

DETECCIÓN DE PATRONES DE DIFICULTAD ACADÉMICA MEDIANTE ANÁLISIS DE SECUENCIAS Y CLUSTERING TEMPORAL

DETECTION OF ACADEMIC DIFFICULTY PATTERNS THROUGH SEQUENCE ANALYSIS AND TEMPORAL CLUSTERING

Autores: ¹Jorge Eduardo Ortiz Maldonado y ²Lidia del Rocío Castro Cepeda.

¹ORCID ID: <https://orcid.org/0009-0002-9439-783X>

²ORCID ID: <https://orcid.org/0000-0002-0471-2879>

¹E-mail de contacto: jorgee.ortiz@unach.edu.ec

²E-mail de contacto: lidia.castro@unach.edu.ec

Afiliación: ¹²Universidad Nacional de Chimborazo, (Ecuador).

Artículo recibido: 15 de Septiembre del 2026.

Artículo revisado: 2 de Octubre del 2026.

Artículo aprobado: 2 de Octubre del 2026.

¹Ingeniero Mecánico, graduado de la Escuela Superior Politécnica del Chimborazo, (Ecuador), Magíster en Administración de Empresas con especialidad en Dirección de Proyectos, graduado de la Universidad Viña del Mar, (Chile), Maestrante de la Maestría en Matemática Aplicada de la Universidad Nacional de Chimborazo, (Ecuador).

²Ingeniera Industrial, graduada de la Escuela Superior Politécnica de Chimborazo, (Ecuador). Máster Universitario en Ingeniería Matemática y Computación, graduado de la Universidad Internacional de la Rioja, (España). Máster Universitario en Ingeniería de la Energía, graduada de la Universidad Politécnica de Madrid, (España).

Resumen

El objetivo del estudio fue identificar patrones de dificultad académica en estudiantes de una universidad pública ecuatoriana a partir de sus trayectorias de rendimiento durante cuatro periodos consecutivos. Se analizaron los registros de 5 897 estudiantes con la ventana completa, codificados por periodo en tres estados: aprobación directa, aprobación mediante examen supletorio y reprobación final. Se aplicó un enfoque en dos etapas. Primero se separó a los estudiantes con aprobación directa sostenida del grupo con al menos un episodio de dificultad; después, sobre este grupo, se calculó una distancia de Hamming ponderada por las tasas de transición entre estados y se aplicó clustering jerárquico de Ward sobre una geometría euclidiana exacta, con el número de grupos fijado por una regla de estabilidad definida de antemano. El 74.7% de los estudiantes mantuvo aprobación directa en los cuatro periodos, con diferencias marcadas entre facultades. En el grupo con dificultad se identificaron siete perfiles que combinan el tipo de dificultad con el momento en que aparece, entre ellos uno de reprobación tardía y otro de estudiantes que cierran la ventana necesitando todavía el examen supletorio. El 63.0% de las reprobaciones finales ocurrió después de rendir ese examen, y casi la mitad del grupo con dificultad no

reprobó en ningún periodo. La separación entre perfiles fue débil y tres de ellos mostraron estabilidad limitada. Se concluye que el examen supletorio constituye la vía dominante del fracaso académico y una señal temprana útil para diseñar sistemas de alerta diferenciados.

Palabras clave: Análisis de secuencias, Clustering jerárquico, Dificultad académica, Trayectorias estudiantiles, Analítica del aprendizaje.

Abstract

The aim of this study was to identify patterns of academic difficulty among students at an Ecuadorian public university from their performance trajectories over four consecutive periods. The records of 5,897 students with a complete observation window were analyzed and coded, for each period, into three states: direct pass, pass through a supplementary examination, and final failure. A two-stage approach was applied. First, students with sustained direct passes were separated from the group with at least one episode of difficulty; then, within this group, a Hamming distance weighted by the transition rates between states was computed, and Ward hierarchical clustering was applied on an exact Euclidean geometry, with the number of groups set by a stability rule defined in advance. Of all students, 74.7% maintained direct passes in all

four periods, with marked differences between faculties. Within the group with difficulty, seven profiles were identified that combine the type of difficulty with the moment it appears, including a late-failure profile and one of students who close the window still needing the supplementary examination. Of all final failures, 63.0% occurred after sitting that examination, and almost half of the group with difficulty did not fail in any period. Separation between profiles was weak and three of them showed limited stability. It is concluded that the supplementary examination is the dominant route to academic failure and a useful early signal for designing differentiated warning systems.

Keywords: Sequence analysis, Hierarchical clustering, Academic difficulty, Student trajectories, Learning analytics.

Sumário

O objetivo do estudo foi identificar padrões de dificuldade acadêmica em estudantes de uma universidade pública equatoriana a partir de suas trajetórias de desempenho ao longo de quatro períodos consecutivos. Foram analisados os registros de 5.897 estudantes com a janela completa, codificados por período em três estados: aprovação direta, aprovação mediante exame de recuperação e reprovação final. Aplicou-se uma abordagem em duas etapas. Primeiro, separaram-se os estudantes com aprovação direta sustentada do grupo com pelo menos um episódio de dificuldade; em seguida, nesse grupo, calculou-se uma distância de Hamming ponderada pelas taxas de transição entre estados e aplicou-se o agrupamento hierárquico de Ward sobre uma geometria euclidiana exata, com o número de grupos definido por uma regra de estabilidade estabelecida previamente. Do total de estudantes, 74,7% mantiveram aprovação direta nos quatro períodos, com diferenças marcantes entre faculdades. No grupo com dificuldade, identificaram-se sete perfis que combinam o tipo de dificuldade com o momento em que ela surge, entre eles um de reprovação tardia e outro de estudantes que encerram a janela ainda necessitando do exame

de recuperação. Do total de reprovações finais, 63,0% ocorreram depois de realizar esse exame, e quase metade do grupo com dificuldade não reprovou em nenhum período. A separação entre perfis foi fraca e três deles mostraram estabilidade limitada. Conclui-se que o exame de recuperação constitui a via dominante do fracasso acadêmico e um sinal precoce útil para desenhar sistemas de alerta diferenciados.

Palavras-chave: Análise de sequências, Agrupamento hierárquico, Dificuldade acadêmica, Trajetórias estudantis, Analítica da aprendizagem.

Introducción

Los sistemas de seguimiento del rendimiento académico en las instituciones de educación superior ecuatorianas se apoyan en indicadores transversales, como la tasa de aprobación, el promedio de calificaciones o la tasa de deserción. Estos indicadores describen al estudiante en un momento dado, pero no informan sobre la dirección de su trayectoria. Un estudiante que reprueba en el primer semestre y luego se estabiliza tiene un perfil muy distinto del de otro que aprueba tres semestres y se desploma en el cuarto, aunque ambos arrojen promedios parecidos (Haas y Hadjar, 2020). En el contexto ecuatoriano, la deserción universitaria oscila entre el 20% y el 26% según la fuente y el periodo (SENESCYT, 2023; Ávila et al., 2024), y el Consejo de Aseguramiento de la Calidad de la Educación Superior exige a las instituciones mecanismos de seguimiento estudiantil como parte de la acreditación (CACES, 2024). Detectar a tiempo las trayectorias de dificultad tiene, por tanto, una aplicación inmediata.

En este estudio, el objeto de análisis son los patrones de dificultad académica, entendidos como toda trayectoria que se aparta de la aprobación directa sostenida. Dentro de ellos, se denomina deterioro a la transición desde la

aprobación directa hacia la necesidad del examen supletorio o hacia la reprobación a lo largo de periodos consecutivos. Esta definición distingue la dificultad sostenida tanto del bajo rendimiento puntual como de la deserción, que implica el abandono de la institución. Lo que interesa aquí es la trayectoria como proceso.

La minería de datos educativos y la analítica del aprendizaje han avanzado considerablemente en la predicción del rendimiento mediante modelos supervisados, desde árboles de decisión y bosques aleatorios hasta redes neuronales profundas (Romero y Ventura, 2020; Namoun y Alshanqiti, 2021). En América Latina, revisiones recientes confirman el peso de las variables socioeconómicas y académicas, pero permanecen concentradas en enfoques transversales (Tene et al., 2024; Real et al., 2024), al igual que los modelos predictivos aplicados en universidades mexicanas (Rico y Gaytán, 2022; Bautista Godínez et al., 2024) y ecuatorianas (Guanin-Fajardo et al., 2024). Los sistemas de alerta temprana derivados de esos modelos identifican con creciente precisión a los estudiantes en riesgo (Çırak et al., 2024), e incluso incorporan el tiempo hasta el abandono mediante análisis de supervivencia (Gutiérrez et al., 2023). Todos ellos comparten, sin embargo, una limitación de fondo: tratan la trayectoria como un problema de predicción de un desenlace, sin describir la forma que adopta la dificultad antes de que ese desenlace ocurra.

El análisis de secuencias, nacido en las ciencias sociales para estudiar trayectorias de vida (Abbott, 1995), ofrece una perspectiva complementaria: toma la secuencia completa de estados como unidad de análisis, mide la similitud entre trayectorias y agrupa las que se parecen para construir tipologías (Gabadinho et al., 2011; Studer y Ritschard, 2016). La disciplina ha incorporado en los últimos años

procedimientos para validar esas tipologías y juzgar su estabilidad (Studer, 2021; Liao et al., 2022). En el ámbito hispanohablante, Madero (2022) advierte lo poco familiar que resulta aún esta herramienta en las ciencias sociales latinoamericanas, y Carbonell y Simó (2022) revisan los métodos cuantitativos disponibles para estudiar trayectorias.

En educación superior, el análisis de secuencias se ha aplicado a trayectorias de inscripción, itinerarios curriculares y secuencias de rendimiento (Haas y Hadjar, 2020; Haas, 2022), por ejemplo, en universidades españolas (Villar y Muñoz, 2024), en la combinación de minería de datos y trayectorias curriculares (Maqsood et al., 2023) y en trayectorias de estudiantes de medicina en México (Canales et al., 2022). Su crecimiento en educación se ha concentrado, no obstante, en el nivel micro temporal, es decir, en las secuencias de acciones de aprendizaje dentro de una misma tarea: la revisión de Valle Torre et al. (2024), con 74 estudios, no incluye ninguno que analice estados académicos semestrales a lo largo de la carrera.

Hasta donde conocen los autores, no existen aplicaciones previas de la combinación que aquí se estudia: análisis de secuencias sobre estados de rendimiento semestrales, con un estado intermedio de tipo supletorio, en educación superior latinoamericana. Una búsqueda exploratoria en Scopus por la etiqueta literal no arroja resultados, si bien expresiones afines como *academic performance trajectories*, *student progression* o *academic risk sequences* devuelven trabajos relevantes; se trata, por tanto, de un vacío de aplicación en este contexto y no de una ausencia de antecedentes metodológicos. El examen supletorio, además, no tiene equivalente directo en los sistemas anglosajones, aunque la literatura sobre exámenes de recuperación muestra que estas

segundas oportunidades influyen en la progresión y en la forma en que los estudiantes distribuyen su esfuerzo (Schmidt et al., 2022; Nguyen et al., 2023).

El objetivo de este estudio es identificar y caracterizar patrones de dificultad académica a partir de las trayectorias de rendimiento de una cohorte universitaria, mediante un enfoque en dos etapas que combina la segmentación del grupo estable con el análisis de secuencias y el clustering del grupo con dificultad. El trabajo responde a tres preguntas: (PI1) ¿existen tipos cualitativamente distintos de trayectorias de dificultad que un indicador transversal no identifica?; (PI2) ¿la propensión a la dificultad y el tipo de trayectoria se asocian con la facultad?; y (PI3) ¿qué dinámicas de transición entre estados caracterizan a cada tipo de trayectoria?

Materiales y Métodos

Se realizó un estudio cuantitativo, retrospectivo y longitudinal, de alcance exploratorio y descriptivo, orientado a descubrir una tipología de trayectorias y no a predecir un desenlace. Se emplearon registros administrativos de calificaciones de una universidad pública ecuatoriana, proporcionados y depurados por su Dirección Académica, que abarcan cuatro periodos académicos consecutivos (2023-2S, 2024-1S, 2024-2S y 2025-1S). La base contiene 219 669 registros (estudiante, asignatura y periodo) de 14 667 estudiantes matriculados en 45 carreras de cuatro facultades. La única exclusión a nivel de registro fueron 135 filas sin situación académica (0.06%).

La población fueron todos los estudiantes con registros en la ventana, y la muestra analítica quedó formada por quienes presentan registros en los cuatro periodos ($n = 5\,897$), porque la distancia empleada compara las trayectorias

posición a posición y solo está definida entre secuencias de igual longitud. Se excluyó a 8 770 estudiantes (59.8%): 3 494 con un solo periodo, 2 748 con dos y 2 528 con tres. Esta exclusión es, en su mayor parte, estructural. La mitad de los excluidos (50.2%) ingresa después del primer periodo, y el 97.5% de ellos lo hace en primer o segundo semestre: son cohortes nuevas. Otro 42.3% deja de aparecer antes del último periodo, y el 79.6% de ellos cursaba entonces los dos últimos semestres de su carrera, lo que apunta a egreso. En conjunto, cerca del 84% de la exclusión se explica por la entrada o la salida regular de la ventana; el 16% restante concentra el riesgo de abandono y es donde reside el sesgo de supervivencia que se cuantifica en los resultados.

Para cada asignatura se asignó un estado según el desenlace del periodo: A, aprobación directa; S, aprobación mediante examen supletorio; y X, reprobación final, con o sin paso previo por el supletorio. El criterio es el desenlace y no el tránsito por la instancia de recuperación, porque agrupar en un solo estado a quienes aprueban el supletorio y a quienes lo reprueban desplazaría hacia la categoría intermedia una parte sustantiva del fracaso: en esta base, el 36.7% de los registros con supletorio termina en reprobación. El estado del periodo corresponde al desenlace más frecuente entre las asignaturas cursadas, y los empates (4.6% de las combinaciones estudiante-periodo) se resuelven por severidad, X antes que S y S antes que A, para privilegiar la detección de dificultad.

Conviene precisar qué resume esta agregación. Un estudiante cursa una mediana de 6 asignaturas por periodo, y el 37.9% de los periodos combina desenlaces distintos; la regla modal codifica como aprobación directa el 65.0% de los periodos con alguna reprobación

y el 74.6% de los que tienen algún supletorio. La tipología caracteriza, por tanto, trayectorias de dificultad predominante. La validez del estado S se sustenta en lo que anticipa: tras un estado S, la probabilidad de presentar dificultad en el periodo siguiente es de 0.300, frente a 0.066 tras una aprobación directa, un riesgo relativo de 4.57. Las secuencias se alinean por periodo calendario, de modo que las posiciones P1 a P4 denotan la ventana común de observación y no el semestre de carrera.

En la primera etapa se separó a los estudiantes con la secuencia AAAA, aprobación directa en los cuatro periodos, del grupo con al menos un episodio de dificultad. Esta segmentación evita que el grupo mayoritario, homogéneo por definición, absorba la solución de clustering y oculte la heterogeneidad de la minoría que interesa. La asociación entre la pertenencia al grupo con dificultad y la facultad se evaluó con la prueba chi-cuadrado, la V de Cramér y los residuos estandarizados ajustados. Como los estudiantes están anidados en carreras, la significación se corrigió mediante el ajuste de Rao y Scott (1984) y se contrastó con ecuaciones de estimación generalizadas con errores robustos (Liang y Zeger, 1986).

En la segunda etapa, sobre el grupo con dificultad, se calculó la distancia entre trayectorias. El costo de sustituir el estado i por el estado j se derivó de las tasas de transición observadas, según la definición empleada en TraMineR (Gabadinho et al., 2011):

$$c(i, j) = 2 - p(j | i) - p(i | j) \quad (1)$$

donde $p(j | i)$ es la probabilidad de pasar del estado i al estado j entre periodos consecutivos. Los costos resultantes fueron $c(A, S) = 1.072$, $c(A, X) = 1.188$ y $c(S, X) = 1.653$, y cumplen la desigualdad triangular. La distancia entre dos

trayectorias x e y de longitud T es la suma de los costos posición a posición:

$$D(x, y) = \sum c(x_t, y_t), \quad t = 1, \dots, T \quad (2)$$

que corresponde a una distancia de Hamming ponderada; con secuencias de igual longitud, el algoritmo de Optimal Matching se reduce a esta misma expresión. El agrupamiento se realizó con el método jerárquico de Ward (1963), que minimiza en cada fusión el incremento de la suma de cuadrados intragrupo. Ese criterio solo es exacto cuando las disimilitudes proceden de una configuración euclidiana (Murtagh y Legendre, 2014), y D no lo es: su matriz doblemente centrada presenta 24 autovalores negativos. D es, en cambio, de tipo negativo. En cada posición, los tres costos forman una métrica sobre tres puntos, y toda métrica sobre a lo sumo cuatro puntos es embebible en ℓ_1 (Deza y Laurent, 1997); la suma conserva esa propiedad y toda métrica ℓ_1 es de tipo negativo, lo que implica que $\sqrt{D}$ es exactamente euclidiana (Schoenberg, 1935).

Se comprobó numéricamente que la matriz doblemente centrada de $\sqrt{D}$ no tiene autovalores negativos, y Ward se aplicó sobre $\sqrt{D}$. Esta especificación equivale algebraicamente a aplicar el método ward. D sobre D en R ; como las secuencias duplicadas producen muchas disimilitudes empatadas, el orden de fusión puede variar entre implementaciones, por lo que los resultados se reportan con la implementación documentada en el repositorio.

El número de grupos se seleccionó con cuatro criterios: el índice de silueta promedio, el estadístico Gap con una distribución de referencia que permuta los estados dentro de cada periodo (Tibshirani et al., 2001), la estabilidad bootstrap medida con el índice de Rand ajustado y la estabilidad de Jaccard por clúster, ambas con 200 remuestreos (Hennig,

2007). Se adoptó de antemano como regla la primera solución cuyo índice de Rand ajustado medio supera 0.65, frontera de recuperación moderada según Steinley (2004). La estabilidad de Jaccard se interpretó con la escala de Hennig: igual o inferior a 0.5, clúster disuelto; por debajo de 0.6, no confiable; entre 0.6 y 0.75, patrón real de composición imprecisa; desde 0.75, estable. La solución se contrastó con validación en mitades aleatorias, k-medoides, clustering espectral (von Luxburg, 2007), un modelo de clases latentes por posición (Muthén y Muthén, 2000) y una mezcla de cadenas de Markov.

Para cada clúster se estimó la matriz de transición de primer orden, con intervalos de confianza bootstrap por percentiles (1 000 remuestreos), y se aplicaron los contrastes de homogeneidad y de orden de Anderson y Goodman (1957). Como los clústeres se construyen a partir de las propias secuencias, las diferencias en sus dinámicas están garantizadas por construcción; por ello, la PI3 se aborda de forma descriptiva y los contrastes se reportan como verificación. La sensibilidad se evaluó frente al sesgo de supervivencia, la métrica de distancia, la regla de codificación del periodo y una clasificación determinista alternativa.

El procesamiento se realizó en Python con NumPy, pandas, SciPy, scikit-learn y statsmodels, con semillas fijas (42 para los remuestreos y 123 para la validación en mitades), y el cuaderno se ejecuta de principio a fin en Google Colab. La implementación de la distancia se verificó numéricamente contra TraMineR en R. Los datos fueron anonimizado antes del análisis mediante la eliminación de nombres, documentos de identidad y cualquier identificador directo, y se trabajó exclusivamente con registros académicos

agregados en el marco de un proyecto de investigación institucional.

Resultados y Discusión

La muestra comprende 5 897 estudiantes distribuidos en Ciencias de la Educación, Humanas y Tecnologías (30.3%), Ciencias de la Salud (28.3%), Ciencias Políticas y Administrativas (21.1%) e Ingeniería (20.3%). De los 219 534 registros válidos, el 84.53% corresponde a aprobación directa, el 8.06% a aprobación tras el supletorio, el 4.67% a reprobación tras el supletorio y el 2.74% a reprobación sin esa instancia. La reprobación final alcanza, por tanto, al 7.41% de los registros, y el 63.0% de ella transita por el examen supletorio.

En la primera etapa, 4 406 estudiantes (74.7%) mantienen la secuencia AAAA, con una nota media de 8.38, y 1 491 (25.3%) conforman el grupo con dificultad, con una nota media de 7.31 (Figura 1). La asociación entre la pertenencia a este grupo y la facultad es de magnitud media-grande ($\chi^2 = 560.13$, $gl = 3$, $p < 0.001$; V de Cramér = 0.308). Ingeniería está sobrerrepresentada en el grupo con dificultad ($z = +23.39$) y Ciencias de la Educación, subrepresentada ($z = -11.19$). La conclusión resiste el anidamiento en carreras: con una correlación intraclase de 0.062 y un efecto de diseño de 14.3, el ajuste de Rao y Scott conserva la significación (χ^2 ajustado = 39.12, $p = 1.6 \times 10^{-8}$), al igual que las ecuaciones de estimación generalizadas (Wald $\chi^2 = 31.10$, $p = 8.1 \times 10^{-7}$).

En la figura 1 el grupo con dificultad contiene 79 secuencias distintas, y las diez más frecuentes cubren el 65.1% de los casos. La Figura 2 presenta tres de los cuatro criterios de selección. La silueta crece con K sin un codo marcado, efecto esperable con secuencias tan duplicadas, y alcanza 0.414 en la solución

adoptada, lo que indica una separación débil. La regla preespecificada selecciona $K = 7$ (índice de Rand ajustado bootstrap = 0.667; $K = 6$

obtiene 0.599), y la validación en mitades arroja 0.664.

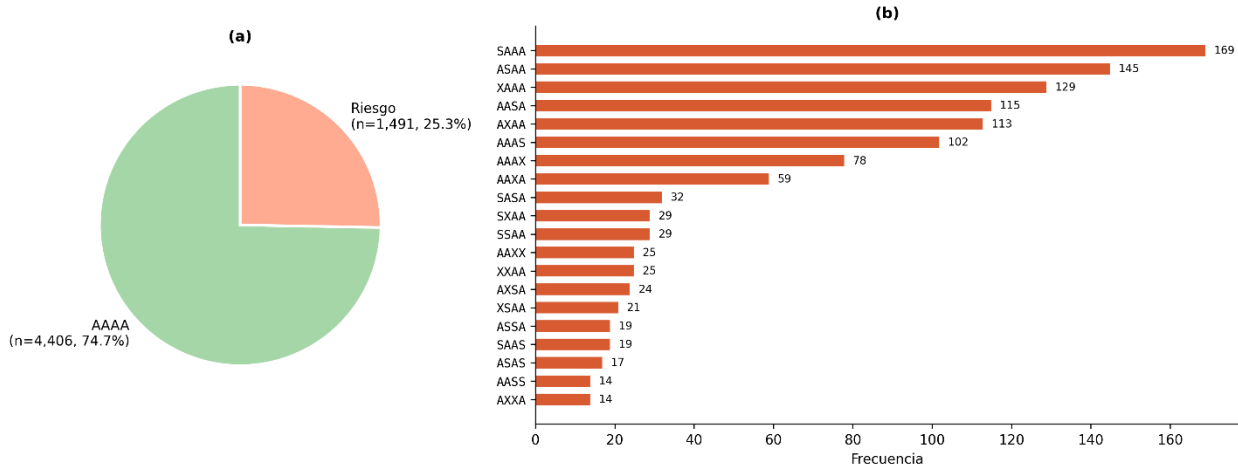


Figura 1. Distribución de las secuencias en la muestra ($n = 5\,897$): (a) clasificación en grupo estable y grupo con dificultad; (b) secuencias de dificultad más frecuentes.

Fuente: Elaboración propia.

El estadístico Gap señala $K = 8$, un refinamiento anidado que conserva seis de los siete clústeres y divide uno de ellos; se mantuvo $K = 7$ por fidelidad a la regla, y el contraste pareado no

distingue de forma concluyente entre ambas soluciones (diferencia de 0.09, IC 95% [-0.04, 0.17]). La tabla completa de criterios se publica como material suplementario.

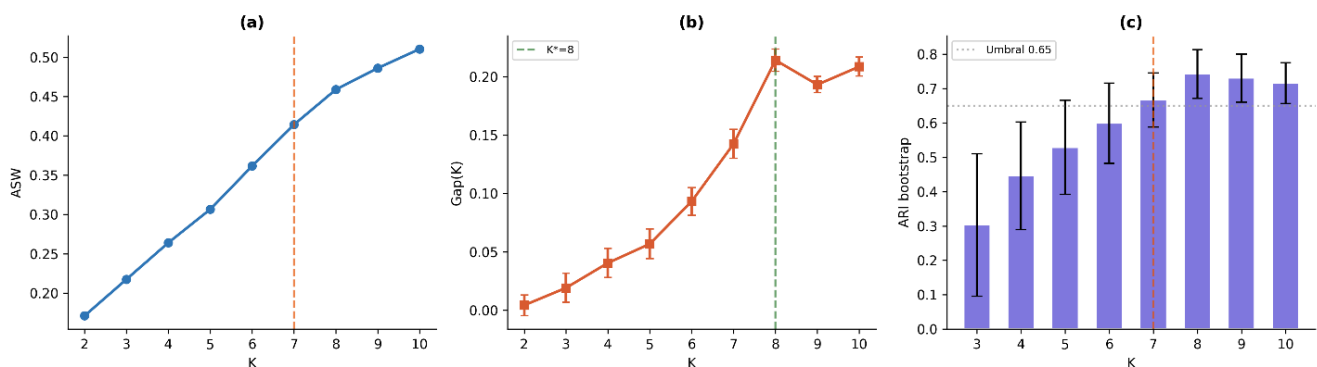


Figura 2. Criterios de selección del número de clústeres ($n = 1\,491$): (a) índice de silueta; (b) estadístico Gap; (c) estabilidad bootstrap.

Fuente: Elaboración propia.

La estabilidad por clúster es heterogénea: cuatro clústeres son estables (Jaccard entre 0.771 y 0.820), dos indican un patrón real de composición imprecisa (0.654 y 0.685) y uno, el de reprobación tardía, queda en la franja no

confiable (0.560). El contraste con otros métodos es instructivo por sus discrepancias. El clustering espectral reproduce la partición en buena medida (índice de Rand ajustado = 0.758), lo que indica que la estructura no

depende del algoritmo jerárquico; el modelo de clases latentes prefiere cuatro clases y concuerda solo moderadamente (0.357), y k-medoides concuerda poco (0.336). La mezcla de cadenas de Markov prefiere dos grupos (0.156), lo que tiene una explicación estructural: una cadena homogénea no puede codificar la posición del episodio, y las secuencias AXAA y AAXA, asignadas a

clústeres distintos, comparten exactamente las mismas transiciones. La tipología debe leerse, por tanto, como una partición útil e interpretable seleccionada con una regla verificable, no como una estructura latente única, en línea con las recomendaciones actuales para validar tipologías de secuencias (Studer, 2021; Liao et al., 2022).

Tabla 1. *Tipología de trayectorias del grupo con dificultad (n = 1 491).*

Cl.	Perfil	n	%	Secuencias típicas	Nota	P4 A (%)	P4 S (%)	P4 X (%)
C1	Reprobación en P1	227	15.2	XAAA, XSAA, XAAS	7.22	80.6	16.3	3.1
C2	Supletorio terminal	191	12.8	AAAS, SAAS, ASAS	7.46	0.0	100.0	0.0
C3	Reprobación tardía	321	21.5	AAAX, AAXA, AAXX	6.94	33.6	5.3	61.1
C4	Reprobación en P2	224	15.0	AXAA, SXAA, XXAA	7.04	92.9	0.0	7.1
C5	Supletorio en P2	174	11.7	ASAA, SSAA	7.63	100.0	0.0	0.0
C6	Supletorio en P1	169	11.3	SAAA	7.75	100.0	0.0	0.0
C7	Supletorio en P3	185	12.4	AASA, SASA, ASSA	7.55	95.1	4.9	0.0

Fuente: Elaboración propia.

La Tabla 1 presenta los siete perfiles. La partición se organiza sobre dos ejes: el tipo de dificultad predominante y el momento de la ventana en que aparece (Figuras 3 y 4). Las notas medias se mueven en un rango estrecho, de 6.94 a 7.75, de modo que un indicador transversal no distinguiría entre perfiles cualitativamente distintos, lo que responde afirmativamente a la PI1. Dos perfiles reúnen la reprobación temprana seguida de recuperación: C1, con la reprobación en el primer periodo, y C4, en el segundo.

En ambos, la mayoría cierra la ventana con aprobación directa (80.6% y 92.9%). C3 es el perfil más preocupante y el más numeroso: la reprobación aparece en la segunda mitad de la ventana y el 61.1% la cierra reprobando. Es también el clúster de frontera menos definida, porque mezcla dos situaciones que el refinamiento K = 8 separa: una reprobación terminal (n = 146), en la que todos cierran

reprobando, y una reprobación en el tercer periodo de la que el 62% se recupera. C2 aísla a quienes cierran la ventana necesitando el examen supletorio; el 84.3% de ellos no ha reprobado nunca, pero ocupa el estado que multiplica por 4.57 la probabilidad de dificultad en el periodo siguiente. Los tres perfiles restantes, C6, C5 y C7, pasan por el supletorio en el primer, segundo o tercer periodo sin reprobado en ninguno. Sumados, los estudiantes sin ninguna reprobación en la ventana son 689, el 46.2% del grupo con dificultad.

Las matrices de transición (Figura 5) completan esta lectura. Los perfiles de supletorio sin reprobación presentan las probabilidades de recuperación más altas (paso de S a A de 1.00 en C6, de 0.86 en C5, IC 95% [0.82, 0.90], y de 0.79 en C7, IC 95% [0.73, 0.84]). Los perfiles de reprobación temprana son los que mejor remontan desde ese estado (paso de X a A de 0.73 en C1 y de 0.66 en C4).

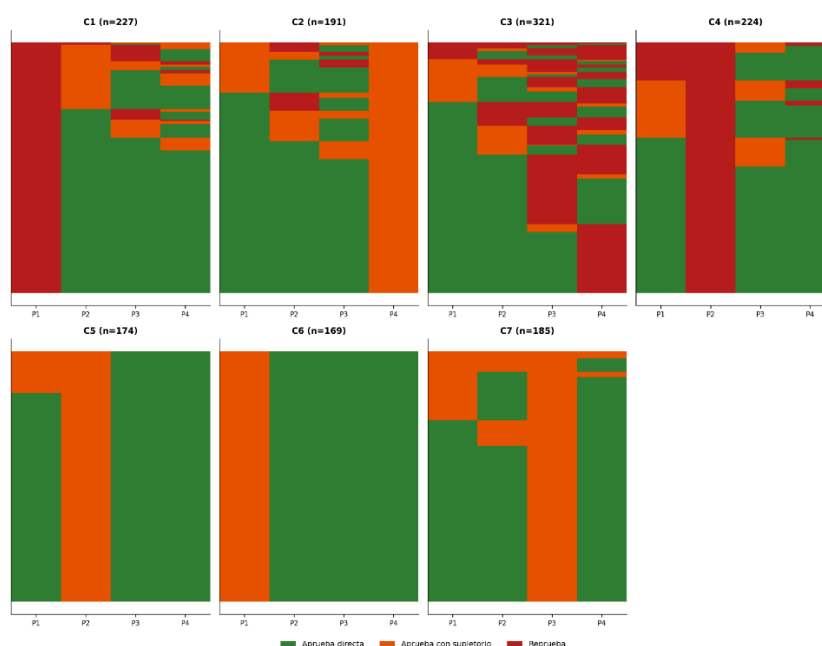


Figura 3. Trayectorias individuales por clúster ($K = 7$). Verde: aprobación directa; naranja: aprobación mediante supletorio; rojo: reprobación final.

Fuente: Elaboración propia.

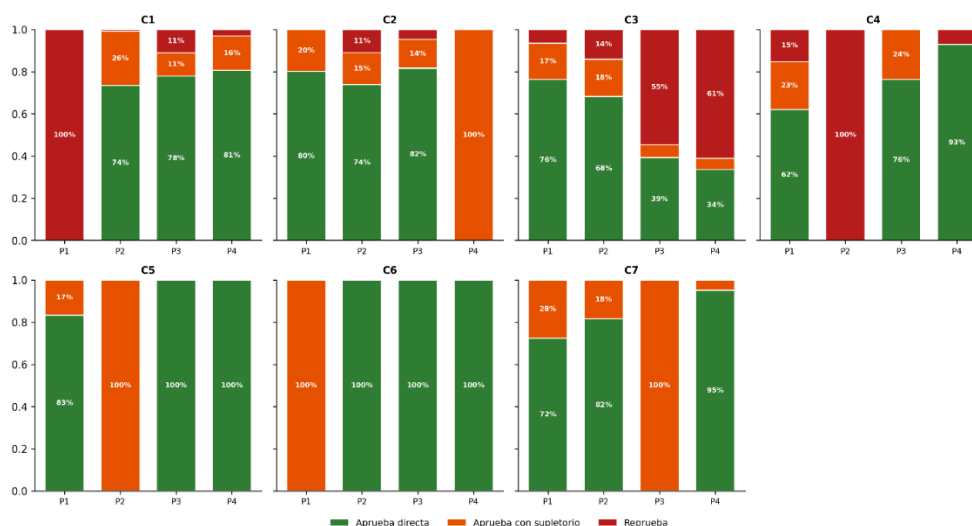


Figura 4. Distribución de estados por periodo en cada clúster.

Fuente: Elaboración propia.

En C3, la reprobación es un destino probable tanto desde la aprobación directa (0.45) como desde el supletorio (0.49), y en C2 la transición hacia el supletorio ocurre íntegramente en el último paso de la ventana. Los contrastes de

Anderson y Goodman confirman que las matrices difieren entre clústeres ($G^2 = 2\,193.30$, $gl = 36$, $p < 0.001$) y detectan dependencia de segundo orden ($G^2 = 250.98$, $gl = 12$, $p < 0.001$), lo que respalda el análisis de la secuencia completa frente a las transiciones aisladas.

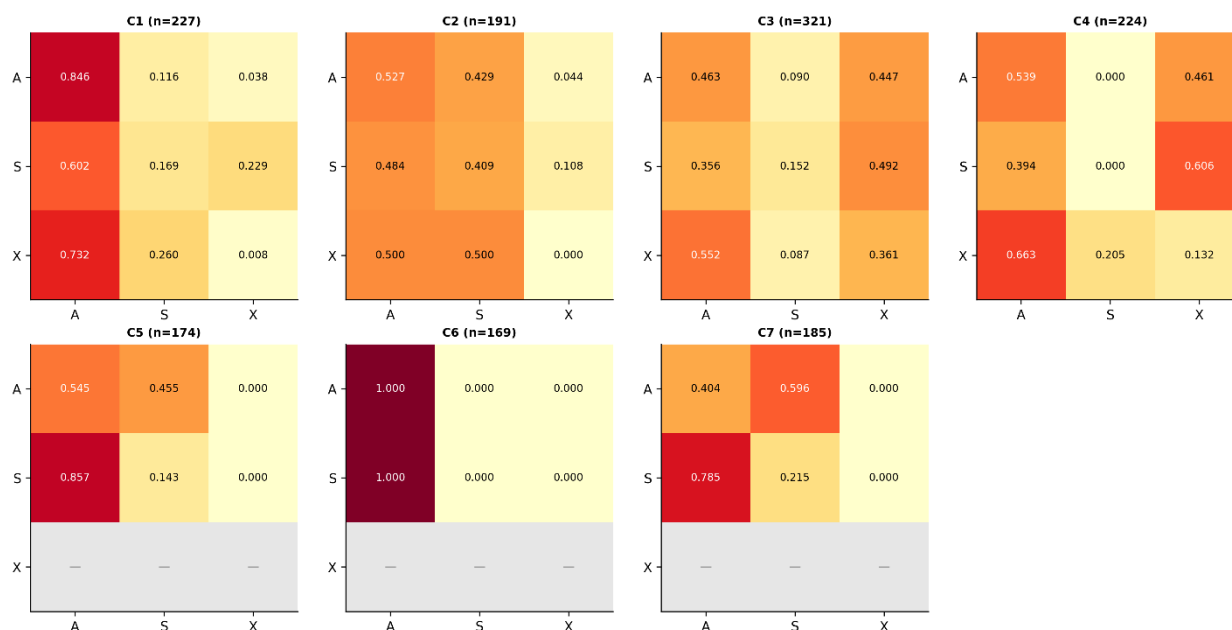


Figura 5. Matrices de transición por clúster. Las filas sin observaciones se muestran como no estimables.

Fuente: Elaboración propia.

Respecto de la PI2, la asociación entre clúster y facultad es débil pero detectable, y sobrevive al ajuste por conglomerados ($\chi^2 = 101.28$, $gl = 18$, $V = 0.150$; Rao-Scott $p = 0.004$). Ingeniería concentra la reprobación en el segundo periodo (C4, $z = +4.17$), Ciencias de la Salud la del primero (C1, $z = +4.76$) y evita el supletorio terminal (C2, $z = -5.51$), mientras que Ciencias Políticas y Administrativas presenta el patrón inverso en ambos perfiles. La facultad se asocia,

por tanto, sobre todo con quién entra en dificultad y, en menor medida, con el tipo de trayectoria. La sobrerepresentación de Ingeniería es consistente con la literatura que atribuye mayores tasas de dificultad a la brecha entre las competencias matemáticas de ingreso y las exigidas por el currículo (Zúñiga et al., 2020), aunque, al no disponer de medidas de conocimientos previos, esa explicación queda como hipótesis.

Tabla 2. Sensibilidad de los resultados a la regla de agregación del periodo.

Regla de agregación	Grupo con dificultad (n)	% de la muestra	Sin reprobación (%)	K por la regla	Concordancia con la tipología
Moda con desempate severo (principal)	1 491	25.3	46.2	7	Referencia
Moda con desempate leniente	982	16.7	54.3	5	0.323
Proporción, umbral 0.4	1 681	28.5	52.1	7	0.612
Proporción, umbral 0.2	3 226	54.7	43.5	Ninguna estable	0.176

Fuente: Elaboración propia.

La sensibilidad de los resultados se resume en la Tabla 2. Bajo cuatro reglas de agregación distintas, entre el 43.5% y el 54.3% de las trayectorias con dificultad no contiene ninguna

reprobación: la proporción de dificultad que se resuelve dentro del sistema de recuperación es estable. Varían, en cambio, la prevalencia del riesgo, el número de perfiles y la pertenencia

individual; la codificación por proporción con umbral de 0.4 vuelve a seleccionar siete perfiles, mientras que la regla leniente reorganiza las asignaciones. Con costos de sustitución unitarios, la estructura general se mantiene (índice de Rand ajustado = 0.518).

La estructura es en buena medida posicional, pero no se reduce a una regla determinista: una clasificación explícita de siete categorías basada en el tipo y la posición del episodio recupera la partición solo de forma parcial (0.673). Respecto del sesgo de supervivencia, los incluidos promedian 8.11 frente a 7.72 de los excluidos (d de Cohen = 0.312, efecto pequeño a mediano); el 24.8% de los excluidos presenta dificultad en su ventana parcial, una prevalencia comparable a la de la muestra, pero su último periodo termina en reprobación en el 14.2% de los casos, casi el doble de la tasa global, huella esperable del abandono que la ventana no alcanza a observar.

El hallazgo más relevante es que el examen supletorio no funciona únicamente como mecanismo de recuperación: es también la vía por la que ocurre la mayor parte del fracaso académico, ya que el 63.0% de las reprobaciones finales se produce después de rendirlo. Al mismo tiempo, casi la mitad del grupo con dificultad atraviesa la ventana sin reprobación, de modo que para muchos estudiantes el supletorio cumple su función. Esta doble condición coincide con la evidencia de que las segundas oportunidades favorecen la progresión, pero alteran la forma en que se distribuye el esfuerzo (Schmidt et al., 2022; Nguyen et al., 2023). Desde la teoría de la persistencia estudiantil (Tinto, 2017), los perfiles de dificultad sostenida pueden leerse como fallas progresivas en la integración académica, y los de recuperación sugieren preguntas para la investigación cualitativa sobre

qué distingue a quienes remontan. Estas implicaciones son directas para los sistemas de alerta temprana. Un estudiante con la secuencia AAAS no aparecería en ningún sistema basado en la reprobación, pero ocupa el estado que más anticipa la dificultad siguiente; por ello, el perfil C2 es el que mejor justifica una intervención preventiva, mientras que C3 requiere apoyo intensivo y los perfiles de supletorio sin reprobación, solo seguimiento. Estos modelos predictivos (Çırak et al., 2024; Gutierrez et al., 2023) y la tipología propuesta son complementarios: los primeros estiman quién está en riesgo y la segunda describe qué forma adopta la dificultad. Conviene recordar que «temprano» y «terminal» describen aquí posiciones en la ventana de observación y no semestres de la carrera.

El estudio tiene limitaciones. Se basa en una sola institución y en una ventana de cuatro periodos, y la exigencia de secuencias completas subestima la prevalencia de la dificultad. La regla modal describe la dificultad predominante y no toda la dificultad, y la pertenencia individual depende de la regla de codificación. La separación entre perfiles es débil y tres de los siete no alcanzan el umbral de estabilidad individual, de modo que la composición de esos grupos debe leerse con cautela. Finalmente, el diseño es observacional y no incluye variables socioeconómicas, de motivación ni de ingreso, por lo que las asociaciones con la facultad no deben interpretarse como relaciones causales.

Conclusiones

De los resultados obtenidos y de su discusión se desprenden las siguientes conclusiones: 1) la dificultad académica en la institución estudiada adopta formas temporales distintas, que se organizan según el tipo de dificultad y el momento en que aparece, y que un indicador

transversal no permite distinguir; 2) el examen supletorio es la vía por la que ocurre la mayor parte del fracaso académico y, al mismo tiempo, el mecanismo que resuelve la dificultad de cerca de la mitad de los estudiantes que la atraviesan, de modo que el paso por esa instancia constituye la señal temprana más informativa disponible; 3) la facultad se asocia con fuerza con la propensión a la dificultad y, en menor medida, con el tipo de trayectoria; 4) el enfoque en dos etapas, con una geometría compatible con el criterio de Ward y una regla de selección definida de antemano, produce una partición interpretable y verificable, aunque de separación débil, que debe leerse como descripción útil y no como estructura única; y 5) estos resultados aportan una base para diseñar sistemas de alerta temprana diferenciados según la forma temporal de la dificultad, cuya validación en otras instituciones y con ventanas más largas queda como línea de trabajo futuro.

Referencias Bibliográficas

- Abbott, A. (1995). Sequence Analysis: New Methods for Old Ideas. *Annual Review of Sociology*, 21(1), 93–113. <https://doi.org/10.1146/annurev.so.21.080195.000521>
- Anderson, T. W., y Goodman, L. A. (1957). Statistical Inference about Markov Chains. *The Annals of Mathematical Statistics*, 28(1), 89–110. <https://doi.org/10.1214/aoms/1177707039>
- Ávila, L., Cepeda, F., y Aucancela, R. (2024). Deserción en la Educación Superior en Ecuador, Causas y Consecuencias. *Ciencia Latina Revista Científica Multidisciplinar*, 8(3), 11475–11490. https://doi.org/10.37811/cl_rcm.v8i3.12472
- Bautista Godínez, T., Castañeda Garza, G., Pérez Mora, R., Ceballos, H. G., et al. (2024). Perspectives and Opportunities for Learning Analytics Integration. *Journal of Learning Analytics*, 11(1), 49–66. <https://doi.org/10.18608/jla.2024.8125>
- CACES. (2024). Modelo de Evaluación Externa 2024 con fines de Acreditación para Universidades y Escuelas Politécnicas. Consejo de Aseguramiento de la Calidad de la Educación Superior. <https://www.caces.gob.ec/universidades-y-escuelas-politecnicas-3/>
- Canales, D., Bautista, T., Moreno-Salinas, J. G., García-Minjares, M., y Sánchez-Mendiola, M. (2022). Academic trajectories analysis with a life-course approach: A case study in medical students. *Cogent Education*, 9(1), 2018118. <https://doi.org/10.1080/2331186X.2021.2018118>
- Carbonell, À., y Simó, M. (2022). El análisis cuantitativo de trayectorias laborales. Un estado del arte. *Papers. Revista de Sociologia*, 107(4). <https://doi.org/10.5565/rev/papers.3079>
- Çırak, C. R., Akıllı, H., y Ekinci, Y. (2024). Development of an early warning system for higher education institutions by predicting first-year student academic performance. *Higher Education Quarterly*, 78, e12539. <https://doi.org/10.1111/hequ.12539>
- Deza, M. M., y Laurent, M. (1997). Geometry of cuts and metrics. Springer.
- Gabardinho, A., Ritschard, G., Müller, N. S., y Studer, M. (2011). Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software*, 40(4), 1–37. <https://doi.org/10.18637/jss.v040.i04>
- Guanin-Fajardo, J. H., Guaña-Moya, J., y Casillas, J. (2024). Predicting academic success of college students using machine learning techniques. *Data*, 9(4), 60. <https://doi.org/10.3390/data9040060>
- Gutierrez, D., Garcia, G., Cuadros, E., Camara-Chavez, G., y Gomez-Nieto, E. (2023). Supporting decision-making process on higher education dropout by analyzing academic, socioeconomic, and equity factors through machine learning and survival analysis methods in the Latin American context. *Education Sciences*, 13(2), 154. <https://doi.org/10.3390/educsci13020154>

- Haas, C. (2022). Applying Sequence Analysis in Higher Education Research. En J. Huisman y M. Tight (Eds.), *Theory and Method in Higher Education Research* (Vol. 8, pp. 127–147). Emerald Publishing.
- Haas, C., y Hadjar, A. (2020). Students' trajectories through higher education: a review of quantitative research. *Higher Education*, 79(2), 223–250. <https://doi.org/10.1007/s10734-019-00458-5>
- Hennig, C. (2007). Cluster-wise assessment of cluster stability. *Computational Statistics & Data Analysis*, 52(1), 258–271. <https://doi.org/10.1016/j.csda.2006.11.025>
- Liang, K.-Y., y Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13–22. <https://doi.org/10.1093/biomet/73.1.13>
- Liao, T., Bolano, D., Brzinsky, C., Cornwell, B., Fasang, A. E., Helske, S., Piccarreta, R., Raab, M., Ritschard, G., Struffolino, E., y Studer, M. (2022). Sequence analysis: Its past, present, and future. *Social Science Research*, 107, 102772. <https://doi.org/10.1016/j.ssresearch.2022.102772>
- Madero-Cabib, I. (2022). Comprendiendo trayectorias de vida crecientemente diversas y complejas: El análisis de secuencias. *Revista de Sociología*, 37(1), 35–51. <https://doi.org/10.5354/0719-529X.2022.68148>
- Maqsood, R., Ceravolo, P., Ahmad, M., et al. (2023). Examining students' course trajectories using data mining and visualization approaches. *International Journal of Educational Technology in Higher Education*, 20, 55. <https://doi.org/10.1186/s41239-023-00423-4>
- Murtagh, F., y Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? *Journal of Classification*, 31(3), 274–295. <https://doi.org/10.1007/s00357-014-9161-z>
- Muthén, L. K., y Muthén, B. O. (2000). Integrating person-centered and variable-centered analyses. *Alcoholism: Clinical and Experimental Research*, 24(6), 882–891. <https://doi.org/10.1111/j.1530-0277.2000.tb02070.x>
- Namoun, A., y Alshanqiti, A. (2021). Predicting student performance using data mining and learning analytics techniques. *Applied Sciences*, 11(1), 237. <https://doi.org/10.3390/app11010237>
- Nguyen, N. T., Clark, C., y Juriansz, J. (2023). Effects of resit exams on student progression: An Australian case study. *Journal of University Teaching and Learning Practice*, 20(1), Article 6. <https://eric.ed.gov/?id=EJ1382379>
- Rao, J., y Scott, A. (1984). On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *The Annals of Statistics*, 12(1), 46–60. <https://doi.org/10.1214/aos/1176346391>
- Real-Delor, R., Guevara Tirado, A., Morales Ojeda, I., et al. (2024). Factores asociados al rendimiento académico en estudiantes universitarios de Latinoamérica en 2023. *Investigación en Educación Médica*, 13(51), 42–52. <https://doi.org/10.22201/fm.20075057e.2024.51.23580>
- Rico, A., y Gaytán, N. (2022). Modelos predictivos del rendimiento académico a partir de características de estudiantes de ingeniería. *IE Revista de Investigación Educativa de la REDIECH*, 13, e1426. https://doi.org/10.33010/ie_rie_rediech.v13i0.1426
- Romero, C., y Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), e1355. <https://doi.org/10.1002/widm.1355>
- Schmidt, H., Baars, G., Hermus, P., van der Molen, H. T., Arnold, I. J. M., y Smeets, G. (2022). Changes in examination practices reduce procrastination in university students. *European Journal of Higher Education*, 12(1), 56–71. <https://doi.org/10.1080/21568235.2021.1875857>

- Schoenberg, I. J. (1935). Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espaces distanciés vectoriellement applicables sur l'espace de Hilbert". *Annals of Mathematics*, 36(3), 724–732.
- SENESCYT. (2023). Estadísticas de Educación Superior, Ciencia, Tecnología e Innovación. Secretaría de Educación Superior del Ecuador. <https://siau.senescyt.gob.ec/portal-de-indicadores-de-educacion-superior/>
- Steinley, D. (2004). Properties of the Hubert-Arabie Adjusted Rand Index. *Psychological Methods*, 9(3), 386–396. <https://doi.org/10.1037/1082-989X.9.3.386>
- Studer, M. (2021). Validating sequence analysis typologies using parametric bootstrap. *Sociological Methodology*, 51(2), 290–318. <https://doi.org/10.1177/00811750211014232>
- Studer, M., y Ritschard, G. (2016). What matters in differences between life trajectories. *Journal of the Royal Statistical Society: Series A*, 179(2), 481–511. <https://doi.org/10.1111/rssa.12125>
- Tene-Tenempaguay, T., Martínez-Abad, F., y Hernández-Ramos, J. P. (2024). Factores asociados al rendimiento académico de los estudiantes latinoamericanos: Una revisión sistemática. *Profesorado, Revista de Currículum y Formación del Profesorado*, 28(3). <https://doi.org/10.30827/profesorado.v28i3.29626>
- Tibshirani, R., Walther, G., y Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B*, 63(2), 411–423. <https://doi.org/10.1111/1467-9868.00293>
- Tinto, V. (2017). Through the Eyes of Students. *Journal of College Student Retention*, 19(3), 254–269. <https://doi.org/10.1177/1521025115621917>
- Valle Torre, M., Oertel, C., y Specht, M. M. (2024). The Sequence Matters in Learning — A Systematic Literature Review. En *Proceedings of the 14th Learning Analytics and Knowledge Conference (LAK '24)* (pp. 263–272). ACM. <https://doi.org/10.1145/3636555.3636880>
- Villar-Aguilés, A., y Muñoz-Rodríguez, D. (2024). Complex trajectories in higher education students. *Journal of New Approaches in Educational Research*, 13(4). <https://doi.org/10.1007/s44322-023-00006-1>
- von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17(4), 395–416. <https://doi.org/10.1007/s11222-007-9033-z>
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236–244. <https://doi.org/10.1080/01621459.1963.10500845>
- Zúñiga, M. G., Vargas, E. A., y Maldonado, K. V. (2020). Análisis del rendimiento académico de los estudiantes de Ingeniería de Sistemas, posibilidades de deserción y propuestas para su retención. *Ingeniare. Revista Chilena de Ingeniería*, 28(4), 668–682. <https://doi.org/10.4067/S0718-33052020000400668>



Esta obra está bajo una licencia de Creative Commons Reconocimiento-No Comercial 4.0 Internacional. Copyright © Jorge Eduardo Ortiz Maldonado y Lidia del Rocío Castro Cepeda.

Declaraciones éticas y editoriales del artículo
Contribución de los autores (Taxonomía CRediT) Jorge Eduardo Ortiz Maldonado: conceptualización de la investigación, diseño metodológico, desarrollo del proceso investigativo, análisis formal de los datos, redacción del borrador original del manuscrito, revisión crítica del contenido científico y supervisión general del estudio. Lidia del Rocío Castro Cepeda: curación y organización de los datos, participación en la recolección de información, validación de los resultados obtenidos y elaboración de representaciones gráficas y visualización de los datos.
Declaración de conflicto de intereses Los autores declaran que no existe conflicto de intereses en relación con la investigación presentada, la autoría del manuscrito ni la publicación del presente artículo.
Declaración de financiamiento La presente investigación no recibió financiamiento específico de agencias públicas, comerciales o de organizaciones sin fines de lucro. En caso de existir financiamiento institucional o externo, este deberá ser declarado explícitamente por los autores en esta sección.
Declaración del editor El editor responsable certifica que el proceso editorial del presente artículo se desarrolló conforme a los principios de integridad científica, transparencia y buenas prácticas editoriales. El manuscrito fue sometido a un proceso de evaluación mediante revisión por pares doble ciego, garantizando la confidencialidad de la identidad de los autores y revisores durante todo el proceso de dictamen académico. Asimismo, el editor declara que el artículo cumple con los criterios científicos, metodológicos y éticos establecidos por la revista.
Declaración de los revisores Los revisores externos que participaron en la evaluación del presente manuscrito declaran haber realizado el proceso de revisión de manera objetiva, independiente y confidencial. Asimismo, manifiestan que no mantienen conflictos de interés con los autores ni con la investigación evaluada, y que sus observaciones y recomendaciones se fundamentan exclusivamente en criterios científicos, metodológicos y académicos.
Declaración ética de la investigación Los autores declaran que la investigación se desarrolló respetando los principios éticos de la investigación científica, garantizando la confidencialidad de los datos y el respeto a los participantes del estudio. En los casos en que la investigación involucre seres humanos, los procedimientos deben ajustarse a los principios éticos establecidos en la Declaración de Helsinki y a las normativas institucionales correspondientes.
Declaración sobre el uso de inteligencia artificial Los autores declaran que el uso de herramientas de inteligencia artificial, en caso de haberse utilizado durante el proceso de investigación o redacción del manuscrito, se realizó únicamente como apoyo técnico para mejorar la claridad del lenguaje o el análisis de información, manteniendo siempre la responsabilidad intelectual sobre el contenido del artículo. Las herramientas de inteligencia artificial no fueron utilizadas como autoras del manuscrito ni sustituyen la responsabilidad académica de los investigadores.
Disponibilidad de datos Los datos que respaldan los resultados de esta investigación estarán disponibles previa solicitud razonable al autor de correspondencia, respetando las normas éticas y de confidencialidad establecidas por la investigación.

