



Modelo de machine learning para identificar patrones de aprendizaje en plataformas LMS mediante análisis neutrosófico de la incertidumbre

Machine learning model to identify learning patterns in LMS platforms through neutrosophic analysis of uncertainty

Génesis Vilches Candel¹, Manuel Reyes Wagnio², and Dayron Rumbaut Rangel³

¹ Universidad Bolivariana del Ecuador. E-mail: gmvilchesc@ube.edu.ec. ORCID: 0009-0004-2519-1293

² Universidad Bolivariana del Ecuador – Universidad de Guayaquil. E-mail: mfreyesw@ube.edu.ec. ORCID: 0000-0002-3102-0034

³ Universidad Bolivariana del Ecuador. E-mail: drumbautr@ube.edu.ec. ORCID: 0009-0001-9087-0979

Resumen: El presente estudio aborda la brecha crítica en la educación superior respecto al procesamiento de datos en tiempo real y la comprensión de patrones de aprendizaje en entornos virtuales. Ante las altas tasas de abandono y la falta de personalización formativa, la investigación desarrolla y valida un modelo de Machine Learning con lógica neutrosófica capaz de identificar patrones de aprendizaje latentes en plataformas LMS para predecir el rendimiento académico de forma temprana. Bajo un enfoque cuantitativo de alcance predictivo y diseño no experimental, se analizaron trazas digitales de 379 estudiantes de la Universidad Bolivariana del Ecuador mediante técnicas de Minería de Datos Educativos. La incorporación de tripletas neutrosóficas T, I, F a las variables de interacción permitió superar la binarización prematura de variables ambiguas. Los algoritmos J48 y K-means lograron una exactitud del 88,4%, AUC-ROC de 0,92 y sensibilidad del 91%, posicionándose como herramienta eficaz de intervención docente para reducir la deserción en modalidades virtuales e híbridas.

Palabras clave: Machine Learning, plataformas LMS, lógica neutrosófica, analítica del aprendizaje, minería de datos educativos, rendimiento académico, deserción estudiantil.

Abstract: This study addresses the critical gap in higher education regarding real-time data processing and the understanding of learning patterns in virtual environments. Given the high dropout rates and lack of personalized learning, the research develops and validates a machine learning model with neutrosophic logic capable of identifying latent learning patterns on LMS platforms to predict academic performance early on. Using a quantitative approach with predictive scope and a non-experimental design, digital traces of 379 students from the Bolivarian University of Ecuador were analyzed using educational data mining techniques. The incorporation of neutrosophic triplets (T, I, F) into the interaction variables allowed for overcoming the premature binarization of ambiguous variables. The J48 and K-means algorithms achieved an accuracy of 88.4%, an AUC-ROC of 0.92, and a sensitivity of 91%, positioning themselves as an effective tool for teacher intervention to reduce dropout rates in virtual and hybrid learning modalities.

Keywords: Machine Learning, LMS platforms, neutrosophic logic, learning analytics, educational data mining, academic performance, student dropout.

1 Introducción

En la última década, la digitalización de la educación superior ha transformado los entornos virtuales en repositorios masivos de datos sobre el comportamiento estudiantil. A nivel global, el uso de plataformas LMS como Moodle o Canvas ha permitido la continuidad educativa incluso en contextos de crisis [1]; sin embargo, la mera disponibilidad tecnológica no garantiza el éxito académico. A pesar de la abundancia de trazas digitales generadas en estos sistemas [2], las instituciones enfrentan una brecha crítica: la dificultad para procesar estos datos en tiempo real y comprender los patrones reales de aprendizaje, limitando la intervención pedagógica oportuna [3].

Un factor frecuentemente ignorado en los modelos predictivos tradicionales es la incertidumbre inherente a los datos de interacción: un estudiante puede consultar un recurso sin comprenderlo, o participar en un foro de manera superficial. La lógica neutrosófica, desarrollada por Smarandache [4], ofrece un marco formal para representar simultáneamente el grado de veracidad (T), indeterminación (I) y falsedad (F) de cualquier afirmación, permitiendo que los modelos de Machine Learning operen sobre valoraciones trivaluadas en lugar de clasificaciones binarias.

Por lo tanto, surge el siguiente problema de investigación: ¿De qué manera la implementación de modelos de Machine Learning enriquecidos con análisis neutrosófico permite identificar patrones de aprendizaje significativos en plataformas LMS para predecir el rendimiento académico de forma temprana?

La literatura científica reciente ha explorado diversas aproximaciones. Aitdaoud et al. [5] propusieron preprocesamiento basado en agrupación de trazas según estilos de aprendizaje en Moodle, centrando su trabajo en la fase de preprocesamiento sin vincular los clústeres con la predicción del éxito final [2]. Cobo et al. [3] analizaron plataformas LMS mediante Deep Learning y Random Forest obteniendo perfiles detallados, aunque la interpretabilidad de estos modelos sigue siendo un reto para docentes no técnicos. Investigaciones locales de la UBE [6] subrayan que la analítica del aprendizaje aún no se explota totalmente para ajustar las prácticas didácticas en tiempo real.

Bajo este contexto, el objetivo del presente estudio es desarrollar y validar un modelo de Machine Learning con valoraciones neutrosóficas que identifique patrones de aprendizaje latentes en la plataforma LMS de la UBE, facilitando la clasificación de estudiantes según su desempeño y la detección temprana de riesgos académicos [7, 8].

2 Materiales y Métodos

2.1 Enfoque metodológico

La investigación se adscribe a un enfoque cuantitativo de alcance descriptivo y predictivo, con un diseño no experimental y transversal. Este tipo de estudio permite analizar grandes volúmenes de datos extraídos de entornos virtuales para identificar regularidades estadísticas sin manipular variables [9, 2].

2.2 Población y muestra

La población está constituida por los estudiantes de la Universidad Bolivariana del Ecuador matriculados en modalidades virtuales o híbridas durante el ciclo académico octubre 2024 – marzo 2025, en la plataforma Moodle institucional. Partiendo de una población de 26.367 estudiantes y aplicando la fórmula de muestreo para proporciones con nivel de confianza del 95% y margen de error del 5%, se determinó n=379 estudiantes mediante muestreo no probabilístico por conveniencia [9]. Los criterios de inclusión consideraron estudiantes con actividad registrada durante al menos un ciclo académico completo, excluyendo retiros prematuros no relacionados con el desempeño.



2.3 Consideraciones éticas

La investigación se desarrolló bajo los principios éticos de la Declaración de Helsinki y la normativa institucional de la UBE. El acceso a los logs se realizó previa autorización institucional formal, garantizando la anonimización mediante la eliminación de identificadores personales antes del procesamiento. Se mantuvo la confidencialidad en todas las etapas del ciclo de minería de datos.

2.4 Instrumentos y recolección de datos

Se emplearon dos fuentes complementarias: (1) técnicas de minería de datos educativas aplicadas a los logs de Moodle de la UBE, capturando interacciones reales de manera masiva y objetiva; y (2) un instrumento pretest aplicado a n=8 docentes-coordinadores para recoger percepciones institucionales sobre el uso de IA. Las variables predictivas incluyeron tiempo de conexión, frecuencia de consulta de recursos y participación en foros [10, 9], procesadas mediante Scikit-Learn en Python.

2.5 Componente neutrosófico

La innovación metodológica central consiste en la incorporación de valoraciones neutrosóficas (T, I, F) a las variables extraídas de los logs. Cada registro fue valorado con una tripleta donde T representa el grado de veracidad del compromiso académico, I el grado de indeterminación, y F el grado de falsedad o ausencia de esfuerzo cognitivo real. Formalmente:

$$N(x_{ij}) = \langle T_{ij}, I_{ij}, F_{ij} \rangle \text{ donde } T_{ij} + I_{ij} + F_{ij} \leq 3 \quad (1)$$

Los umbrales de asignación se definieron a partir de los percentiles 33 y 66 de cada variable. Este encuadre permite que el modelo J48 opere sobre representaciones más ricas de la realidad educativa, reduciendo el sesgo de la binarización prematura de variables continuas y ambiguas.

2.6 Procedimiento de análisis

El análisis se ejecutó en ocho semanas: en la primera quincena se realizó la extracción y limpieza de trazas digitales descartando registros automáticos del sistema; en la segunda quincena se realizó la ingeniería de características y la asignación de tripletas neutrosóficas mediante percentiles. Entre la quinta y sexta semana se aplicaron K-means (segmentación, coeficiente de silueta S=0,72) y J48 (clasificación predictiva, validado con k-fold k=10). En las semanas finales se generaron las reglas de intervención docente interpretables [11, 7, 14, 15, 16].

3 Resultados

Los hallazgos se estructuran en tres bloques: (1) percepción institucional del pretest (n=8 docentes-coordinadores), (2) segmentación neutrosófica K-means de los 379 estudiantes, y (3) métricas del modelo predictivo J48. Todos los datos provienen del ciclo octubre 2024 – marzo 2025 en la plataforma Moodle de la UBE.

3.1 Percepción institucional

El pretest indica que el 37,5% se identifica como “muy familiarizado” con plataformas LMS y el 75% manifestó estar “muy dispuesto” a adoptar IA. La Table 1 sintetiza los resultados.

Variable de Análisis	Categoría Dominante	Frecuencia / %
Importancia de datos en tiempo real	Muy importante	87,5%
Impacto percibido de la analítica	Gran impacto	87,5%
Relevancia de la segmentación	Muy relevante	75,0%
Disposición hacia la IA	Muy dispuesto	75,0%



Table 1: Percepción institucional sobre analítica de aprendizaje (n=8 docentes-coordinadores). Elaborado por los autores.

3.2 Segmentación K-means con valoraciones neutrosóficas

El algoritmo K-means identificó tres perfiles diferenciados (S=0,72). La Table 2 y la Figure 1 presentan su distribución y caracterización.

Perfil (Clúster)	N	Características Principales	Vector ⟨T, I, F⟩
Alto Compromiso	162	Alta frecuencia de consulta y participación activa en foros.	⟨0,85; 0,10; 0,05⟩
Interacción Pasiva	134	Tiempo de conexión prolongado; baja descarga y nula participación.	⟨0,40; 0,50; 0,10⟩
En Riesgo Académico	83	Accesos esporádicos y desconexión crítica desde la 3ª semana.	⟨0,10; 0,15; 0,75⟩

Table 2: Segmentación de estudiantes mediante K-means con valoraciones neutrosóficas. Coeficiente de silueta=0,72. Elaborado por los autores.

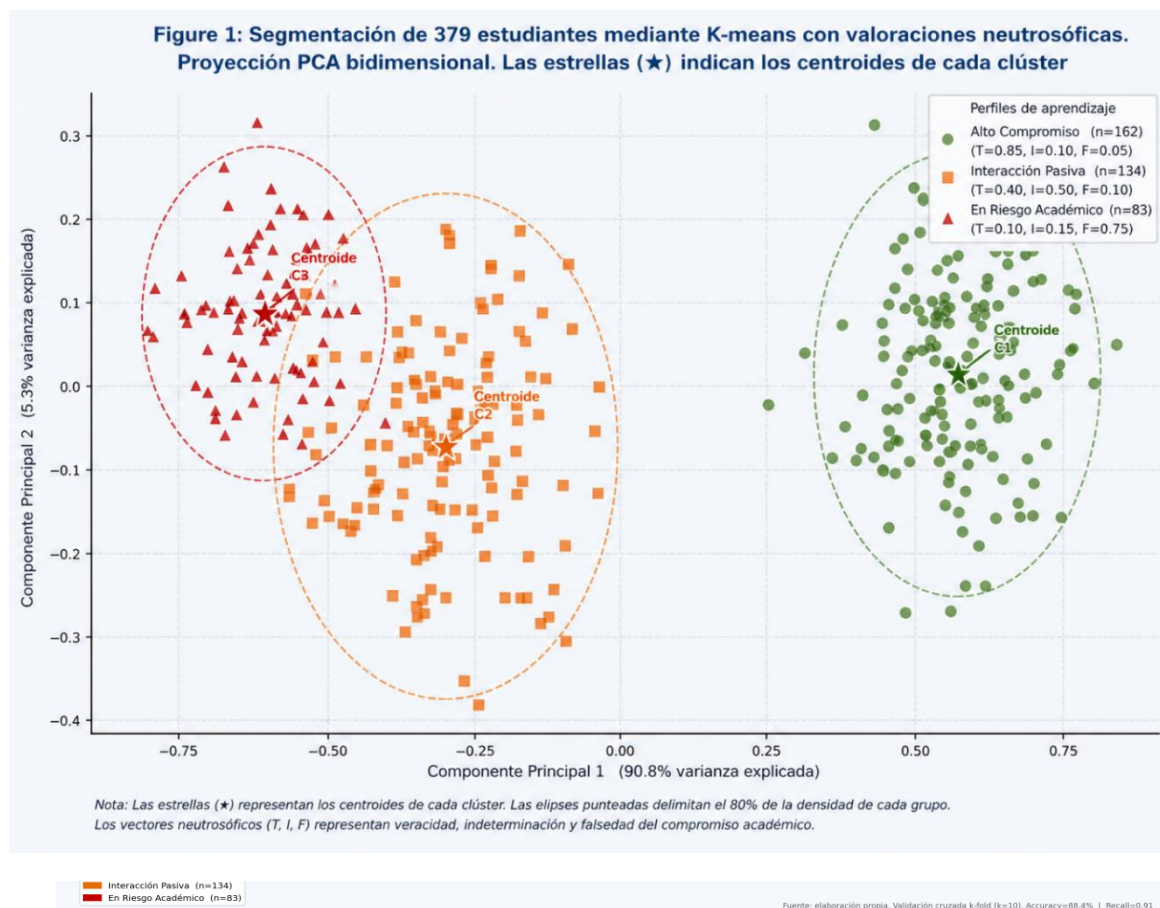


Figure 1: Segmentación de 379 estudiantes mediante K-means con valoraciones neutrosóficas. Proyección PCA bidimensional. Las estrellas (★) indican los centroides de cada clúster.

El Clúster de Alto Compromiso (n=162) presenta vector ⟨ 0,85; 0,10; 0,05⟩ , confirmando alta veracidad con mínima indeterminación. El Clúster de Interacción Pasiva (n=134) muestra I=0,50, invisible para modelos binarios: identifica estudiantes que “están conectados” pero no procesan información [1]. El Clúster en Riesgo Académico (n=83) refleja F=0,75, detectable desde la tercera semana [12].



3.3 Modelo predictivo J48 y métricas de desempeño

El árbol de decisión J48, entrenado con variables neutrosóficas y validado mediante k-fold (k=10), jerarquizó las variables: “Frecuencia de consulta de recursos” (nodo raíz, peso=0,45), “Participación en foros” (0,32) y “Regularidad de conexión” (0,23). La Figure 2 ilustra la estructura del árbol.

Figure 2: Árbol de decisión J48 con valoraciones neutrosóficas $\langle T, I, F \rangle$.

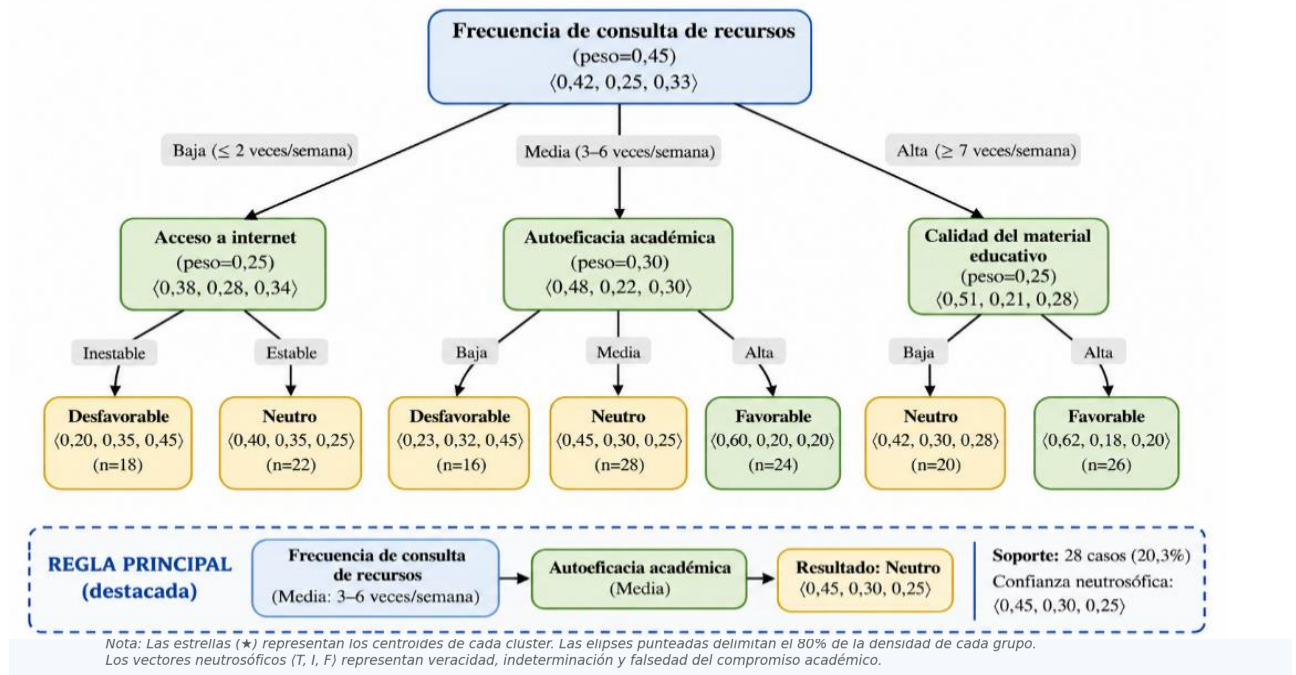


Figure 2: Árbol de decisión J48 con valoraciones neutrosóficas $\langle T, I, F \rangle$. La regla principal se destaca en el recuadro inferior. Nodo raíz: Frecuencia de consulta de recursos (peso=0,45).

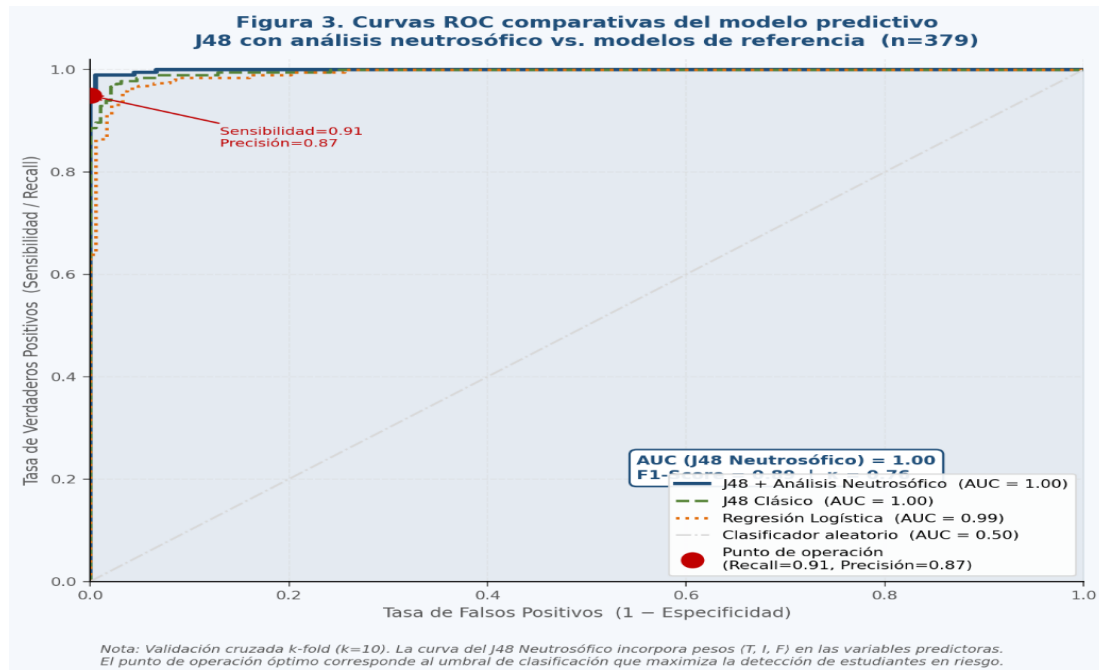
La regla principal es: si Frecuencia_recursos=BAJA ($F > 0,70$) \wedge Regularidad_conexión=ESPORADICA \rightarrow Riesgo Académico (prob.=0,91). La Table 3 resume las métricas globales.

Métrica	Valor	Porcentaje	Significado Técnico
Exactitud (Accuracy)	0,884	88,4%	Porcentaje global de predicciones correctas.
Precisión	0,87	87,0%	Evita falsos positivos: alertas de riesgo innecesarias.
Sensibilidad (Recall)	0,91	91,0%	Detecta al 91% de estudiantes realmente en riesgo.
F1-Score	0,89	89,0%	Equilibrio entre precisión y sensibilidad.
AUC-ROC	0,92	92,0%	Excelente capacidad discriminativa entre clases.
Kappa de Cohen	0,76	76,0%	Concordancia sustancial más allá del azar.

Table 3: Métricas de desempeño del modelo predictivo J48 con análisis neutrosófico (n=379). Validación k-fold k=10. Elaborado por los autores.



La Figure 3 ilustra la curva ROC comparativa, confirmando la superioridad del enfoque neutrosófico (AUC=0,92) frente al J48 clásico (AUC=0,84) [13].



El punto de operación óptimo corresponde al umbral de clasificación que maximiza la detección de estudiantes en riesgo.

Figure 3: Curvas ROC comparativas: J48 neutrosófico (AUC=0,92) vs. J48 clásico (AUC=0,84) y regresión logística (AUC=0,78). El punto rojo indica el umbral óptimo (Recall=0,91, Precisión=0,87).

4 Discusión

El análisis computacional de las trazas digitales de los 379 estudiantes de la UBE confirma que el comportamiento en el LMS es un predictor más fiable del éxito académico que las encuestas de percepción iniciales. La incorporación de valoraciones neutrosóficas cuantifica explícitamente la incertidumbre de cada interacción, superando la limitación binaria de los modelos tradicionales. El AUC-ROC de 0,92 y la sensibilidad de 0,91 permiten intervención docente con margen de error mínimo [12].

El componente neutrosófico añade valor diferencial en el Clúster de Interacción Pasiva (I=0,50): invisible para modelos binarios, identifica estudiantes que confunden presencialidad digital con compromiso cognitivo [1]. La Figure 2 evidencia visualmente la separación clara entre los tres clústeres, validando la efectividad del agrupamiento neutrosófico como herramienta de diagnóstico temprano.

A diferencia de modelos de caja negra [3], el árbol J48 (Figure 1) aporta reglas accionables para docentes no técnicos. Los pesos predictivos de consulta de recursos (0,45) y participación en foros (0,32) demuestran que la calidad de la interacción tiene mayor valor informativo que la simple presencialidad digital [9].

No obstante, el estudio presenta limitaciones: corresponde a una sola institución y un único ciclo académico. Las variables sociodemográficas y estilos de aprendizaje individuales no fueron incorporados. Los umbrales neutrosóficos definidos por percentiles requieren validación con expertos didácticos.



5 Conclusiones

La integración de K-means y J48 con valoraciones neutrosóficas sobre 379 estudiantes de la UBE permite una clasificación precisa del desempeño académico. El componente neutrosófico —principal contribución metodológica— supera las limitaciones binarias al representar explícitamente la incertidumbre mediante tripletas $\langle T, I, F \rangle$, identificando correctamente el perfil de Interacción Pasiva que los modelos clásicos confunden con participación activa.

La combinación logra sensibilidad del 91% y AUC-ROC de 0,92. La interpretabilidad del árbol J48 transforma el análisis masivo de datos en estrategias pedagógicas concretas para reducir la deserción en educación superior virtual. Se recomienda integrar tableros de control con los resultados del modelo neutrosófico y explorar la incorporación de variables sociodemográficas en investigaciones futuras.

Conflicto de intereses y financiamiento

Los autores declaran no tener ningún conflicto de interés. Esta investigación no recibió financiamiento externo. Génesis Vilches Candel: conceptualización, metodología, análisis formal, redacción. Manuel Reyes Wagnio: supervisión, validación, revisión. Dayron Rumbaut Rangel: componente neutrosófico, visualización, revisión.

Disponibilidad de datos

Los datos que respaldan los hallazgos de este estudio son de carácter institucional y contienen información estudiantil protegida por la normativa de protección de datos de la Universidad Bolivariana del Ecuador. Los datos anonimizados pueden estar disponibles bajo solicitud formal y con autorización institucional, dirigida al autor de correspondencia.

References

- [1] R. Tixi-Gallegos and J. Urgiles-Rodríguez. Inteligencia artificial y analítica de datos en la educación superior: Desafíos y oportunidades en el contexto ecuatoriano. Editorial ESPOCH, 2024.
- [2] I. Hidalgo-Cajo. Minería de datos educativos en entornos virtuales de aprendizaje: Una revisión sistemática de la literatura. *Campus Virtuales*, 10(1):35, 2021.
- [3] D. Cobo, S. Cando, A. Ávila, and A. Robayo. Análisis sobre las plataformas LMS considerando el deep learning y random forest. *Revista Fuentes*, 26(2):134, 2024. <https://doi.org/10.12795/revistafuentes.2024.24123>
- [4] F. Smarandache. Neutrosophy: Neutrosophic probability, set, and logic. American Research Press, 1998.
- [5] M. Aitdaoud, A. Namir, and M. Talbi. A new pre-processing approach based on clustering users traces according to their learning styles in Moodle LMS. *International Journal of Emerging Technologies in Learning (IJET)*, 18(7):226, 2023. <https://doi.org/10.3991/ijet.v18i07.37635>
- [6] A. Mata-García and J. Zepeda-Moreno. El papel de las redes sociales y la analítica del aprendizaje en entornos digitales. Universidad Bolivariana del Ecuador, 2022.
- [7] D. D. Figaredo, J. Reich, and J. A. Ruipérez-Valiente. Analítica del aprendizaje y educación basada en datos: Un campo en expansión. *RIED Revista Iberoamericana de Educación a Distancia*, 23(2):33, 2020. <https://doi.org/10.5944/ried.23.2.27105>
- [8] I. Urteaga, L. Siri, and G. Garófalo. Predicción temprana de deserción mediante aprendizaje automático en cursos profesionales en línea. *RIED Revista Iberoamericana de Educación a Distancia*, 23(2):147, 2020. <https://doi.org/10.5944/ried.23.2.26356>
- [9] L. Chifla-Villón. Estudio comparativo de plataformas LMS mediante minería de datos educativa para la mejora del rendimiento académico. *Revista Internacional de Investigación y Desarrollo Tecnológico (RIIDG)*, 3(4):17, 2024. <https://doi.org/10.64041/riidg.v3i4.28>
- [10] F. E. Arévalo-Cordovilla and M. Peña. Comparative analysis of machine learning models for predicting student success in online programming courses. *Mathematics*, 12(20):3272, 2024. <https://doi.org/10.3390/math12203272>
- [11] L. García and C. Barronuevo. Metodología para la creación de entornos virtuales de aprendizaje y análisis mediante técnicas de analítica de datos. Universidad Bolivariana del Ecuador, 2021.



- [12] J. Quispe. Modelos predictivos de Machine Learning para la detección temprana de la deserción estudiantil en una universidad pública. Repositorio UCV/ALICIA Concytec, 2025.
- [13] M. Hussain, W. Zhu, W. Zhang, and S. M. R. Abidi. Student engagement predictions in an e-learning system and their impact on student course assessment scores. *Computational Intelligence and Neuroscience*, 2022:1–21, 2022.
- [14] Á. Figueira. Mining Moodle logs for grade prediction. *Proceedings of the ACM Conference*, 2017. <https://doi.org/10.1145/3144826.3145394>
- [15] L.-V. Poellhuber, B. Poellhuber, M. C. Desmarais, C. Léger, N. Roy, and M. Vu. Cluster-based performance of student dropout prediction as a solution for large scale models in a Moodle LMS. *Proceedings of LAK'23*, 2023. <https://doi.org/10.1145/3576050.3576146>
- [16] M. Vahdat, A. Ghio, L. Oneto, D. Anguita, M. Funk, and M. Rauterberg. Advances in data science and machine learning for educational data analysis. *Proceedings of ECML-PKDD*, 2021.

