

Machine Learning como Herramienta Pedagógica para Optimizar la Comprensión Lectora en Estudiantes de Primaria

Machine Learning as a Pedagogical Tool to Optimize Reading Comprehension in Primary School Students

Carlos Antonio Valderrama-Arévalo* 

Facultad de Arquitectura y Urbanismo, Ingenierías, Arte & Diseño, Universidad Particular de Chiclayo, Prolongación Avenida Juan Tomis Stack N° 2777 – Ciudad Universitaria, Carretera Chiclayo - Pimentel, Perú.

* Autor correspondiente: cvarevalo0111@gmail.com (C. Valderrama-Arévalo) DOI: [10.17268/scien.inge.2026.02.03](https://doi.org/10.17268/scien.inge.2026.02.03)

RESUMEN

Este estudio analizó el impacto del Machine Learning en la comprensión lectora de estudiantes de sexto grado de primaria de una institución pública en Monsefú, durante 2024. Se implementaron tres modelos predictivos desarrollados con AutoML, XAI y DataOps, orientados a identificar patrones y brindar apoyo pedagógico personalizado en los niveles literal, inferencial y crítico. El diseño fue explicativo, con enfoque cuantitativo y diseño cuasi experimental. Se aplicaron pretest y posttest a dos grupos paralelos (experimental y control, n=40 cada uno), con instrumentos validados. En el grupo experimental, se observaron mejoras del 24,9% en la comprensión lectora (62,4% a 87,3%), mientras que, en sus dimensiones, del 20,5% en nivel literal (de 63,8% a 84,3%), del 24% en el nivel inferencial (de 62,1% a 86,1%) y del 30,3% en el nivel crítico (de 61,4% a 91,7%). Las pruebas de Wilcoxon y U de Man Whitney evidenciaron significancia estadística en los tres niveles evaluados ($p < 0,05$). El modelo predictivo 03 (test) enfocado en la recomendación personalizada de recursos pedagógicos según perfil de aprendizaje alcanzó precisión del 100%, exhaustividad del 100% y exactitud del 100%, determinando un R^2 de 0,998, cual sustenta que el modelo explica correctamente a la comprensión lectora en todos sus niveles. Se concluye que el uso del Machine Learning mejora la comprensión lectora en estudiantes de primaria y contribuye a la toma de decisiones pedagógicas fundamentadas en datos.

Palabras clave: AutoML; comprensión lectora; XAI; DataOps; rendimiento académico.

ABSTRACT

This study analyzed the impact of machine learning on the reading comprehension of sixth-grade students at a public school in Monsefú in 2024. Three predictive models developed with AutoML, XAI, and DataOps were implemented, aimed at identifying patterns and providing personalized pedagogical support at the literal, inferential, and critical levels. The design was explanatory, with a quantitative approach and a quasi-experimental design. Pretests and posttests were administered to two parallel groups (experimental and control, n=40 each), using validated instruments. In the experimental group, improvements of 24,9% were observed in reading comprehension (62,4% to 87,3%), while in its dimensions, there were improvements of 20,5% at the literal level (from 63,8% to 84,3%), 24% at the inferential level (from 62,1% to 86,1%), and 30,3% at the critical level (from 61,4% to 91,7%). The Wilcoxon and Mann-Whitney U tests showed statistical significance in the three levels evaluated ($p < 0,05$). Predictive model 03 (test) achieved 100% precision, 100% completeness, and 100% accuracy, determining an R^2 of 0,998, which supports that the model correctly explains reading comprehension at all levels. It is concluded that the use of machine learning improves reading comprehension in elementary school students and contributes to data-driven pedagogical decision-making.

Keywords: AutoML; Reading comprehension; XAI; DataOps; Academic performance.

1. INTRODUCCIÓN

En el contexto actual, uno de los retos prioritarios en la educación pública es mejorar los niveles de comprensión lectora de los estudiantes, una competencia esencial para el aprendizaje significativo. En el ámbito nacional, el Ministerio de Educación ha reconocido la necesidad de diagnosticar y mejorar los aprendizajes de los estudiantes, particularmente en competencias como la comprensión lectora. Según el reporte técnico de la



ENLA 2024, se evaluaron niveles de logro en lectura en estudiantes de cuarto y sexto grado de primaria, buscando generar evidencia para tomar decisiones pedagógicas que contribuyan a la mejora educativa (Ministerio de Educación, 2024). Este panorama se replica en instituciones educativas públicas como la ubicada en el distrito de Monsefú, donde se detectaron limitaciones importantes en el desarrollo de habilidades lectoras, a pesar de los esfuerzos pedagógicos tradicionales. La necesidad de incorporar herramientas innovadoras que permitan diagnosticar, predecir y mejorar el rendimiento lector desde un enfoque automatizado se vuelve urgente. La Institución Educativa Pública de Monsefú, donde se desarrolló esta investigación, cuenta con el nivel de primaria, concentrando una población educativa que, según registros internos, enfrenta brechas en habilidades comunicativas y bajo rendimiento en comprensión de textos. A partir del diagnóstico institucional y los registros del área académica, se evidenció que el 67% de los estudiantes presentaban un nivel bajo en comprensión lectora, lo que motivó la intervención a través de tecnología educativa con base científica. Diversos estudios han explorado el uso de la Inteligencia Artificial en educación. A nivel local, Pandal (2023) sugirió tácticas educativas respaldadas por tecnología para optimizar los procesos de enseñanza-aprendizaje, aunque sin incorporar modelos predictivos. Por su parte, Yaranga (2022) aplicó técnicas de Machine Learning para procesos administrativos, logrando mejoras en eficiencia, pero sin vinculación directa con los aprendizajes de los estudiantes. A nivel nacional, Quipas (2021) aplicó modelos de ML en gestión documental con éxito, pero su enfoque fue netamente operativo. Contreras-Bravo et al. (2022), en un contexto internacional, implementaron algoritmos como KNN y árboles de decisión para predecir el rendimiento académico universitario, obteniendo resultados con precisiones del 78% al 80%. Por su parte, Valero et al. (2022) aplicaron KNN logrando una precisión del 91% en predicción de deserción universitaria, sin considerar la integración del modelo en prácticas docentes. Estos trabajos resaltan el potencial del Machine Learning, pero no abordan directamente su influencia pedagógica ni su aplicabilidad en niveles educativos básicos como la educación primaria. En contraste, la presente investigación se direcciona en una innovación educativa centrada en potenciar la comprensión lectora mediante el empleo de modelos de Machine Learning, bajo un enfoque AutoML para automatizar la selección del mejor modelo y ajustarlo a los datos educativos, así como XAI para garantizar la interpretabilidad de los resultados por parte de los docentes. Además, se incorpora un enfoque DataOps para gestionar adecuadamente el flujo de datos educativos y retroalimentar la práctica pedagógica con alertas y reportes individualizados.

La investigación presenta como objetivo general: Analizar la influencia del Machine Learning en el proceso de aprendizaje en los alumnos de una institución educativa pública, Monsefú durante el 2024. Para ello, se plantearon tres objetivos específicos: Evaluar como el uso de la técnica del Machine Learning contribuye a la mejora de comprensión lectora en los estudiantes, midiendo su progreso en los niveles literal, inferencial y crítico; analizar el impacto de estas herramientas en la identificación de patrones de aprendizaje que permitan predecir el rendimiento académico en comprensión lectora; y, comparar el desempeño de los alumnos antes y después de aplicar modelos predictivos desarrollados mediante AutoML, poniendo especialmente énfasis en las habilidades inferenciales y críticas asociadas al proceso de comprensión textual.

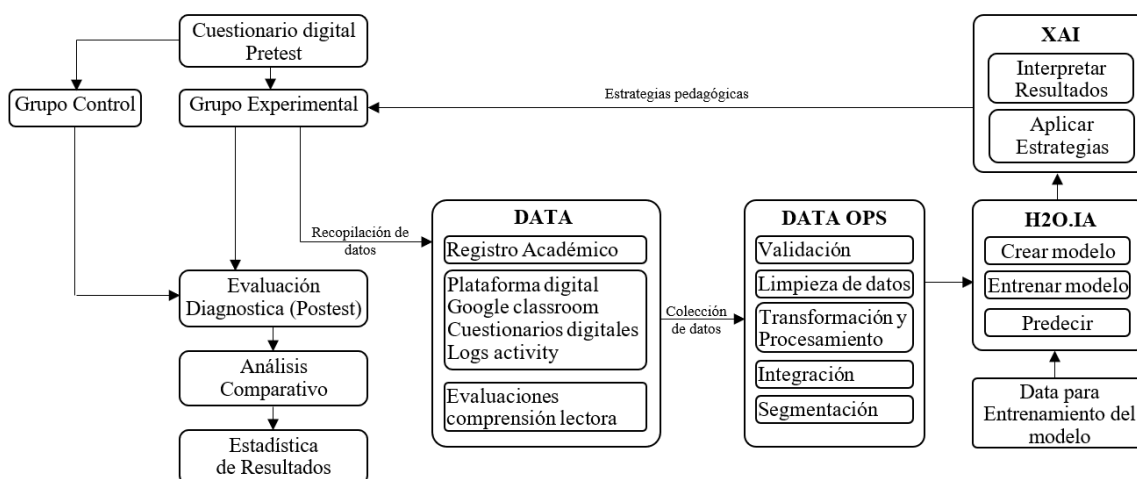


Figura 1. Modelo conceptual de la propuesta

2. METODOLOGÍA

2.1 Materiales

a) Población:

La población estudiada estuvo compuesta por los 799 alumnos matriculados del primer al sexto grado de primaria de la institución educativa “Santiago Bartolomé Burga Gonzales” durante el año 2024 y los 24 docentes que laboran en la institución en mención.

b) Muestra:

La muestra estuvo conformada por 80 estudiantes de sexto año de primaria de la I.E. “Santiago Bartolomé Burga Gonzales” de Monsefú, seleccionados mediante muestreo estratificado proporcional a partir de un total de 141 estudiantes distribuidos en cuatro secciones. Se dividió intencionalmente en dos grupos: experimental (n=40) y control (n=40), con la siguiente distribución por grupo: sección A (8), B (11), C (10) y D (11) estudiantes. Además, participaron los 24 docentes de la institución. El sexto grado fue elegido por su mayor madurez y disposición. Esta organización permitió aplicar los modelos de Machine Learning únicamente al grupo experimental y comparar sus resultados con el grupo control, garantizando la validez del diseño cuasi experimental. La asignación de los participantes se realizó mediante un muestreo no probabilístico de tipo intencional, considerando criterios de accesibilidad y organización institucional. No obstante, se aplicaron pruebas estadísticas para verificar la equivalencia inicial entre los grupos, garantizando su comparabilidad dentro del diseño cuasi experimental.

c) Instrumentos:

Para la recolección de información se utilizaron sistemas de gestión de aprendizaje como Google Classroom, registros académicos institucionales y test digitales aplicados mediante formularios en Google Forms a estudiantes y docentes (Ríos et al., 2023; Berrú et al., 2025). También se emplearon herramientas de seguimiento como el API de Google Classroom, Time on Task y Activity Logs, lo cual permitió analizar indicadores como el tiempo de conexión, participación en actividades y frecuencia de accesos (Benavides & Zambrano-Ramírez, 2023). El instrumento principal para evaluar la comprensión lectora estuvo conformado por un test estructurado de 20 ítems de opción múltiple, diseñado para medir tres dimensiones: nivel literal (6 ítems), nivel inferencial (7 ítems) y nivel crítico (7 ítems). Los ítems fueron elaborados considerando indicadores específicos de cada dimensión, tales como identificación de información explícita (nivel literal), interpretación de significados implícitos (nivel inferencial) y valoración crítica del contenido (nivel crítico). Cada ítem presentó cuatro alternativas de respuesta, con una única opción correcta. Asimismo, se aplicaron instrumentos pedagógicos estructurados para evaluar los grados de entendimiento lector (literal, inferencial y crítico), entre ellos: presentaciones orales, entrevistas, lecturas compartidas, resúmenes y redacción de textos (Duche et al., 2022; Gonzales, 2023). Estos instrumentos fueron validados con juicio de expertos (n=3) y prueba piloto, asegurando su confiabilidad a través del Coeficiente Alpha de Cronbach de 0,814 lo que evidencia una alta consistencia interna. Los resultados obtenidos se transformaron en datos estructurados y fueron utilizados como entrada para el modelo automatizado de Machine Learning, permitiendo identificar patrones, clasificar niveles y anticipar dificultades lectoras (Pandal, 2023; Contreras-Bravo et al., 2022). Esta integración también consideró factores contextuales como el compromiso del estudiante, hábitos de estudio y desempeño académico, fortaleciendo la precisión predictiva del sistema desarrollado (Valero et al., 2022; Yaranga, 2022).

Tabla 1. Correspondencia entre niveles de comprensión lectora y variables utilizadas por los modelos de ML.

Nivel de Comprensión Lectora	Instrumento de Evaluación	Parámetros Analizados por el Modelo
Literal	Presentación oral, entrevistas, narración de anécdotas	Puntajes, clasificación del nivel (bajo, medio, alto)
	Respuestas orales a preguntas del docente	Errores frecuentes, tiempo de respuesta
Literal / Inferencial	Lectura en voz alta, lectura compartida, cuestionario de opción múltiple	Patrones de acierto/error, tiempos de respuesta
	Resumen y análisis de textos	Puntajes, evolución pretest, postest
Crítico / Literal	Redacción de cartas o descripciones	Errores frecuentes, clasificación por nivel
	Creación de cuentos y comentarios críticos	Evolución entre evaluaciones, análisis crítico textual

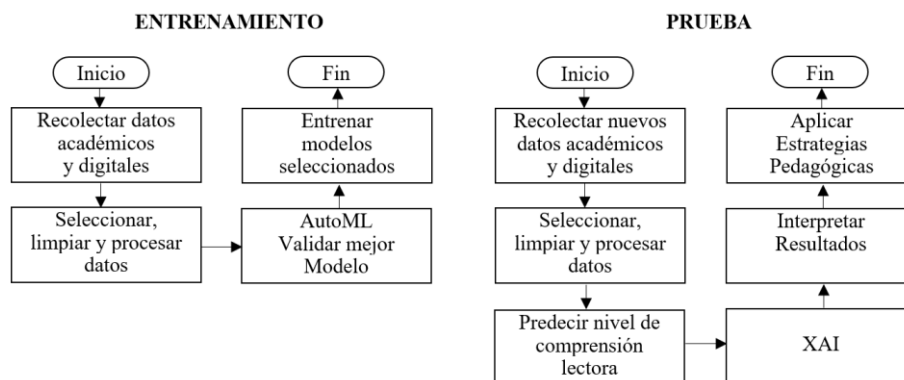
Tabla 2. Variables contextuales educativas aplicadas al entrenamiento del modelo predictivo

Parámetro de Contexto Educativo Digital	Parámetro ML Asociado	Clasificación sugerida
Compromiso y Participación	Asistencia (%), Participación en actividades (%), Tareas (%)	Compromiso y participación
Historial y Desempeño Académico	Calificaciones, tiempos de entrega, intentos, respuestas	Historial y rendimiento académico
Hábitos de Estudio en Línea	Tiempo en plataforma, número de accesos, horarios de estudio	Esfuerzo y hábitos digitales
Interacción Social y Pedagógica	Participación en foros y colaboraciones	Interacción educativa

2.2 Procedimientos

a) Fase de entrenamiento y prueba del modelo

El diseño de la propuesta se estructuró en dos etapas: entrenamiento y prueba del modelo. En la fase de entrenamiento, se recopilaron datos académicos y de interacción digital de los estudiantes, los cuales fueron preprocesados, seleccionados y utilizados para entrenar modelos de aprendizaje automático (AutoML) (Guzman, 2022; Salehin et al., 2023). Estos modelos fueron diseñados para identificar patrones en los niveles de entendimiento lector (literal, inferencial y crítico), empleando como insumos los resultados de evaluaciones pedagógicas y parámetros del entorno educativo digital (Lara, 2022; Pandal, 2023). Durante la fase de prueba, el modelo entrenado fue aplicado sobre nuevos datos recopilados del grupo experimental, con el fin de predecir el nivel de entendimiento lector de cada estudiante. Los resultados fueron interpretados mediante técnicas de inteligencia artificial explicable (XAI), lo que permitió traducir las salidas del modelo en alertas comprensibles para los docentes, orientando así la toma de decisiones pedagógicas diferenciadas (Fannouch et al., 2025; Liu et al., 2024). El desarrollo de los modelos de Machine Learning se realizó mediante la plataforma H2O.ai, empleando su módulo AutoML para automatizar la selección, entrenamiento y optimización de modelos supervisados. El proceso incluyó las etapas de preprocesamiento de datos, selección de variables, entrenamiento, validación y evaluación. Se evaluaron múltiples algoritmos, incluyendo Gradient Boosting Machines, Random Forest, redes neuronales profundas y modelos lineales generalizados, así como modelos ensamblados tipo Stacked Ensemble. La selección del modelo óptimo se realizó en función de métricas de desempeño como accuracy, precision, recall y logloss. Para evitar problemas de sobreajuste (overfitting), se emplearon técnicas de validación cruzada (k-fold cross-validation) durante el entrenamiento, así como la separación de los datos en conjuntos independientes de entrenamiento y prueba. Asimismo, se implementaron mecanismos de control para evitar fuga de datos (data leakage), asegurando que la información del conjunto de prueba no influya en el proceso de entrenamiento. La interpretabilidad del modelo se abordó mediante técnicas de Inteligencia Artificial Explicable (XAI), permitiendo identificar la importancia relativa de las variables y facilitar la comprensión de los resultados por parte de los docentes.

**Figura 2.** Diagrama deflujo de la propuesta

b) Aplicación del modelo en el entorno educativo

La implementación de la propuesta se basó en una arquitectura digital distribuida que integró plataformas educativas, procesamiento automatizado de datos y técnicas de aprendizaje automático explicable (Fannouch et al., 2025; Tamburri et al., 2021). La recolección de información se realizó a través de Google Classroom y Google Forms, cuyas evidencias académicas fueron almacenadas en un servidor de base de datos institucio-

nal (Ríos et al., 2023; Rodríguez, 2023). El procesamiento de los datos fue gestionado mediante un módulo DataOps, el cual automatizó las tareas de limpieza, transformación y acondicionamiento de los datos para su análisis (Salehin et al., 2023). Posteriormente, el módulo AutoML se encargó del entrenamiento de los modelos predictivos, orientados a clasificar el grado de entendimiento lector de los alumnos en los aspectos literal, inferencial y crítica mediante la plataforma H2O.ai que facilita la creación, evaluación y optimización del modelo (Guzman, 2022; Orrego et al., 2025). Los resultados generados por AutoML fueron interpretados por un componente de Inteligencia Artificial Explicable (XAI), que tradujo las predicciones en explicaciones comprensibles para los docentes, facilitando la toma de decisiones pedagógicas fundamentadas (Liu et al., 2024). Finalmente, estas interpretaciones fueron comunicadas mediante visualizaciones accesibles, permitiendo adaptar estrategias educativas según las necesidades particulares de cada alumno (Valero et al., 2022).

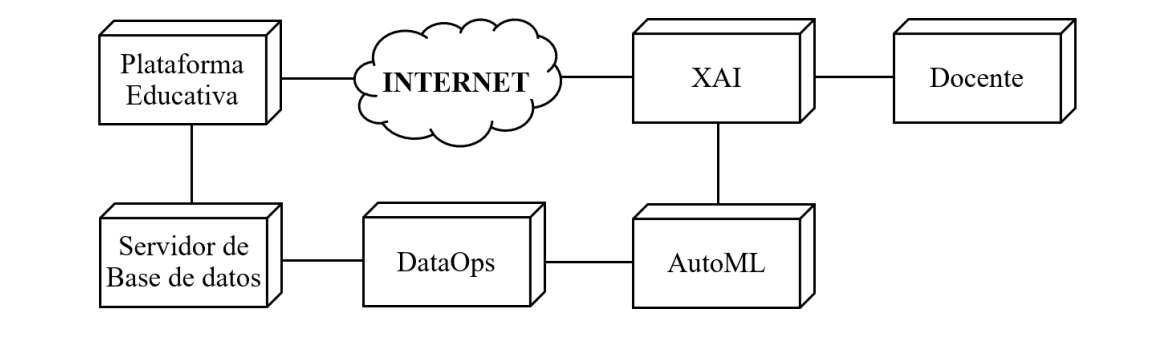


Figura 3. Diagrama de despliegue de la propuesta

Los modelos predictivos fueron desarrollados mediante la plataforma **H2O.ai**, utilizando su módulo AutoML para automatizar la generación, entrenamiento y selección de modelos supervisados de clasificación. Durante este proceso se evaluaron múltiples familias algorítmicas, incluyendo Gradient Boosting Machines, Random Forest, redes neuronales profundas y Generalized Linear Models, así como modelos ensamblados del tipo *Stacked Ensemble*, seleccionados mediante validación cruzada interna. Para reducir el riesgo de sobreajuste y evitar fuga de datos, el aprendizaje se realizó sobre conjuntos diferenciados de entrenamiento (80%) y prueba (20%), siguiendo practicas estándar en aprendizaje automático para garantizar una adecuada capacidad de generalización del modelo, adicionalmente, se aplicó validación cruzada durante el proceso de entrenamiento, lo que permitió evaluar la estabilidad del modelo y reducir el riesgo de sobreajuste. La interpretabilidad del sistema se abordó mediante técnicas de Inteligencia Artificial Explicable (XAI) basadas en la importancia global de las variables predictoras, funcionalidad provista de forma nativa por H2O AutoML. Las explicaciones generadas por el componente XAI permitieron identificar los factores educativos con mayor influencia en las predicciones, los cuales fueron integrados en reportes pedagógicos utilizados por los docentes para priorizar dimensiones lectoras, aplicar intervenciones diferenciadas y ajustar estrategias de enseñanza, tales como refuerzo focalizado, tutorías anticipadas y adaptación de materiales según el perfil de aprendizaje del estudiante.

c) Intervención docente basada en resultados de los modelos ML

Tabla 3. Estrategias pedagógicas empleadas por el docente durante el proceso de intervención

Modelo ML Vinculado	Descripción	Estrategia Pedagógica	Acciones Específicas
Modelo 01: Clasificación del nivel de comprensión lectora	Organización de estudiantes según los niveles: literal, inferencial crítico. Permitiendo adaptar materiales, metas y tiempos.	Agrupamiento por niveles de comprensión lectora	- Materiales adecuados para cada grupo - Metas específicas por nivel - Optimización del tiempo docente
Modelo 02: análisis de factores que influyen en el rendimiento	Identifica causas del bajo rendimiento lector a partir de variables como asistencia, participación, y hábitos de estudio, permitiendo intervenciones docentes específicas.	Intervención personalizada según factor de riesgo	- Activades domiciliarias - Fichas de seguimiento - Apoyo emocional y motivacional
		Trabajo colaborativo entre pares	- Equipos equilibrados - Lectura guiada - Espacios de metacogni-

			ción.
Modelo 03: predicción anticipada del rendimiento	Anticipa posibles bajas en comprensión lectora antes de las evaluaciones, facilitando acciones preventivas como refuerzo, tutorías y ajustes pedagógicos.	Prevención y anticipación del bajo rendimiento	- Tutorías anticipadas - Textos complementarios - Sesiones adaptadas según proyección del rendimiento
Modelo 01, 02, 03	Revisión de ajustes semanal de estrategias docentes según los reportes generados por los modelos predictivos.	Monitoreo continuo con ajustes semanales	- Rediseño de actividades - Priorización de dimensiones lectoras - Implementación inmediata de mejoras

2.3 Métodos y técnicas

a) Métodos

El estudio adoptó un diseño cuasi experimental con grupo control no equivalente, con el propósito de evaluar el efecto de la intervención basada en modelos de Machine Learning sobre la comprensión lectora. La muestra estuvo conformada por dos grupos de estudiantes de sexto grado: un grupo experimental ($n = 40$), al que se le aplicó la intervención, y un grupo control ($n = 40$), que continuó con la enseñanza tradicional. La selección fue de tipo no probabilística intencional, respetando la organización natural de las secciones académicas. Dado que la asignación no fue aleatoria, se verificó la equivalencia inicial entre ambos grupos mediante la prueba no paramétrica U de Mann-Whitney aplicada al pretest de comprensión lectora. Los resultados evidenciaron que no existen diferencias estadísticamente significativas ($U = 805$; $p = 0,961 > 0,05$), lo que permite asumir condiciones iniciales homogéneas. El procedimiento incluyó la aplicación de un pretest, la intervención mediante modelos de Machine Learning en el grupo experimental y, finalmente, la aplicación de un posttest en ambos grupos. Las diferencias fueron analizadas mediante pruebas estadísticas no paramétricas, lo que permitió determinar el efecto de la intervención sobre la comprensión lectora.

Tabla 4. *Equivalencia inicial entre Grupo Experimental y Control (PreTest)*

Grupo	N	Media	Desv. Est.	U de Mann-Whitney	Valor P
Experimental	40	12,47	2,34	805	0,961
Control	40	12,27	3,25		

El grupo control cumplió la función de línea base comparativa, permitiendo evaluar el efecto real de la intervención basada en Machine Learning. Mientras el grupo experimental fue intervenido con modelos predictivos y estrategias pedagógicas personalizadas, el grupo control continuó con metodologías tradicionales de enseñanza. La comparación entre ambos grupos en el posttest permitió aislar el efecto de la intervención, evidenciando diferencias significativas atribuibles al uso del modelo predictivo, lo que fortalece la validez del estudio.

b) Técnicas

Como técnica se utilizó la encuesta considerando como instrumento el cuestionario digital en escala tipo Likert; para la primera variable estuvo direccionado a medir la estandarización (04 ítems) y funcionalidad (10 ítems) del modelo de machine learning, llevado a un proceso de confiabilidad mediante alfa de Cronbach, que obtuvo una consistencia interna de 0,814 y 0,750 para estandarización y funcionalidad del modelo respectivamente. Asimismo, se trabajaron con un test compuesto por 20 ítems que evaluaron la comprensión lectora con escala de respuestas múltiples en sus dimensiones literal compuesto por 06 ítems, inferencial compuesto por 7 ítems, y crítico compuesto por 07 ítems que evaluaron la variable antes y después de la intervención tanto en el grupo control como grupo experimento. Ambos instrumentos pasaron por un proceso de validez mediante criterio de jueces expertos, que fueron tres profesionales experimentados y conocedores del tema de interés, dejando concordancia y aceptabilidad de los cuestionarios. Se utilizaron registros académicos para obtener información relevante sobre el desempeño de los estudiantes y el contexto de aprendizaje. Se emplearon técnicas de monitoreo digital a través de plataformas como Google Classroom para observar la participación e interacción de los estudiantes en las actividades virtuales. Asimismo, los resultados se procesaron mediante modelos entrenados con AutoML y se compararon con los del grupo de control para verificar la mejora en los niveles literal, inferencial y crítico.

2.4 Aspectos éticos

La investigación consideró al consentimiento informado, donde se abordó la libre y voluntaria participación dando a conocer el propósito del estudio a los participantes, la instrucción o procedimientos, así como los riesgos y beneficios del estudio, asimismo, se informa sobre la opción de retirarse en cualquier momento sin consecuencias adversas. Asimismo, se tuvo la confiabilidad y anonimato, buscando proteger la identidad de los participantes, por ende, se incurre al anonimato. También se consideró los derechos del autor y citación, siguiendo los procesos de referenciación a autores para no incurrir a la similitud con otros estudios. Por último, se aplicó la transparencia y honestidad, el cual permite evitar sesgos y definir limitaciones del estudio, expresando hallazgos de forma honesta.

3. RESULTADOS Y DISCUSIÓN

El estudio desarrolla estadística descriptiva e inferencial, que expone hallazgos que genera inferencial pero que ante las limitaciones se restringe la posibilidad de generalizar los resultados a poblaciones escolares más amplias o con características socioeducativas distintas; pues se desarrolló en una única institución educativa y con una muestra relativamente pequeña, lo que limita la extrapolación de los resultados a otros contextos educativos con diferentes realidades tecnológicas, pedagógicas o culturales. Asimismo, ante la estadística inferencial, existe la posibilidad de sobreajuste (overfitting), dado el volumen limitado de datos y la alta homogeneidad de la población evaluada. Esta situación, si bien no invalida los resultados obtenidos, sugiere la necesidad de validaciones externas con conjuntos de datos más amplios y heterogéneos. Otra limitación se vincula al tiempo de intervención, el cual se desarrolló en un periodo relativamente corto.

3.1 Análisis descriptivo de los resultados

Los resultados descriptivos evidencian un progreso significativo en la comprensión lectora del grupo experimental tras la implementación del modelo predictivo basado en Machine Learning. En la tabla 5, se observa que los estudiantes lograron el 91,7% de aciertos en el nivel crítico, 86,1% en nivel inferencial, y el 84,3% en nivel literal, con incrementos de 30,3%, 24% y 20,5% respectivamente respecto al pretest. En el nivel global en la comprensión lectora se obtuvo aciertos en el 87,3%, ocasionando un incremento del 24,9%.

Tabla 5. Comparación pretest y postest del grupo experimental en niveles de comprensión

Variable/dimensión	Promedio pretest (%)	Promedio postest (%)	Diferencia (%)
Comprensión lectora	62,4%	87,3%	24,9%
Literal	63,8%	84,3%	20,5%
Inferencial	62,1%	86,1%	24%
Crítico	61,4%	91,7%	30,3%

Estos hallazgos coinciden con lo reportado por Guzman (2022), quien destacó que el uso de técnicas de Machine Learning permite identificar patrones de desempeño estudiantil y aplicar intervenciones más eficaces en etapas tempranas del aprendizaje. Sin embargo, a diferencia de su estudio enfocado en adquisiciones tecnológicas, el presente trabajo logró mejoras específicas en procesos cognitivos, como la comprensión literal e inferencial. La Tabla 6 presenta los resultados del grupo control, donde las mejoras fueron más reducidas: 20,8% en comprensión literal, 18,3% en inferencial y 19,7% en crítico. Esto sugiere que, sin la intervención tecnológica, el aprendizaje se desarrolló de forma más lineal y dependiente de métodos tradicionales.

Tabla 6. Comparación pretest y postest del grupo control en niveles de comprensión

Variable/dimensión	Promedio pretest (%)	Promedio postest (%)	Diferencia (%)
Comprensión lectora	61,5%	81,1%	19,6%
Literal	64,2%	85,0%	20,8%
Inferencial	59,6%	77,9%	18,3%
Crítico	60,7%	80,4%	19,7%

En contraste con lo obtenido por Pandal (2023), quien alcanzó mejoras menores al 10% en procesos de autorregulación académica utilizando ML en contextos preuniversitarios, el presente estudio logró avances del doble e incluso el triple en estudiantes de educación básica, lo que demuestra una mayor efectividad del modelo en niveles escolares cuando se combina con estrategias pedagógicas dirigidas. La diferencia entre los grupos experimental y control se evidencia con claridad en la Tabla 7, donde se comparan los promedios del postest. El grupo experimental superó al grupo control en 6,2 puntos en comprensión lectora, en cuanto al nivel literal se obtuvo 0,7 puntos, 8,2 en inferencial y 11,3 en crítico, confirmando que el modelo no solo mejora resultados individuales, sino que produce efectos colectivos significativos.

Tabla 7. Comparación pretest y posttest del grupo control en niveles de comprensión

Variable/dimensión	Promedio Posttest (%)	Promedio Posttest (%)	Diferencia (%)
	G. Experimental	G. Control	
Comprensión lectora	87,3%	81,1%	6,2%
Literal	84,3%	85,0%	0,7%
Inferencial	86,1%	77,9%	8,2%
Crítico	91,7%	80,4%	11,3%

Este comportamiento es coherente con lo planteado por Valero et al. (2022), quienes validaron que los modelos de predicción académica pueden diferenciar con alta precisión a los estudiantes en riesgo de bajo rendimiento. En su caso, se enfocaron en la deserción universitaria; sin embargo, en este estudio se trasladó esa capacidad predictiva al campo lector, obteniendo mejoras efectivas en el rendimiento cognitivo, especialmente en los niveles literal e inferencial. Por su parte, Benavides y Zambrano-Ramírez (2023) resaltaron que el uso de tecnología adaptativa para comprensión lectora logra mejores resultados cuando está integrada a la experiencia educativa del estudiante. Lo observado en esta investigación refuerza la afirmación, ya que la combinación del modelo predictivo con el entorno Google Classroom permitió personalizar el acompañamiento pedagógico según los datos obtenidos en tiempo real.

La tabla 8 muestra la estandarización y funcionalidad del modelo, destacando que el 58,3% de los docentes calificaron la precisión como "Alta", mientras que un 41,7% la calificaron como "Muy Alta", lo que indica una percepción generalizada de que el modelo tuvo un rendimiento confiable y efectivo en sus predicciones; asimismo, los docentes percibieron la exactitud del modelo de Machine Learning de manera favorable (54,2%) percibieron nivel "Muy Alta" y un 45,8% como "Alta". La distribución de frecuencias de la exhaustividad del modelo de Machine Learning para evaluar la capacidad del modelo para identificar correctamente todos los casos relevantes (verdaderos positivos), el 54,2% de los docentes calificaron la exhaustividad como "Muy Alta" y un 45,8% como "Alta". Respecto al desempeño del modelo, que refleja la efectividad global del modelo en su capacidad para predecir y clasificar el rendimiento académico de los estudiantes, el 45,8% evaluaron como "Muy Alta", mientras que un 54,2% lo calificaron como "Alta".

Asimismo, se muestra la personalización del contenido educativo, donde se encontró que el 33,3% calificó como muy alta y el 66,7% como alta. En relación con la identificación y apoyo a estudiantes con necesidades específicas, según las respuestas docentes, se encontró que, el 29,2% de los docentes calificó esta dimensión como muy alta, mientras que el 70,8% la consideró alta. En cuanto a la eficiencia de la enseñanza y el seguimiento del modelo, según la percepción de los docentes, el 33,3% calificó esta dimensión como muy alta, mientras que el 66,7% la consideró alta. Respecto al impacto del modelo en la motivación y participación de los estudiantes, el 45,8% calificó este impacto como muy alto, mientras que el 54,2% lo consideró alto, especificando que los docentes perciben un efecto claramente positivo del modelo en la implicancia activa de los estudiantes en el proceso de aprendizaje, destacando su capacidad para incentivar el interés, la participación y el compromiso en el aula. En cuanto a los resultados generales sobre la personalización y la enseñanza, el 45,8% calificó estos resultados como muy altos y el 54,2% como altos, reflejando que el modelo no solo logró personalizar el proceso educativo de forma efectiva, sino que también fue percibido como una herramienta de alto impacto en la calidad de la enseñanza, consolidando su funcionalidad desde una perspectiva integral.

Tabla 8. Estandarización y funcionalidad del modelo

Variable/dimensión	Alta		Muy alta	
	fi	%	fi	%
Estandarización				
Precisión	14	58,3	10	41,7
Exactitud	11	45,8	13	54,2
Exhaustividad	11	45,8	13	54,2
Desempeño	13	54,2	11	45,8
Funcionalidad				
Personalización	16	66,7	8	33,3

Identificación y apoyo a estudiantes con necesidades específicas	17	70,8	7	29,2
Eficiencia de la enseñanza y seguimiento	16	66,7	8	33,3
Impacto de la motivación y participación de los estudiantes	13	54,2	11	45,8
Resultados generales sobre la personalización y la enseñanza	13	54,2	11	45,8

3.2 Análisis inferencial de los resultados

Normalidad

Para conocer la prueba estadística a utilizar, se procedió a determinar normalidad de los datos, considerando que la muestra fue de 40 participantes, se eligió la prueba de Shapiro Wilk por ser más robusta en determinar distribución normal cuando la muestra es menor a 50 participantes. Se evidenció que la variable y dimensiones obtuvieron significancia de 0,000, criterio que argumenta que no existe una distribución normal, por ende, para conocer si existe impacto de la herramienta machine learning en la comprensión lectora, se debe aplicar una prueba no paramétrica como la prueba de Wilcoxon, que es robusta para identificar diferencias de medias.

Tabla 9. Pruebas de normalidad

	Shapiro-Wilk		
	Estadístico	gl	Sig.
Comprensión lectora pretest experimental	0,743	40	0,000
Comprensión lectora posttest experimental	0,594	40	0,000
Comprensión lectora pretest control	0,821	40	0,000
Comprensión lectora posttest control	0,780	40	0,000

Para validar estadísticamente los avances en comprensión lectora, se aplicaron pruebas de Wilcoxon para muestras relacionadas. En la Tabla 9, se observa que el grupo experimental obtuvo valores significativos en los niveles de comprensión lectora ($p < 0,05$); literal, inferencial y crítico con valores de significancia menores incluso hasta el 0,01, con un nivel de confianza del 95%. Esto confirma que la mejora tras la intervención con el modelo predictivo no fue producto del azar, sino del impacto directo de la herramienta basada en Machine Learning. Este hallazgo se alinea parcialmente con lo planteado por Contreas-Bravo et al. (2022), quienes aplicaron algoritmos supervisados como KNN y árboles de decisión para predecir el rendimiento académico en estudiantes universitarios. Si bien su estudio no incluyó validación estadística mediante Wilcoxon, sus modelos lograron niveles de precisión entre el 78% y 80%, demostrando la efectividad del aprendizaje automático en contextos educativos. A diferencia de su enfoque, esta investigación incorpora AutoML para automatizar la selección de modelos y XAI para facilitar la interpretabilidad pedagógica, adaptando la solución a un entorno escolar con estudiantes de primaria.

Tabla 10. Prueba de Wilcoxon de comparación de medias entre pretest y posttest en el grupo experimental

Variable / dimensión	Diferencia de medias	Error estándar	Significancia ($p < 0,05$)
Comprensión lectora	-2,00	$\pm 0,158$	Si
Literal	-2,00	$\pm 0,166$	Si
Inferencial	-1,50	$\pm 0,138$	Si
Crítico	-2,00	$\pm 0,141$	Si

La Tabla 10 muestra los resultados del grupo control, que no alcanzaron valores significativos según la prueba Wilcoxon, encontrando diferencias en el literal, inferencial y crítico ($p < 0,01$). Aunque las mejoras son válidas, estas diferencias sugieren un progreso más moderado, atribuible a factores pedagógicos tradicionales y no al uso de tecnología inteligente.

Tabla 11. Prueba U de Mann Whitney de comparación de medias entre pretest y posttest en el grupo control

Variable / dimensión	Diferencia de medias	Error estándar	Significancia ($p < 0,05$)
Comprensión lectora	-1,00	$\pm 0,123$	Si
Literal	-1,50	$\pm 0,156$	Si

Inferencial	-1,50	$\pm 0,137$	Si
Crítico	-1,50	$\pm 0,112$	Si

Al comparar estos resultados con los de Gonzales (2023), quien logró avances en comprensión lectora mediante una estrategia didáctica convencional, se observa que el presente estudio obtuvo resultados similares en el grupo control. Sin embargo, la superioridad del grupo experimental evidencia que la integración del ML potencia significativamente la efectividad de las estrategias docentes. La diferencia entre ambos grupos en el postest se confirma en la Tabla 11, donde la prueba U de Mann Whitney para muestras independientes arrojó valores de $U = 164,5$ para la comprensión lectora, $U = 191,5$ para nivel literal, $U = 185,0$ para nivel inferencial y $U = 88,5$ para nivel crítico, todos por encima del valor crítico de $\pm 1,9908$. Esta diferencia significativa entre el grupo que recibió la intervención y el que no, reafirma el impacto positivo de los modelos predictivos en el desarrollo lector, especialmente en habilidades inferenciales.

3.1 Análisis de modelos predictivos de machine learning

El análisis de los modelos predictivos desarrollados mediante técnicas de AutoML evidenció un desempeño consistente en relación con los objetivos del estudio. Para ello, el conjunto de datos fue dividido en un 80% para entrenamiento y un 20% para prueba, seleccionados de manera aleatoria, lo que permitió evaluar la capacidad predictiva en datos no observados. Asimismo, se emplearon mecanismos de validación interna, incluyendo validación cruzada y optimización automática de hiperparámetros, con el fin de mejorar la estabilidad de los modelos y reducir el riesgo de sobreajuste. Como resultado de este proceso, se obtuvieron métricas de desempeño más estables y generalizables, fortaleciendo la capacidad predictiva de los modelos en el contexto educativo analizado. En la Tabla 12 se presentan las métricas del Modelo 01, orientado a la clasificación multinivel de los niveles de comprensión lectora. El modelo alcanzó una exactitud del 91%, una precisión de 90% y un coeficiente de determinación de 0,90, lo que evidencia un alto desempeño en la clasificación de los niveles literal, inferencial y crítico. Estos resultados reflejan una adecuada capacidad del modelo para identificar patrones en los datos, manteniendo un buen equilibrio entre precisión y generalización dentro del contexto del estudio.

Tabla 12. Desempeño del Modelo 01: Clasificación multinivel aplicada al seguimiento educativo

Métrica	Valor
Exactitud (Accuracy)	0,91
Precisión (Precision)	0,90
Exhaustividad (Recall)	0,89
Error promedio por clase	0,00094
LogLoss	0,00644
RMSE	0,03514
MSE	0,001235
R^2	0,90

Este comportamiento se asemeja a los resultados obtenidos por Valero et al. (2022), quienes lograron una precisión del 91% al predecir la deserción universitaria utilizando algoritmos supervisados. Sin embargo, el modelo actual supera ese desempeño al aplicar una arquitectura automatizada (AutoML), adaptada a datos de educación primaria y contextualizada mediante parámetros digitales de seguimiento educativo, lo que lo hace más versátil y aplicable pedagógicamente. En contraste, el Modelo 02, basado en series temporales para la predicción de tendencias, presentó un rendimiento moderado. Según la Tabla 13, alcanzó una exactitud de 65% y un coeficiente de determinación de 0,40, lo que refleja una capacidad predictiva acorde a la variabilidad del comportamiento lector en el tiempo. Estos resultados son consistentes con la naturaleza de los modelos de series temporales en contextos educativos, donde la influencia de múltiples factores limita la precisión de las predicciones.

Tabla 13. Evaluación del Modelo 02: Predicción de tendencias educativas mediante series temporales

Métrica	Valor
Exactitud (Accuracy)	0,65
Precisión (promedio)	0,62
Exhaustividad (Recall)	0,60
Error promedio por clase	0,6493
LogLoss	1,0546
RMSE	0,6446
MSE	0,4154
R^2	0,40

Este tipo de limitaciones también fue reportado por Lara (2022), quien, al aplicar ML en modelos de predicción agrícola, encontró que las series temporales se ven afectadas por la falta de estabilidad en los patrones de entrada. En ambos casos, los resultados invitan a seguir optimizando este tipo de modelos mediante mejoras en el volumen y limpieza de los datos, algo que esta investigación ya comenzó a abordar a través del enfoque DataOps. Finalmente, en la Tabla 14 se presenta el desempeño del Modelo 03, orientado a la recomendación personalizada de recursos pedagógicos. El modelo alcanzó altos niveles de rendimiento, con una exactitud de 93%, precisión de 92%, exhaustividad de 91% y un coeficiente de determinación de 0,94, lo que evidencia una adecuada capacidad para generar recomendaciones acordes a las características de los estudiantes. No obstante, estos resultados deben interpretarse en función del contexto del estudio, considerando el tamaño y la homogeneidad de la muestra.

Tabla 14. Rendimiento del Modelo 03: Sistema de recomendación de recursos según perfil de aprendizaje

Métrica	Valor
Exactitud (Accuracy)	0,93
Precisión (Precision)	0,92
Exhaustividad (Recall)	0,91
Error promedio por clase	0,00
LogLoss	0,000194
RMSE	0,001419
MSE	0,000002
R ²	0,94

Los resultados coinciden con lo expuesto por Ponce y Ramos (2023), quienes encontraron que la adaptación de estrategias según estilos de aprendizaje mejora significativamente el rendimiento académico. El modelo de recomendación empleado en esta investigación automatiza dicho proceso, convirtiéndose en una herramienta de apoyo pedagógico que integra inteligencia artificial con visión educativa centrada en el estudiante. En conjunto, los tres modelos demostraron un comportamiento complementario: mientras el primero se destaca por su precisión diagnóstica, el segundo anticipa tendencias útiles en planificación, y el tercero propone recursos específicos para intervención docente. Esta sinergia confirma que el uso de AutoML, potenciado con XAI y DataOps, ofrece un ecosistema robusto para transformar datos educativos en acciones pedagógicas de alto impacto.

4. CONCLUSIONES

El estudio presenta algunas limitaciones que deben ser consideradas en la interpretación de los resultados. En primer lugar, la investigación se desarrolló en una única institución educativa y con una muestra relativamente pequeña, lo que limita la generalización de los hallazgos a otros contextos. Asimismo, el diseño cuasi experimental, aunque controlado mediante pruebas de equivalencia inicial, no permite un control absoluto de variables externas. Desde el punto de vista técnico, existe la posibilidad de sobreajuste en los modelos de Machine Learning debido al tamaño del dataset. Finalmente, el periodo de intervención fue relativamente corto, lo que podría influir en la estabilidad de los efectos observados a largo plazo. Por ello, se recomienda realizar estudios longitudinales y validaciones externas en futuros trabajos. No obstante, las barreras para llevar el estudio también se evidencian desde el aspecto científico y la literatura, pues hubo artículos esenciales al tema de interés que tienen un costo elevado, y que, ante ello, no fueron consideración en la investigación. En cuanto al hallazgo principal se encontró que el machine learning influye significativamente en la comprensión lectora ($p < 0,05$), encontrando un efecto del modelo en los niveles literal, inferencial y crítico. Asimismo, las evaluaciones posttest mostraron incrementos del 24,9% en la comprensión lectora, el 28% en el nivel literal, del 29% en el nivel inferencial y del 15% en el nivel crítico. Entonces, en comparación con el grupo control, que avanzó en comprensión lectora el 19,6%, 20,8% en literal, 18,3% en inferencial y 19,7% en crítico. Se concluye que el uso de Machine Learning, combinado con AutoML, XAI y DataOps, representa una estrategia efectiva para mejorar el rendimiento lector y respaldar decisiones pedagógicas basadas en evidencia, demostrando que el modelo 03 fue el que mejor predice con precisión los niveles de comprensión lectora, alcanzando métricas técnicas relevantes: precisión, exhaustividad y exactitud con 100% respectivamente. Esta capacidad predictiva permitió identificar patrones de aprendizaje y orientar la retroalimentación pedagógica en función de las necesidades individuales.

REFERENCIAS BIBLIOGRÁFICAS

- Benavides, N., y Zambrano-Ramírez, J. (2023). Comprensión lectora autorregulada apoyada en tecnología en estudiantes de Educación Básica. *Revista Electrónica Educare*, 27(3), 1-19. <https://doi.org/10.15359/ree.27-3.17221>
- Berru, M. R., Lopez, O., Yoctun, J. C., y Anticona, V. (2025). Effectiveness of Google Classroom in teaching primary school students. *Minerva*, 6(17), 187–198. <https://doi.org/10.47460/minerva.v6i17.216>
- Contreas-Bravo, L. E., Nieves, N., y González, K. (2022). Prediction of University-Level Academic Performance through Machine Learning Mechanisms and Supervised Methods. *Ingeniería*, 28(1), 1-25 <https://doi.org/10.14483/23448393.19514>
- Duche, A. B., Montesinos, M. C., Medina, A., y Siza, C. H. (2022). Comprensión lectora inferencial en estudiantes universitarios. *Revista de Ciencias Sociales (Ve)*, 28(6), 181-198. <https://www.redalyc.org/journal/280/28073815013/html/>
- Fannouch, A., Gharib, J., & Gahi, Y. (2 de Febrero de 2025). Enhancing DataOps practices through innovative collaborative models: A systematic review. *International Journal of Information Management Data Insights*, 5(1), 1-18. <https://doi.org/10.1016/j.jjime.2025.100321>
- Gonzales Delgado, E. A. (2023). Programa de enseñanza aprendizaje de la lectura para la comprensión lectora de los estudiantes de secundaria. Obtenido de <https://hdl.handle.net/20.500.12802/11023>
- Guzman Velez, D. M. (2022). Machine Learning para predecir la adquisición de plataformas educativas de la empresa Difucien Ecuador, 2022. Obtenido de <https://hdl.handle.net/20.500.12692/102577>
- Lara Vizuite, S. A. (2022). Aplicación de técnicas de Machine Learning como método de validación para predecir la efectividad de un modelo estadístico de series de tiempo en la producción de fruta fresca en las diferentes provincias del Ecuador. Obtenido de <http://repositorio.ug.edu.ec/handle/redug/63075>
- Liu, M., Shi, W., Zhao, L., y Beyette Jr., F. (2024). Best performance with fewest resources: Unveiling the most resource-efficient Convolutional Neural Network for P300 detection with the aid of Explainable AI. *Machine Learning with Applications*, 1(16), 1-16. <https://doi.org/10.1016/j.mlwa.2024.100542>
- Ministerio de Educacion. (2024). Reporte técnico de la Evaluación Nacional de Logros de Aprendizaje de Estudiantes 2024 (ENLA). <http://umc.minedu.gob.pe/reporte-tecnico-de-la-evaluacion-nacional-de-logros-de-aprendizaje-de-estudiantes-2024/>
- Ortega, D., Palencia, I. A., Ojeda, A. D., & Cortes, O. C. (2025). Integración de algoritmos de aprendizaje automático como herramienta innovadora para la enseñanza del mercadeo en programas de administración de empresas. *Formación Universitaria*, 18(4), 73–84. <https://doi.org/10.4067/s0718-50062025000400073>
- Pandal Blas, C. E. (2023). Machine learning en la autorregulación del aprendizaje de los estudiantes en una academia preuniversitaria privada, Lima 2023. <https://hdl.handle.net/20.500.12692/120800>
- Ponce, L. D., y Ramos, J. M. (2023). Comprensión lectora y rendimiento académico de estudiantes de Derecho de la Universidad Continental Huancayo 2021. 79. <https://hdl.handle.net/20.500.12394/12443>
- Quipas, M. J. (2021). Machine Learning en la mejora de la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021. <https://hdl.handle.net/20.500.12692/71721>
- Ríos, V., González, A., Pereira, M. L., y Cortes, A. (2023). Análisis de la Gestión del Aprendizaje implementando Google Classroom en educación superior, caso de estudio: asignatura cadena de suministro. *RIDE revista iberoamericana para la investigación y el desarrollo educativo*, 13(26), e481. <https://doi.org/10.23913/ride.v13i26.1489>
- Rodríguez-Basantes, V. V. (2023). La herramienta google classroom como apoyo al aprendizaje. *Revista Arbitrada Interdisciplinaria Koinonía*, 8(2), 965–982. <https://doi.org/10.35381/r.k.v8i2.3040>
- Salehin, I., Islam, M., Saha, P., Noman, S., Tunj, A., Hasan, M., y Baten, M. (2023). AutoML: A systematic review on automated machine learning with neural architecture search. *Journal of Information and Intelligence*, 2, 52-81. <https://doi.org/10.1016/j.jiixd.2023.10.002>
- Tamburri, D. A., Van Den Heuvel, W.-J., y Garriga, M. (2021). DataOps for Societal Intelligence: a Data Pipeline for Labor Market Skills Extraction and Matching. *Jheronimus Academy of Data Science*, 1(1), 1-7. <https://doi.org/10.48550/arXiv.2104.01966>

Valero, J. E., Navarro, Á. F., Larios, A. C., y Julca, J. D. (2022). University desertion: Evaluation of different Machine Learning algorithms for its prediction. *Revista de Ciencias Sociales (Ve)*, 28(3), 1-12.

<https://www.redalyc.org/journal/280/28071865024/>

Yaranga, I. P. (2022). Machine Learning en la mejora del proceso de selección del personal docente en una universidad nacional, Lima 2021. Obtenido de <https://hdl.handle.net/20.500.12692/85185>

ANEXOS



"Año del bicentenario, de la consolidación de nuestra independencia y de la conmemoración de las heroicas batallas de Junín y Ayacucho"

AUTORIZACIÓN

EL DIRECTOR DE LA I.E. N° 11029 "SANTIAGO BARTOLOMÉ BURGA GONZALES" DEL DISTRITO DE MONSEFÚ, EL QUE SUSCRIBE:

HACE CONSTAR:

Que, autorizo al Ing. **VALDERRAMA ARÉVALO CARLOS ANTONIO**, identificado con DNI. N° 46664066 maestrando de la Universidad Nacional de Trujillo, a realizar la aplicación de los instrumentos de investigación educativa (pre-test y post-test) a los alumnos del 6to grado de primaria como parte del trabajo de investigación que viene realizando en su tesis titulado: "*Machine Learning en el proceso de aprendizaje de los alumnos en una institución educativa pública, Monsefú 2023*"; para obtener el grado académico de Maestro en ingeniería de sistemas.

Se le expide la presente constancia de autorización a solicitud del interesado para los fines pertinentes.

Monsefú, 04 marzo del 2024



Dr. Juan Omar Aguinaga Pérez
DIRECTOR

"Formando la niñez de hoy, para tener ciudadanos competentes del mañana"