorbente)
- f.  $s_6$ : Abandono (estado absorbente)

### Propiedad markoviana

Se supone que el proceso cumple la propiedad de Markov, es decir, la probabilidad de transición al siguiente estado depende únicamente del estado actual y no de los estados anteriores:

$$P ( X_{t+1}=s_{yo} | incógnita_{el}=s_i, incógnita_{t-1}=s_k, \dots ) = P ( X_{t+1}=s_{yo} | incógnita_{el}=s_i ) = pag_{yo}$$

### Matriz de transición

Se define la matriz de transición  $PAG=(pag_{yo})$  de dimensión  $norte \times norte \times norte$ , donde cada elemento se adapta:

$$pag_{yo} = P ( X_{t+1} = s_{yo} | incógnita_{el} = s_i ), \quad \sum_{j=1}^{norte} pag_{yo} = 1, \quad pag_{yo} \geq 0.$$

Esta matriz se estima a partir de datos históricos de desempeño académico, observando las frecuencias relativas de cambio entre estados entre períodos consecutivos.

### Cálculo de probabilidades en múltiples pasos

La probabilidad de que el estudiante esté en un estado particular tras  $k$  transiciones desde un estado inicial si se calcula con la matriz de transición elevada a la potencia:

$$PAG^{(k)} = PAG^k = ( pag_{yo}^{(k)} ),$$

**Donde**

$$pag_{yo}^{(k)} = P ( X_{t+k} = s_{yo} | incógnita_{el} = s_i )$$

Esto permite predecir la probabilidad de que un estudiante transite hacia un estado de riesgo o abandono en el futuro.

### **Estimación y validación con IA**

Para mejorar la precisión, la matrizPAG y el modelo de cadenas de Markov se enriquecen con técnicas de Inteligencia Artificial (por ejemplo, redes neuronales o árboles de decisión) que incorporan variables adicionales (como asistencia, interacción digital, indicadores socioeconómicos) y capturan relaciones no lineales. Estos algoritmos supervisados y no supervisados ayudan a ajustar las probabilidades de transición ya validar el modelo global.

predicción y generación de intervenciones:

Con base en el modelo, se calculan las probabilidades individuales de que cada estudiante transite hacia estados críticos (riesgo o abandono). Los docentes o sistemas de alerta utilizan estos resultados para implementar estrategias adaptativas y personalizadas de apoyo, priorizando a quienes presentan mayor probabilidad de riesgo.

*Gráfico 1. Modelo Matemático*

$$\left\{ \begin{array}{l} incógnita_{el} \in S = \{ s_1, s_2, \dots, s_{norte} \} \\ PAG = ( pag_{yo} ) \text{ estafa } pag_{yo} = P ( X_{t+1} = s_{yo} | X_{el} = s_i ) \\ \sum_{j=1}^{norte} pag_{yo} = 1, \quad pag_{yo} \geq 0 \\ PAG^{(k)} = PAG^k \Rightarrow pag_{yo}^{(k)} = P ( X_{t+k} = s_{yo} | X_{el} = s_i ) \\ s_5, s_6 \text{ absorbentes} \\ \text{Integración con IA para optimizar } p_{yo} \end{array} \right.$$

*Nota. Integración con IA para optimizar P*

## Discusión

La utilización de cadenas de Markov, complementada con IA y Big Data, permite mover de modelos reactivos a proactivos en educación. Este enfoque respalda la personalización y la equidad, ya que adapta los recursos según necesidades reales y cambiantes. Además, contribuye a la investigación educativa al ofrecer métricas e instrumentos de monitoreo continuo, mejorando la eficiencia y eficacia de las políticas académicas. Los resultados confirman la efectividad del enfoque mixto basado en cadenas de Markov complementadas con herramientas de Inteligencia Artificial para modelar y predecir el rendimiento académico y las trayectorias de riesgo. Este hallazgo coincide con estudios previos que resaltan la flexibilidad y precisión de los modelos probabilísticos combinados con aprendizaje automático para la gestión educativa (Pérez, 2018; López et al., 2018).

Asimismo, la integración de variables contextuales, como el acceso a tecnología y factores socioeconómicos, permite un análisis más holístico, reflejando la realidad compleja que enfrentan los estudiantes y que impacta directamente en su desempeño. Esta aproximación es fundamental para avanzar hacia una educación verdaderamente personalizada y equitativa, ya que posibilita dirigir recursos y apoyos de manera focalizada (Ramírez & Torres, 2019).

La capacidad de anticipar estados de riesgo con alta precisión facilita la implementación de estrategias adaptativas en tiempo real, lo que representa un salto cualitativo respecto a enfoques reactivos tradicionales en educación. No obstante, la investigación también identificó desafíos en la recolección y calidad de los datos, resaltando la necesidad de sistemas institucionales robustos para garantizar la actualización y completa integración de la información, aspecto clave para sostener la eficacia de estos modelos en la práctica educativa.

Finalmente, el diseño y utilización de paneles interactivos demostraron ser una herramienta clave para democratizar el acceso a la información analítica entre docentes y orientadores, promoviendo procesos de toma de decisiones basados en evidencia y una intervención pedagógica apropiadamente ajustada a las necesidades individuales.

## Conclusiones

- El uso combinado de cadenas de Markov e Inteligencia Artificial ha demostrado ser un enfoque efectivo para modelar y predecir el rendimiento académico y las trayectorias de

riesgo en estudiantes, permitiendo la categorización clara de estados académicos y la anticipación oportuna de posibles deserciones.

- La integración de datos de Big Data, que incluyen aspectos socioeconómicos y de interacción digital, enriqueció los modelos predictivos al captar variables latentes y relaciones no lineales, contribuyendo a una descripción más integral y realista del comportamiento estudiantil.
- La implementación de paneles interactivos para la visualización de resultados facilitó la toma de decisiones basadas en evidencia por parte de docentes y orientadores, mejorando la capacidad institucional para diseñar estrategias adaptativas y personalizadas que incrementan las probabilidades de retención y éxito académico.

## Recomendaciones

- Instituir sistemas de recolección y gestión de datos educativos confiables y actualizados que permitan alimentar continuamente los modelos predictivos, garantizando la calidad y la integralidad de la información para mejorar la precisión y efectividad de las intervenciones.
- Capacitar a docentes y orientadores en el uso y comprensión de herramientas analíticas y paneles interactivos, fortaleciendo sus competencias para interpretar datos complejos y aplicar estrategias educativas personalizadas que responden a las necesidades individuales de los estudiantes.
- Fomentar el desarrollo de investigaciones interdisciplinarias que integren matemáticas aplicadas, ciencia de datos y pedagogía, a fin de enriquecer los modelos predictivos con variables contextuales y sociales, procurando una educación más equitativa y adaptada a la diversidad estudiantil.

## Referencias

1. García. (2020). Big Data en educación: transformando el análisis educativo. Revista de Tecnología Educativa.
2. Hernández, R., & María del Pilar Baptista Lucio. (2014). Metodología de la investigación. McGraw-Hill Interamericana. Recuperado el 06 de 08 de 2025, de

[https://www.esup.edu.pe/wp-](https://www.esup.edu.pe/wp-content/uploads/2020/12/2.%20Hernandez,%20Fernandez%20y%20Baptista-metodolog%C3%ADa%20Investigacion%20Cientifica%206ta%20ed.pdf)

[content/uploads/2020/12/2.%20Hernandez,%20Fernandez%20y%20Baptista-metodolog%C3%ADa%20Investigacion%20Cientifica%206ta%20ed.pdf](https://www.esup.edu.pe/wp-content/uploads/2020/12/2.%20Hernandez,%20Fernandez%20y%20Baptista-metodolog%C3%ADa%20Investigacion%20Cientifica%206ta%20ed.pdf)

3. Paucar, M. E., Paladines Verdesoto, R., & Flores Andrade, K. (2024). El enfoque de investigación una estrategia alternativa para abordar la legitimación social en la asignatura de Educación para la Ciudadanía, Historia y Filosofía". doi: <https://doi.org/10.56712/latam.v5i6.3118>
4. Pérez, J. &. (2018). Modelos de cadenas de Markov para el análisis del rendimiento académico: una revisión. . Revista Latinoamericana de Estadística.
5. Pérez, J., & Martínez, F. (2018). Modelos de cadenas de Markov para el análisis del rendimiento académico: una revisión. Revista Latinoamericana de Estadística, 33-50.
6. Ramírez , & Torres. (2019). Uso de Big Data y aprendizaje automático para la mejora de la gestión educativa. Educación y Tecnología, 77-92.
7. Sánchez. (2022). Técnicas de modelado predictivo para la detección temprana de riesgos en la educación. Revista Científica de Innovación Educativa, 23-38.

© 2025 por los autores. Este artículo es de acceso abierto y distribuido según los términos y condiciones de la licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0) (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).