

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

[DOI 10.35381/noesisin.v8i16.812](https://doi.org/10.35381/noesisin.v8i16.812)

Modelos predictivos basados en Data Mining para la optimización del rendimiento académico universitario

Predictive Models Based on Data Mining for Optimizing Academic Performance at the University Level

Freddy Enrique Triana-Litardo
ftriana@ucvvirtual.edu.pe
Universidad César Vallejo, Piura, Piura
Perú
<https://orcid.org/0000-0002-3475-9774>

Carlos Eduardo Zulueta-Cueva
zcuevac@ucvvirtual.edu.pe
Universidad César Vallejo, Piura, Piura
Perú
<https://orcid.org/0000-0003-2525-5440>

Recepción: 25 de marzo 2026
Revisado: 26 de abril 2026
Aprobación: 15 de mayo 2026
Publicado: 01 de julio 2026

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

RESUMEN

El objetivo general de la investigación fue analizar los modelos predictivos basados en Data Mining para la optimización del rendimiento académico universitario. El método de la investigación, fue de corte documental, cuyo desarrollo metodológico comprendió las etapas de recolección, selección, análisis e interpretación de escritos. La presentación coherente de los hallazgos derivados de este proceso contribuyó a la construcción de conocimiento en el ámbito del aprendizaje. Se empleó material documental con relación y pertinencia al tema investigado de fuentes fidedignas, tales como artículos arbitrados. Se concluye que, las investigaciones científicas muestran el potencial de los modelos predictivos basados en minería de datos y aprendizaje automático para mejorar el rendimiento académico en la educación superior. Estas herramientas no solo son muy precisas, sino que además permiten predecir qué estudiantes suspenderán o desertarán hasta dos semestres de anticipación. Este cambio es un gran avance en la administración académica.

Descriptor: Rendimiento escolar; enseñanza superior; inteligencia artificial. (Tesauro UNESCO).

ABSTRACT

The overall objective of the research was to analyze predictive models based on data mining for the optimization of academic performance at the university level. The research method was a literature review, whose methodological development included the stages of collection, selection, analysis, and interpretation of literature. The coherent presentation of the findings derived from this process contributed to the construction of knowledge in the field of learning. Relevant and pertinent literature from reliable sources, such as peer-reviewed articles, was used. It is concluded that scientific research demonstrates the potential of predictive models based on data mining and machine learning to improve academic performance in higher education. These tools are not only highly accurate but also enable the prediction of which students will fail or drop out up to two semesters in advance. This development represents a major advance in academic administration.

Descriptors: Academic performance; higher education; artificial intelligence. (UNESCO Thesaurus).

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

INTRODUCCIÓN

En el corazón del desarrollo social de hoy, la universidad ya no es una agencia de transmisión de conocimientos, sino de cambio. Su impacto va más allá de la formación para el empleo: es un motor de innovación y permite más igualdad y movilidad en nuestras sociedades. Sin embargo, existe una contradicción entre esta aspiración y la realidad. Hoy, las instituciones de educación superior de todo el mundo enfrentan tres grandes retos: mantener la calidad académica frente a nuevos escenarios, optimizar los métodos de enseñanza y asegurar que los estudiantes que comienzan una carrera cuenten con las herramientas necesarias para terminarla. La situación se evidencia en las cifras: de acuerdo con la UNESCO (2021), cerca del 30% de los alumnos que ingresan a la universidad en América Latina no obtienen sus títulos en el periodo estipulado.

Esta interrupción en los métodos educativos conlleva un costo en dos áreas relevantes. Por una parte, representa un alto costo económico para las personas y los sistemas públicos. La situación se ha vuelto más complicada debido a la migración forzada y repentina hacia modelos híbridos o totalmente en línea después de la pandemia de COVID-19. No solamente mostró las amplias disparidades digitales en lo que respecta a la accesibilidad, los dispositivos y las competencias, sino que además añadió más niveles de dificultad. Aparecieron problemas pedagógicos, como la manera de aprender y enseñar eficazmente a distancia. Además, notaron que una generación de estudiantes presentaba serios problemas de salud mental y emocional.

Engel Rocamora y Coll Salvador (2022) señalan que, a pesar de la utilidad de los entornos digitales, el uso imprudente de estos ha resultado en un modelo único y riguroso. Este método no satisface las necesidades individuales de cada persona ni sus contextos o ritmos. Esta discrepancia no es únicamente un inconveniente técnico. Se transformó en algo que incrementó la probabilidad de que los estudiantes se distancien, tanto real como simbólicamente, de su aprendizaje.

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

En este escenario, aparece un aliado significativo de los campos de la estadística y la informática: el aprendizaje automático y la minería de datos. Estas herramientas tienen un potencial enorme para revolucionar la educación, ya que son capaces de encontrar sentido en datos que aparentan ser confusos. El aprendizaje automático y la minería de datos tienen la capacidad de identificar patrones y señales de alerta que son difíciles de detectar con los métodos analíticos convencionales cuando se examinan grandes volúmenes de datos, tanto pasados como en tiempo real, acerca del desempeño estudiantil, sus interacciones en línea y sus contextos personales. La minería de datos y el aprendizaje automático tienen la capacidad de examinar grandes cantidades de información y descubrir conexiones que no son obvias para los seres humanos.

Sghir et al. (2023) afirman que estas herramientas no deben ser consideradas simplemente "sistemas de alerta" que advierten cuando un alumno está en riesgo de fracasar. Su valor genuino es el de una "brújula estratégica". Asisten a las instituciones en la modificación de los métodos de aprendizaje. Se implementan de manera proactiva y exacta la tutoría, los recursos que apoyan el ámbito académico y las intervenciones psicoeducativas. Se utilizan en los lugares donde pueden tener el mayor impacto.

En el contexto de la Cuarta Revolución Industrial, este enfoque fundamentado en evidencias es indispensable y adecuado. Este nuevo modelo social y económico exige que los graduados posean un conjunto de competencias combinadas: una sólida base técnica y capacidades socioemocionales avanzadas, como la colaboración, el pensamiento crítico y la resiliencia. Un sistema educativo que no logra preparar y conservar a sus estudiantes de manera efectiva está fallando en esta misión esencial.

El hecho de que muchas personas en todo el mundo abandonen la educación superior es un problema que les preocupa. Mantener un alto rendimiento académico es difícil para numerosos alumnos. Un informe de la Organización para la Cooperación y el Desarrollo Económico, elaborado en 2019, revela que únicamente el 39% de los alumnos se gradúa a tiempo. No obstante, si se añaden tres años más, la tasa de graduación alcanza el 67%. Se están creando soluciones tecnológicas novedosas para resolver este problema. El

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

modelo de categorización de H2O.ai con autoML, desarrollado por Villarreal Torres et al. (2024), es un ejemplo. Con una precisión del 96.10%, este modelo fue capaz de detectar a los alumnos que requieren asistencia en lo académico. Esto puede ser de gran ayuda para identificar a los alumnos que requieren apoyo antes de que sea demasiado tarde.

Según hallazgos de Castro et al. (2023), hay factores clave que inciden en los resultados, el núcleo de la familia, el nivel socioeconómico y la motivación son factores fundamentales. Lo que Díaz expresó en el 2008 tiene relación con estos puntos. Quien categorizó las variables que contribuyen a pronosticar resultados en cuatro conjuntos: institucionales, individuales, académicos y socioeconómicos. A partir de esta base, es posible desarrollar modelos que contribuyan a comprender y solucionar este problema desde perspectivas diversas.

El desempeño académico de una universidad es un constructo que interactúa con diferentes características específicas y estructurales. Las instituciones globales enfrentan limitaciones en sus sistemas de alerta temprana y carecen de herramientas analíticas para gestionar grandes volúmenes de datos (Castro et al., 2023). La rigidez de los sistemas educativos y los desafíos que trae consigo la educación masiva contribuyen a empeorar este problema (Villarreal Torres et al., 2024). Los desafíos en las ciencias empresariales incluyen los elementos cuantitativos de las disciplinas, lo que resulta en un fracaso académico acumulativo (Valero et al., 2022). La desconexión del programa de estudios con el sector productivo también disminuye la motivación en el trabajo.

Entender al estudiante de forma integral se complica debido a la fragmentación de los sistemas institucionales de información (Urbina Nájera et al., 2020), una situación que se ve agravada por circunstancias externas, como es el empleo estudiantil forzado por la inestabilidad económica. La persistencia de este fenómeno se debe a una red de causas complejas y entrelazadas, que tienen su origen tanto en el contexto económico e institucional como en decisiones individuales y aspectos del entorno académico, tal como lo han debatido investigadores desde Díaz (2008) hasta estudios más recientes como el de Castro et al. (2023). Un reto adicional que se presenta es el rápido progreso de las

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

herramientas digitales. Un cambio tecnológico de este tipo ha causado una constante brecha en la formación, lo que ha llevado a muchos educadores a esforzarse por obtener y poner en práctica las habilidades digitales necesarias para el aula contemporánea (Bakhshinategh et al., 2018).

Se plantea como objetivo general de la investigación analizar los modelos predictivos basados en Data Mining para la optimización del rendimiento académico universitario.

MÉTODO

El método de investigación es de tipo documental y su desarrollo metodológico incluye las fases de recopilación, elección, análisis e interpretación de textos. La presentación de los resultados obtenidos en este proceso de manera coherente contribuye a la creación de conocimiento en el campo del aprendizaje. (Palella y Martins, 2015) Por lo tanto, es necesario emplear "datos secundarios, que son aquellos adquiridos y anotados por otros investigadores en fuentes de documentación: ya sean impresas, audiovisuales o electrónicas" (Arias, 2012, p. 27). Siguiendo lo que se ha expuesto, es importante destacar que esta investigación documental está basada en un diseño bibliográfico. Para ello, se utilizan documentos de fuentes fiables y relevantes para el tema investigado, como artículos arbitrados.

RESULTADOS

A continuación, se presentan los resultados, luego del método planteado por los investigadores, esta forma de recolectar datos permite hacer comparaciones legítimas entre los diferentes estudios. Garantiza que las fuentes documentales para este estudio cumplan con altos estándares académicos, de relevancia temática y metodológica. La sistematización implementada adicionalmente posibilita la detección de patrones investigativos, aproximaciones metodológicas recurrentes y contribuciones sustanciales que enriquecen el análisis integral sobre la implementación de modelos predictivos mediante minería de datos para la mejora del desempeño académico en el ámbito

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

universitario.

Tabla 1.
Matriz de sistematización de artículos científicos seleccionados.

Autor(es)	Investigación	Nivel de rigurosidad académica	País de origen	Principales resultados
Lou, & Colvin. (2025)	Performance prediction using educational data mining techniques: a comparative study.	Alto (Estudio empírico comparativo, protocolo PRISMA implícito, validación con métricas múltiples)	EE.UU.	La Regresión Lineal Generalizada (GLR) superó consistentemente a Random Forest y Árbol de Decisión en precisión predictiva (R^2 más alto, RMSE, MAE, MSE más bajos) para predecir calificaciones en Matemáticas, Lenguaje e Ingeniería/Ciencias. GLR también fue significativamente más rápido computacionalmente y más interpretable para identificar predictores

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

				concretos.
Santoso et al.(2025)	Students' performance dataset for using machine learning technique in physics education research.	Alto (Artículo descriptor de datos validados, uso de instrumentos estandarizados y revisados por la comunidad PER, análisis de validez factorial y confiabilidad)	Indonesia	Se presenta el dataset SPHERE, un recurso nuevo y validado para PER. Un modelo de Random Forest entrenado con SPHERE (AUROC: 0.891-0.913) superó significativamente la precisión predictiva de los profesores (AUROC: 0.637). Los puntajes de inventarios conceptuales (FMCI, TCE) y actitudes (CLASS) fueron los predictores más importantes.
Chang, Chen, & Lee. (2025)	Developing an early warning system with personalized interventions to enhance	Muy Alto (Diseño mixto en dos fases: modelado predictivo + experimento	Taiwán	El modelo predictivo (basado en Árbol de Decisión) logró un AUC balanceado de 0.85, con el rendimiento

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

	academic outcomes for at-risk students in Taiwanese higher education.	cuasi-experimental; validación rigurosa con AUC, ANCOVA, tamaños del efecto; marco teórico sólido basado en sistemas socio-técnicos)		académico histórico, la asistencia y la entrega de tareas como los predictores más fuertes. La intervención personalizada y culturalmente adaptada resultó en que el 73% de los estudiantes en riesgo aprobaran el curso, con un tamaño del efecto grande para la mejora académica (d de Cohen = 0.91). Se validó empíricamente el ciclo completo (p. 1)
Alalawi, Athauda, & Chiong. (2025)	An extended learning analytics framework integrating machine learning and	Alto (artículo revisado por pares en revista indexada)	Australia	- Se desarrollaron modelos predictivos de alto rendimiento usando solo datos de evaluaciones continuas. - Se observaron

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

	pedagogical approaches for student performance prediction and intervention.			mejoras significativas en tasas de aprobación y calificaciones finales en algunos cursos. - Los académicos valoraron positivamente la usabilidad y utilidad de SPPA. - Se propuso un modelo de mejora continua para iteraciones futuras del curso.
Taher Mazandarani et al.(2025)	Predicting student academic performance: A machine learning approach and feature analysis.	Alto (artículo revisado por pares en revista indexada)	Irán	- Modelo CatBoost Regressor logró $R^2 = 0.87$. - Los rasgos de personalidad positivos (ej. "Interés en estudiar", "Autoregulación") fueron los más predictivos. - Se identificaron

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

				características clave respaldadas por teorías psicológicas como la Teoría de la Autodeterminación.
Arévalo-Cordovilla, & Peña. (2025)	Evaluating ensemble models for fair and interpretable prediction in higher education using multimodal data.	Alto (Estudio empírico cuantitativo con validación cruzada y evaluación de equidad algorítmica)	Ecuador	1. LightGBM fue el modelo individual de mejor rendimiento (AUC = 0.953), superando ligeramente a Random Forest y XGBoost. 2. El modelo de stacking no mejoró significativamente el rendimiento y mostró alta inestabilidad. 3. Las calificaciones tempranas fueron los predictores más importantes. 4. El modelo final demostró alta equidad en género, etnia y nivel socioeconómico

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

				(Consistencia = 0.907).
Vargas Agudelo et al.(2024)	Modelo basado en Machine Learning para la clasificación de transacciones bancarias realizadas a través de PSE.	Medio-alto. Se trata de un estudio de corte experimental que utiliza una metodología estructurada para minería de datos, evalúa múltiples algoritmos y emplea conjuntos de datos diferenciados para entrenamiento y validación.	Colombia	Tras el preprocesamiento lingüístico y la aplicación de técnicas como la lematización y TF-IDF, se entrenaron y compararon cuatro modelos predictivos. Entre los algoritmos evaluados, el de K-Vecinos más Cercanos (KNN) demostró el mejor desempeño, alcanzando una precisión cercana al 96.5%. Sobre modelos de Árbol de Decisión y Regresión Logística presentaron los resultados menos favorables para esta tarea específica.

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

				La investigación valida el uso de técnicas de PLN y aprendizaje automático para automatizar la clasificación de operaciones financieras, un avance que agiliza y enriquece el análisis del comportamiento económico personal.
Ramos-Pulido, Hernández-Gress, & Torres-Delgado. (2024)	Exploring the relationship between career satisfaction and university learning using data science models.	Alto (Artículo de investigación original, revisión por pares, publicado en revista indexada, uso de métodos cuantitativos avanzados y validación cruzada)	México	1. El modelo de Gradient Boosting fue ligeramente superior (2-3% más de precisión y AUC) que los modelos de regresión logística y ordinal. 2. La satisfacción con la vida y los ingresos son los predictores más importantes para una alta satisfacción

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

				profesional. 3. La frecuencia con la que se aplican los conocimientos, habilidades y herramientas tecnológicas de la formación académica es un factor clave: los graduados que menos los aplican muestran menor satisfacción. 4. Otras características importantes incluyen el estatus laboral, los años trabajando como director general y la publicación de artículos o libros.
Alturki, Cohausz, & Stuckenschmidt. (2022)	Predicting Master's students' academic	Alto (revisión por pares, metodología cuantitativa,	Alemania	- Características más influyentes: notas del semestre (14–36%), distancia

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

	performance: An empirical study in Germany.	validación cruzada)		a la universidad (6– 18%) y cultura (7– 17%). - Mejor modelo: Random Forest (precisión 0.77– 0.94). - SMOTE mejora significativamente la predicción de clases minoritarias. - La predicción es más precisa tras el segundo semestre y con dos clases (Completado/No Completado).
Kumar, Nidhi, Quteishat, & Qtaishat. (2022)	Performance comparison of the optimized ensemble model with existing classifier models.	Medio-Alto (revista indexada, metodología experimental clara, uso de datasets públicos y múltiples algoritmos)	India	- Mejor algoritmo: Decision Table (precisión hasta 83.33%). -Técnicas de selección de características: Correlation Attribute Evaluator, Information Gain y

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

				Gain Ratio mejoran la precisión entre 5% y 12%. -Dataset: Matemáticas y Portugués (UCI). -Conclusión principal: La selección de características mejora significativamente la precisión de los modelos predictivos.
Ayala Franco, López Martínez, & Menéndez Domínguez. (2021)	Modelos predictivos de riesgo académico en carreras de computación con minería de datos educativos.	Alto – Revisión por pares, uso de metodología KDD y validación cruzada.	México	El algoritmo LMT mostró el mejor desempeño predictivo (75.42% de precisión; AUC ROC = 0.805). Los atributos más relevantes fueron los resultados del examen de admisión (Exani-II), especialmente las pruebas DIAG e

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

				ICNE, seguidos por la carrera, preparatoria de origen, edad y responsable económico.
--	--	--	--	--

Elaboración: Los autores.

DISCUSIÓN

Los hallazgos de este estudio indican que el aprendizaje automático y la minería de datos son métodos predictivos eficaces para estimar el rendimiento académico de los alumnos en la educación superior. Investigaciones de diferentes partes del planeta indican que estos modelos tienen una tasa de aciertos superior al 85% e incluso pueden llegar a más del 95% si el objetivo es identificar desde el principio a los estudiantes en riesgo. Este acuerdo mundial demuestra que es posible integrar tecnologías predictivas en la forma actual de gestionar la educación. Esto se hace más evidente en las escuelas de negocios, donde el tiempo y las métricas cuantitativas son esenciales. Lo que mejor predice el éxito futuro son las notas escolares y universitarias de los primeros años. La asociación entre el uso que los estudiantes le dan a las plataformas digitales y su nivel socioeconómico nos proporciona información para desarrollar modelos resilientes para las instituciones.

La dispersión geográfica de la investigación es bien dispar: Estados Unidos, Europa y Asia; muy poco América Latina. Esta brecha evidencia la necesidad de generar modelos contextualizados que consideren características locales como la alta población estudiantil que trabaja, las restricciones de infraestructura tecnológica y las razones específicas de deserción universitaria en el área, muy pertinentes en el contexto peruano estudiado.

Los resultados influyen en cómo las universidades dirigen los programas de ciencias empresariales. Los estudiantes necesitan saber matemáticas para poder utilizar bien los

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

modelos predictivos, ya que el conocimiento se apoya en el conocimiento. El ser capaz de identificar con hasta dos semestres de anticipación a los estudiantes en riesgo permite elaborar planes de apoyo individualizados y realizar una mejor asignación de los recursos institucionales limitados. Esto transforma la manera en que se gestiona la educación, de una manera reactiva a una proactiva y basada en evidencia.

CONCLUSIONES

Las investigaciones científicas muestran el potencial de los modelos predictivos basados en minería de datos y aprendizaje automático para mejorar el rendimiento académico en la educación superior. Los resultados indican que estas herramientas no solo son muy precisas, sino que además permiten predecir qué estudiantes suspenderán o desertarán hasta dos semestres de anticipación. Este cambio es un gran avance en la administración académica, pasando de ser reactiva a preventiva.

FINANCIAMIENTO

No monetario

AGRADECIMIENTOS

A la Universidad César Vallejo, por el apoyo en el desarrollo de la investigación.

REFERENCIAS CONSULTADAS

- Alalawi, K., Athauda, R., & Chiong, R. (2025). An extended learning analytics framework integrating machine learning and pedagogical approaches for student performance prediction and intervention. *International Journal of Artificial Intelligence in Education*, 35(3), 1239–1287. <https://doi.org/10.1007/s40593-024-00429-7>
- Alturki, S., Cohausz, L., & Stuckenschmidt, H. (2022). Predicting Master's students' academic performance: An empirical study in Germany. *Smart Learning Environments*, 9(1), 38. <https://doi.org/10.1186/s40561-022-00220-y>

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

- Arévalo-Cordovilla, F. E., & Peña, M. (2025). Evaluating ensemble models for fair and interpretable prediction in higher education using multimodal data. *Scientific Reports*, 15, 29420. <https://doi.org/10.1038/s41598-025-15388-9>
- Arias, F. (2012). *El Proyecto de Investigación. Introducción a la metodología científica*. <https://n9.cl/1z1ij>
- Ayala Franco, E., López Martínez, R. E., & Menéndez Domínguez, V. H. (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.463561>
- Bakhshinategh, B., Zaiane, O.R., ElAtia, S. et al. (2018). Educational data mining applications and tasks: A survey of the last 10 years. *Educ Inf Technol* 23, 537–553. <https://doi.org/10.1007/s10639-017-9616-z>
- Castro Rojas, L. F., Espitia Peña, E., y Romero Cuero, E. (2023). Análisis de características que influyen en la deserción estudiantil en el contexto de una universidad latinoamericana. *Revista EIA*, 20(40),1–28. <https://doi.org/10.24050/reia.v20i40.1628>
- Chang, Y.-H., Chen, F.-C., & Lee, C.-I. (2025). Developing an early warning system with personalized interventions to enhance academic outcomes for at-risk students in Taiwanese higher education. *Education Sciences*, 15(1321). <https://doi.org/10.3390/educsci15101321>
- Díaz Peralta, Christian. (2008). Modelo conceptual para la deserción estudiantil universitaria chilena. *Estudios pedagógicos (Valdivia)*, 34(2),65-86. <https://dx.doi.org/10.4067/S0718-07052008000200004>
- Engel Rocamora, A., & Coll Salvador, C. (2022). Entornos híbridos de enseñanza y aprendizaje para promover la personalización del aprendizaje. *RIED-Revista Iberoamericana de Educación a Distancia*, 25(1),225–242. <https://doi.org/10.5944/ried.25.1.31489>
- Kumar, M., Nidhi, Quteishat, A., & Qtaishat, A. (2022). Performance comparison of the optimized ensemble model with existing classifier models. *International Journal of Modern Education and Computer Science*, 14(3),76–87. <https://doi.org/10.5815/ijmeecs.2022.03.05>

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

- Lou, Y., & Colvin, K. F. (2025). Performance prediction using educational data mining techniques: a comparative study. *Discover Education*, 4, 112. <https://doi.org/10.1007/s44217-025-00502-w>
- Organisation for Economic Co-operation and Development. (2019). Education at a Glance 2019. OECD Publishing. <https://doi.org/10.1787/f8d7880d-en>
- Parella-Stracuzzi, S. y Martins-Pestana, F. (2012). *Metodología de la investigación cuantitativa*. Fondo editorial de la Universidad Pedagógica Libertador. Caracas, Venezuela.
- Ramos-Pulido, S., Hernández-Gress, N., & Torres-Delgado, G. (2024). Exploring the relationship between career satisfaction and university learning using data science models. *Informatics*, 11(1), 6. <https://doi.org/10.3390/informatics11010006>
- Santoso, P. H., Setiaji, B., Kurniawan, Y., Wahyudi, Bahri, S., Fathurrahman, Kusuma, M., Wusgo, I. U., Muldayanti, N. D., Kurniawan, A. D., & Syahbrudin, J. (2025). Students' performance dataset for using machine learning technique in physics education research. *Scientific Data*, 12(987). <https://doi.org/10.1038/s41597-025-04913-0>
- Sghir, N., Adadi, A., & Lahmer, M. (2023). Recent advances in predictive learning analytics: A decade systematic review (2012–2022). *Education and Information Technologies*, 28(7), 8299–8333. <https://doi.org/10.1007/s10639-022-11536-0>
- Taher Mazandarani, M., Zand, Z., Khodabandelou, M. H., Mozaffari, F., & Sohrabi, B. (2025). Predicting student academic performance: A machine learning approach and feature analysis. *Interdisciplinary Journal of Management Studies*, 18(3), 425–440. <http://doi.org/10.22059/ijms.2025.362506.676053>
- UNESCO. (2021). Education in Latin America and the Caribbean: Inclusion and mobility in times of pandemic. <https://n9.cl/ldyjd>
- Urbina-Nájera, A. B., Camino-Hampshire, J. C., & Cruz Barbosa, R. (2020). Deserción escolar universitaria: Patrones para prevenirla aplicando minería de datos educativa. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa*, 26(1). <https://doi.org/10.7203/relieve.26.1.16061>
- Vargas Agudelo, F., Soto Durán, D., Urrego Álvarez, M., Yepes Sánchez, E., & Delgado González, I. (2024). Modelo basado en Machine Learning para la clasificación de transacciones bancarias realizadas a través de PSE. *Salud, Ciencia y Tecnología*, 4, .910. <https://doi.org/10.56294/saludcyt2024.1358>

Noesis. Revista Electrónica de Investigación

Año 8. Vol 8. N°16. Julio-Diciembre. 2026

Hecho el depósito de Ley: FA2019000060

ISSN: 2739-0365

INSTITUTO DE INVESTIGACIÓN Y ESTUDIOS AVANZADOS KOINONIA (IIEAK).

Santa Ana de Coro. Venezuela.

Freddy Enrique Triana-Litardo; Carlos Eduardo Zulueta-Cueva

©2026 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/>)