

UNIVERSIDAD PRIVADA LÍDER PERUANA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E
INFORMÁTICA



UNIVERSIDAD
LÍDER PERUANA

Tesis

Implementación de un Algoritmo Predictivo para Mejorar el Rendimiento
Deportivo en Promesas Kumpirusiato, 2025

Para obtener título de Ingeniero de Sistemas e Informática

Autores:

Miguel Ángel Cornejo Torres
Hugo Ricardo Cortez Umeres

Asesor:

Dr. Jeronimo Soria Mormontoy

Santa Ana, La Convención, Cusco
2026

Título

“Implementación de un Algoritmo Predictivo para Mejorar el Rendimiento Deportivo en Promesas Kumpirusiato, 2025.”

Línea de investigación

Gestión de los Sistemas de Información

grupo 141.docx

INFORME DE ORIGINALIDAD

12%	12%	6%	5%
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	repositorio.ulp.edu.pe Fuente de Internet	3%
2	revistas.unamad.edu.pe Fuente de Internet	1%
3	alicia.concytec.gob.pe Fuente de Internet	1%
4	repositorio.ucsm.edu.pe Fuente de Internet	<1%
5	repositorio.undac.edu.pe Fuente de Internet	<1%
6	repositorio.uwiener.edu.pe Fuente de Internet	<1%
7	repositorio.unsm.edu.pe Fuente de Internet	<1%
8	Submitted to Ana G. Méndez University Trabajo del estudiante	<1%
9	repositorio.unc.edu.pe Fuente de Internet	<1%

Hoja de firma de jurados

UNIVERSIDAD PRIVADA LIDER PERUANA

ESCUELA PROFESIONAL DE INGENIERIA DE SISTEMAS



UNIVERSIDAD
LÍDER PERUANA

TESIS

**“Implementación de un Algoritmo Predictivo para Mejorar el Rendimiento
Deportivo en Promesas Kumpirusiato, 2025”**

Presentado por **Miguel Angel Cornejo Torres y Hugo Ricardo Cortez Umeres**, para
optar el Título de:

Ingeniero de Sistemas e Informática

Presidente: _____

Dr. Edgar Quispe Ccapacca

Primer Miembro: _____

Mg. Moises Gustavo García Jimenez

Segundo Miembro: _____

Mg. Rauil Huillca Huallparimachi

Asesor (es): *Dr. Jeronimo Soria Mormontoy*

Dedicatoria

A mis padres, por ser mi pilar fundamental, su amor incondicional y su esfuerzo constante que me han guiado en cada paso de mi formación académica y personal. A mis hermanos y amigos, por su apoyo moral y por motivarme a culminar con éxito este gran objetivo profesional.

Miguel Angel Cornejo

A Dios por darme la fortaleza necesaria en los momentos de cansancio. A mi familia, cuyo respaldo y sabios consejos han sido la luz en mi camino hacia la superación. Este logro es el resultado de su fe y confianza en mí.

Hugo Ricardo Cortez

Agradecimientos

A Dios por habernos bendecido con salud, sabiduría y perseverancia para afrontar cada uno de los retos académicos y culminar este proyecto de investigación.

A la Universidad Líder Peruana, por abrirnos sus puertas y brindarnos una formación profesional de calidad, dotándonos de las herramientas necesarias para el desarrollo científico y tecnológico.

A nuestro asesor, el Dr. Jeronimo Soria, por su valiosa orientación, paciencia y por compartir sus conocimientos fundamentales que guiaron de forma acertada el desarrollo de la tesis.

A la institución Promesas Kumpirusiato, por brindarnos las facilidades, el acceso a la información y el apoyo constante para la implementación de nuestro algoritmo predictivo, permitiendo validar el impacto real de la ingeniería de sistemas en el rendimiento deportivo de los estudiantes.

Miguel Angel Cornejo y Hugo Ricardo Cortez

Índice de tablas

Tabla 1 Resumen de aportes de antecedentes internacionales.....	14
Tabla 3 Dimensiones de la variable independiente	27

Índice de figuras

No se encuentran elementos de tabla de ilustraciones.

Resumen

La presente investigación analiza el impacto de la implementación de un algoritmo predictivo basado en Machine Learning, desplegado como plugin de WordPress, sobre el Rendimiento Deportivo de los estudiantes de la I.E. Promesas Kumpirusiato durante el año 2025. El problema identificado radica en la ausencia de herramientas digitales para el seguimiento objetivo del desempeño físico en una institución educativa ubicada en una zona rural amazónica de la región Cusco, lo que limita la capacidad del cuerpo docente para identificar estudiantes con bajo rendimiento y orientar intervenciones pedagógicas oportunas.

El objetivo principal es determinar en qué medida la implementación del sistema predictivo mejora el Rendimiento Deportivo, evaluado a través de dos dimensiones: competencia motora y dedicación y compromiso atlético. La investigación adopta un enfoque cuantitativo con diseño cuasiexperimental de pretest y posttest en un solo grupo. Los datos son recolectados mediante ficha de evaluación física y cuestionario estructurado, procesados con el software R mediante prueba t de Student o prueba de Wilcoxon según la distribución de los datos.

Se espera evidenciar una mejora significativa en ambas dimensiones del Rendimiento Deportivo tras la implementación del sistema, aportando un modelo replicable para otras instituciones educativas rurales del Perú que busquen modernizar su gestión del desempeño físico estudiantil mediante tecnologías de bajo costo.

Palabras clave: Algoritmo predictivo, Machine Learning, Rendimiento Deportivo, WordPress, plugin, shortcode, Random Forest, educación rural.

Abstract

This research analyzes the impact of implementing a predictive algorithm based on Machine Learning, deployed as a WordPress plugin, on the Sports Performance of students at I.E. Promesas Kumpirusiato during 2025. The identified problem lies in the absence of digital tools for objective physical performance monitoring in an educational institution located in a rural Amazonian area of the Cusco region, which limits teachers' ability to identify low-performing students and guide timely pedagogical interventions.

The main objective is to determine to what extent the implementation of the predictive system improves Sports Performance, evaluated through two dimensions: motor competence and athletic dedication and commitment. The study follows a quantitative approach with a quasi-experimental pretest-posttest single-group design. Data are collected using a physical evaluation form and a structured questionnaire, processed with R software using Student's t-test or Wilcoxon test depending on data distribution.

The findings are expected to show a significant improvement in both dimensions of Sports Performance after system implementation, providing a replicable model for other rural educational institutions in Peru seeking to modernize their student physical performance management through low-cost technologies.

Keywords: Predictive algorithm, Machine Learning, Sports Performance, WordPress, plugin, shortcode, Random Forest, rural education.

INDICE

Dedicatoria.....	v
Agradecimientos	vi
Índice de tablas	vii
Índice de figuras	viii
No se encuentran elementos de tabla de ilustraciones.....	Error! Bookmark not defined.
Resumen	ix
Abstract.....	x
Introducción.....	xiv
CAPÍTULO I. Planteamiento del problema	1
1.1 Antecedentes del problema.....	1
1.2 Formulación del problema.....	4
1.2.1 Problema general	4
1.2.2 Problemas específicos.....	4
1.3 Objetivo de investigación	4
1.3.1 Objetivo general	4
1.3.2 Objetivos específicos.....	4
1.4 Formulación de hipótesis.....	5
1.4.1 Hipótesis general	5
1.4.2 Hipótesis específicas.....	5
1.5 Justificación de la investigación	5
1.5.1 Justificación teórica	5
1.5.2 Justificación práctica	6
1.5.3 Justificación económica.....	6
1.5.4 Justificación Metodológica.....	6
1.6 Delimitaciones de la investigación	7
1.6.1 Espacial.....	7
1.6.2 Temporal.....	7
1.6.3 Teórico.....	8
CAPÍTULO II. Marco teórico	9
2.1 Antecedentes de la investigación.....	9
2.1.1 Antecedentes internacionales	9
2.1.2 Antecedentes nacionales.....	15

2.2	Bases teóricas	22
2.2.1	Variable independiente: Algoritmo Predictivo en Plugin WordPress	22
2.2.2	Variable dependiente: Rendimiento Deportivo	24
2.2.3	Identificación y conceptualización de variables	25
2.2.4	Operacionalización de variables	32
2.3	Marco conceptual	33
2.3.1	Algoritmo predictivo	33
2.3.2	Machine Learning (Aprendizaje automático).....	33
2.3.3	Random Forest.....	33
2.3.4	Plugin WordPress	33
2.3.5	Competencia motora.....	33
2.3.6	Compromiso atlético	33
2.3.7	Inteligencia artificial en educación física	33
2.3.8	Minería de datos educativos (Educational Data Mining)	34
2.3.9	Compromiso atlético	34
CAPÍTULO III. Metodología de la investigación		35
3.1	Tipo de investigación	35
3.2	Enfoque de la investigación.....	35
3.3	Diseño de la investigación.....	35
3.4	Alcance de investigación	35
3.5	Población y muestra	36
3.5.1	Población	36
3.5.2	Descripción de muestra	36
3.6	Técnica e instrumento.....	36
3.6.1	Técnica.....	36
3.6.2	Instrumento	37
3.7	Técnicas de procesamiento y análisis de datos.....	37
3.7.1	Técnicas de análisis de datos	38
3.8	Procedimiento de procesamiento de datos	38
3.9	Confiabilidad y validez.....	39
CAPÍTULO IV. Resultados, contrastación de hipótesis y discusión.....		40
4.1	Entorno de pruebas	40
4.2	Resultados.....	42

4.3	Contrastación de hipótesis	43
4.3.1	Prueba de hipótesis	Error! Bookmark not defined.
4.4	Discusión	45
CONCLUSIONES		47
RECOMENDACIONES		49
REFERENCIAS BIBLIOGRÁFICAS		50
ANEXOS		56

Introducción

La incorporación de tecnologías de inteligencia artificial en los procesos educativos representa una de las transformaciones más significativas del siglo XXI, y su aplicación en el ámbito de la educación física y el rendimiento deportivo escolar constituye un campo emergente con alto potencial para mejorar la calidad de la enseñanza en contextos de recursos limitados. Sin embargo, en zonas rurales del Perú, los docentes aún evalúan el rendimiento con métodos tradicionales, sin herramientas digitales que permitan registrar, analizar o predecir resultados de forma objetiva y sistemática.

La presente investigación surge ante esta brecha concreta en la I.E. Promesas Kumpirusiato, ubicada en el centro poblado de Kumpirusiato, distrito de Santa Ana, provincia de La Convención, región Cusco. La institución atiende a estudiantes de en una zona geográficamente aislada de la región amazónica, donde la ausencia de herramientas digitales de seguimiento deportivo limita la identificación oportuna de estudiantes con bajo rendimiento físico y la orientación de intervenciones pedagógicas basadas en evidencia.

La investigación propone implementar un algoritmo predictivo basado en Machine Learning, específicamente Random Forest, integrado como plugin en WordPress. Este sistema permite a los docentes registrar datos físicos de los estudiantes, procesarlos automáticamente y obtener predicciones y alertas tempranas para apoyar la toma de decisiones pedagógicas en tiempo real, sin necesidad de conocimientos técnicos avanzados.

El documento se organiza en cuatro capítulos. El Capítulo I presenta el planteamiento del problema, la formulación de la pregunta de investigación, los objetivos, las hipótesis, la justificación y las delimitaciones del estudio. El Capítulo II desarrolla el marco teórico, incluyendo los antecedentes internacionales, nacionales y regionales, las bases teóricas que sustentan la investigación y el marco conceptual. El Capítulo III describe la metodología, el diseño, la población, los instrumentos y las técnicas de análisis de datos. El Capítulo IV presenta los resultados, la contrastación de hipótesis y la discusión. Finalmente, se incluyen las conclusiones, recomendaciones, referencias bibliográficas y anexos.

CAPÍTULO I. Planteamiento del problema

1.1 Antecedentes del problema

La inactividad física en la población escolar constituye uno de los desafíos de salud pública más urgentes del siglo XXI. La Organización Mundial de la Salud (OMS, 2022) advierte que el 81 % de los adolescentes entre 11 y 17 años no alcanza los niveles recomendados de actividad física. Además, estima que entre 2020 y 2030 podrían registrarse casi 500 millones de nuevos casos prevenibles de enfermedades no transmisibles si esta tendencia no se revierte. Este escenario evidencia la necesidad de fomentar la actividad física desde la etapa escolar como prioridad para la salud y el desarrollo social (OMS, 2022).

En lo referente al rendimiento deportivo como expresión específica del desarrollo físico escolar, la evidencia científica muestra que la condición física influye directamente en el desempeño de los estudiantes durante actividades competitivas. Portillo et al. (2022) en un estudio realizado con 43 escolares de secundaria clasificados según su nivel de aptitud física, encontraron que quienes presentaban mejor condición física obtuvieron resultados superiores en carrera, desplazamientos de alta velocidad, sprints y distancia total recorrida durante una competencia simulada de ultimate frisbee. Asimismo, registraron mayor participación técnica y mejores decisiones ofensivas. De forma complementaria, Peterson et al. (2022), dentro del estudio longitudinal Malmö Youth Sport Study, identificaron que aproximadamente el 28 % de los estudiantes de una escuela con perfil deportivo integraban el grupo de élite competitiva luego de tres años de seguimiento, resaltando la importancia de evaluar el rendimiento atlético de manera continua desde edades tempranas.

La relación entre la práctica deportiva y el rendimiento estudiantil también ha sido analizada desde una perspectiva integral. Loturco et al. (2022), mediante una revisión sistemática con metaanálisis de 31 estudios, concluyeron que los programas de actividad física escolar generan mejoras significativas en el rendimiento académico general, siendo el área de matemáticas una de las más favorecidas. En la misma línea, Romero-Carazas et al. (2025), al evaluar a 680 estudiantes de

secundaria en Perú mediante modelado de ecuaciones estructurales, hallaron correlaciones altas entre el rendimiento académico y variables vinculadas con la práctica deportiva, como la intensidad (0.868), la frecuencia (0.862) y la duración (0.890). Estos resultados sugieren que el deporte contribuye al fortalecimiento de capacidades cognitivas, motoras, emocionales y sociales necesarias para un mejor desempeño escolar.

Del mismo modo, Porto Maciel et al. (2023), en una revisión sistemática de 17 estudios sobre participación deportiva y rendimiento escolar en jóvenes atletas, identificaron una relación positiva entre deporte y estudio, así como una influencia favorable de la práctica deportiva en el rendimiento académico. Además, señalaron que, en general, los estudiantes deportistas obtenían mejores resultados escolares que los no deportistas. Estos hallazgos refuerzan el valor del deporte como factor de desarrollo integral en la etapa escolar.

Actualmente, la evaluación objetiva del desempeño físico en escolares ha encontrado en la inteligencia artificial y en los dispositivos tecnológicos herramientas de gran utilidad. Ahmed et al. (2022) desarrollaron un estudio observacional con 29 estudiantes de secundaria en Qatar utilizando dispositivos wearables Fitbit Charge 5 durante tres semanas reportaron que solo el 38 % cumplía con la recomendación diaria de actividad física, con diferencias entre varones (50 %) y mujeres (29 %). Los autores destacan que estas tecnologías permiten recopilar datos con mayor precisión, favoreciendo diagnósticos más confiables y mejores estrategias de intervención.

En esa misma línea, la aplicación de la inteligencia artificial en la educación física viene adquiriendo creciente relevancia a nivel internacional. Wang y Wang (2024), señalan que el uso de la tecnología como la IA puede apoyar la planificación pedagógica, el seguimiento formativo y la evaluación de resultados, mejorando la calidad de la enseñanza. Sin embargo, advierten que este campo aún se encuentra en fase inicial y que muchos docentes no cuentan con la preparación necesaria para aplicarla de forma efectiva, evidenciando una brecha entre el potencial tecnológico y su uso real en las instituciones educativas.

En el contexto peruano, la situación también resulta preocupante. Un estudio de Valladares-Garrido et al. (2025) en adolescentes de cinco colegios de Chiclayo reportó que solo el 33.8 % alcanzaba niveles adecuados de actividad física. Además, se observó mayor práctica en varones y en estudiantes con mejor apoyo familiar, mientras que el uso excesivo de redes sociales y la disfunción familiar se asociaron con menor actividad física. Estos hallazgos evidencian que el problema responde a múltiples factores y que la escuela y la familia cumplen un rol protector importante.

Frente a esta realidad, el Ministerio de Salud del Perú (2025) aprobó la Guía de Actividad Física para una Vida Saludable, orientada a promover la práctica regular en todas las etapas de vida. El documento advierte que la inactividad física incrementa el riesgo de sobrepeso y enfermedades no transmisibles, con mayor incidencia en jóvenes (42 %) y adolescentes (23.9 %). Estos datos evidencian la necesidad de fortalecer acciones preventivas desde la etapa escolar.

A pesar de los avances normativos y de la evidencia científica existente, muchas instituciones educativas ubicadas en contextos alejados continúan enfrentando limitaciones importantes. La I.E. Promesas Kumpirusiato, situada en una zona geográficamente aislada de la región amazónica, no dispone de herramientas digitales orientadas al seguimiento, evaluación ni predicción del rendimiento deportivo de sus estudiantes. Los docentes de educación física no cuentan con sistemas que permitan registrar información objetiva, identificar progresos o retrocesos, ni diseñar intervenciones personalizadas basadas en datos. Esta carencia tecnológica dificulta una valoración rigurosa del desempeño físico y limita la detección oportuna de estudiantes que requieren apoyo o que poseen potencial deportivo destacado.

En este escenario, la implementación de un algoritmo predictivo basado en Machine Learning, desplegado como plugin de WordPress, surge como una alternativa técnicamente viable, accesible y adaptable a la realidad institucional. Una herramienta de esta naturaleza permitiría procesar información relacionada con el rendimiento físico de los estudiantes y generar predicciones útiles para orientar decisiones pedagógicas y estrategias de mejora en tiempo real.

Ante lo expuesto, surge la siguiente interrogante de investigación:

1.2 Formulación del problema

1.2.1 Problema general

¿En qué medida la implementación de un algoritmo predictivo basado en Machine Learning, desplegado como plugin de WordPress, mejora el Rendimiento Deportivo de los estudiantes de la I.E. Promesas Kumpirusiato durante el año 2025?

1.2.2 Problemas específicos

- PE1.** ¿Cómo influye la implementación del algoritmo predictivo basado en Machine Learning en el nivel de competencia motora de los estudiantes de la I.E. Promesas Kumpirusiato antes y después de su aplicación durante el año 2025?
- PE2.** ¿Cómo influye la implementación del algoritmo predictivo basado en Machine Learning en el nivel de dedicación y compromiso atlético de los estudiantes de la I.E. Promesas Kumpirusiato antes y después de su aplicación durante el año 2025?

1.3 Objetivo de investigación

1.3.1 Objetivo general

Determinar en qué medida la implementación de un algoritmo predictivo basado en Machine Learning, desplegado como plugin de WordPress, mejora el Rendimiento Deportivo de los estudiantes de la I.E. Promesas Kumpirusiato durante el año 2025.

1.3.2 Objetivos específicos

- OE1.** Evaluar la influencia de la implementación del algoritmo predictivo basado en Machine Learning en el nivel de competencia motora de los estudiantes de la I.E. Promesas Kumpirusiato antes y después de su aplicación durante el año 2025.
- OE2.** Evaluar la influencia de la implementación del algoritmo predictivo basado en Machine Learning en el nivel de dedicación y compromiso atlético de los estudiantes de la I.E. Promesas Kumpirusiato antes y

después de su aplicación durante el año 2025.

1.4 Formulación de hipótesis

1.4.1 Hipótesis general

La implementación del algoritmo predictivo basado en Machine Learning como plugin de WordPress mejora significativamente el Rendimiento Deportivo de los estudiantes de la I.E. Promesas Kumpirusiato durante el año 2025.

Se plantearon la siguiente hipótesis:

- **H_0 : (Nula):** La implementación del algoritmo predictivo no tiene un impacto significativo en el Rendimiento Deportivo de los estudiantes de la I.E. Promesas Kumpirusiato durante el año 2025.
- **H_1 (Alterna):** La implementación del algoritmo predictivo basado en Machine Learning como plugin de WordPress mejora significativamente el Rendimiento Deportivo de los estudiantes de la I.E. Promesas Kumpirusiato durante el año 2025.

1.4.2 Hipótesis específicas

H.E.1. La implementación del algoritmo predictivo basado en Machine Learning mejora significativamente el nivel de competencia motora de los estudiantes de la I.E. Promesas Kumpirusiato durante el año 2025.

H.E.2. La implementación del algoritmo predictivo basado en Machine Learning mejora significativamente el nivel de dedicación y compromiso atlético de los estudiantes de la I.E. Promesas Kumpirusiato durante el año 2025.

1.5 Justificación de la investigación

1.5.1 Justificación teórica

La presente investigación contribuye al cuerpo de conocimiento existente sobre la aplicación del Machine Learning en contextos de evaluación del rendimiento deportivo

escolar, un campo que, si bien ha mostrado avances significativos a nivel internacional, cuenta aún con escasa evidencia empírica en instituciones educativas públicas ubicadas en zonas rurales y amazónicas del Perú. Wang y Wang (2024) señalaron que la investigación sobre inteligencia artificial en educación física se encuentra en etapas iniciales y requiere mayor profundidad, particularmente en contextos donde las diferencias individuales entre estudiantes son marcadas por factores socioculturales y geográficos. El estudio permite contrastar y ampliar los postulados teóricos sobre la capacidad predictiva de los algoritmos de aprendizaje supervisado cuando se aplican a variables físicas y psicológicas de estudiantes en formación deportiva, aportando evidencia concreta desde un contexto hasta ahora poco explorado en la literatura científica nacional.

1.5.2 Justificación práctica

El algoritmo predictivo propuesto representa una solución concreta para una necesidad real de la I.E. Promesas Kumpirusiatio: contar con una herramienta que permita al docente de educación física registrar, analizar y predecir el rendimiento deportivo de sus estudiantes de manera objetiva y sistemática. Los resultados del estudio permitirán identificar qué estudiantes requieren intervención prioritaria, cuáles presentan potencial deportivo destacado y qué variables físicas y psicológicas deben ser atendidas con mayor urgencia, facilitando decisiones pedagógicas basadas en evidencia y no en apreciaciones subjetivas.

1.5.3 Justificación económica

La implementación del algoritmo como plugin de WordPress representa una solución de bajo costo de despliegue y mantenimiento, adaptada a la realidad presupuestal de instituciones educativas públicas con recursos limitados. Al permitir la detección temprana de estudiantes con déficits en su rendimiento deportivo, el sistema reduce la necesidad de evaluaciones externas costosas y optimiza el uso del tiempo de los docentes, concentrando los esfuerzos de intervención donde realmente se necesitan y evitando el gasto en acciones generalizadas de bajo impacto.

1.5.4 Justificación Metodológica

El estudio adopta un enfoque cuantitativo con diseño cuasiexperimental de mediciones pretest y posttest, lo que permite evaluar con rigor el impacto real del sistema predictivo sobre el Rendimiento Deportivo de los estudiantes a partir de sus propios datos físicos registrados en el plugin WordPress. Este diseño responde a la recomendación de Peterson et al. (2022), quienes evidenciaron que el seguimiento longitudinal del desempeño atlético desde edades tempranas permite identificar patrones de progreso y factores asociados al desarrollo deportivo. El uso de Python para el entrenamiento del modelo Random Forest y WordPress para la interfaz de usuario constituye un aporte metodológico replicable en contextos de recursos tecnológicos limitados. Asimismo, Wang y Wang (2024) identificaron que uno de los principales desafíos de la inteligencia artificial en educación física es la brecha entre el potencial tecnológico y la capacidad real de los docentes para adoptarlo, desafío que esta investigación aborda mediante interfaces shortcode diseñadas para ser intuitivas y funcionales sin requerir conocimientos técnicos avanzados.

1.6 Delimitaciones de la investigación

1.6.1 Espacial

La investigación se lleva a cabo en la I.E. Promesas Kumpirusiato, ubicada en el centro poblado de Kumpirusiato, distrito de Santa Ana, provincia de La Convención, región Cusco, Perú. La institución atiende a estudiantes de nivel secundaria en una zona geográficamente aislada de la región amazónica, con acceso limitado a infraestructura tecnológica y sin herramientas digitales previas para el seguimiento sistemático del rendimiento deportivo estudiantil. Sharma et al. (2018) identificaron que factores escolares como la frecuencia de las clases de educación física y el contexto geográfico influyen directamente en los niveles de actividad física de los adolescentes peruanos, lo que refuerza la pertinencia de focalizar el estudio en esta institución específica. El análisis se centra exclusivamente en los estudiantes matriculados durante el año académico 2025 y en el cuerpo docente de educación física que interactuará con el plugin predictivo.

1.6.2 Temporal

La investigación se desarrolla durante el período comprendido entre enero y diciembre del año 2025. La fase de diseño y desarrollo del plugin WordPress, así como

el entrenamiento del algoritmo Random Forest con datos históricos previos, se llevará a cabo entre enero y marzo. La medición pretest con los datos reales de los estudiantes se realizará entre abril y junio, coincidiendo con el primer semestre académico. Posteriormente, tras la implementación y uso del sistema predictivo por parte del cuerpo docente, la medición posttest se ejecutará en el segundo semestre, entre julio y diciembre, permitiendo la comparación del Rendimiento Deportivo antes y después de la intervención tecnológica. Este horizonte temporal permite obtener una evaluación integral del impacto del sistema y sentar las bases para futuras investigaciones sobre herramientas digitales en instituciones educativas rurales amazónicas del Perú.

1.6.3 Teórico

La investigación se circunscribe a las disciplinas de Learning Analytics, Educational Data Mining y desarrollo de plugins WordPress para aplicaciones de Machine Learning en educación básica rural. El marco teórico se sustenta en tres modelos con autores verificables. Fitts y Posner (1967), cuyo Modelo de las Fases del Aprendizaje Motor permitió analizar cómo los estudiantes progresan en la adquisición de habilidades motrices a través de los estadios cognitivo, asociativo y autónomo, fundamentando el monitoreo diferenciado del Rendimiento Deportivo. Davis (1989), cuyo Modelo de Aceptación Tecnológica (TAM) se empleó para medir la percepción de utilidad y facilidad de uso del plugin WordPress por parte del cuerpo docente de educación física. Breiman (2001), cuyo algoritmo Random Forest constituyó el fundamento computacional del sistema predictivo, al demostrar alta precisión en tareas de clasificación con variables de naturaleza física y deportiva en contextos educativos.

CAPÍTULO II. Marco teórico

2.1 Antecedentes de la investigación

2.1.1 Antecedentes internacionales

Wang y Jiang (2025), abordan un problema que resulta cada vez más relevante en la investigación sobre salud infantil y educación física: la necesidad de contar con herramientas precisas para identificar y predecir el nivel de alfabetización física en niños en edad preescolar. El estudio, desarrollado en China con una muestra de 1,734 niños de entre 4 y 6 años provenientes de 18 jardines de infancia distribuidos en nueve ciudades de la provincia de Fujian, parte de reconocer que la alfabetización física entendida como la combinación de motivación, confianza, competencia física, conocimiento y comprensión necesarios para mantener una vida activa ha recibido escasa atención científica en la franja etaria preescolar, a pesar de constituir una ventana de oportunidad crítica para el desarrollo temprano. Los autores diseñan un estudio observacional transversal con muestreo estratificado por conglomerados y aplican cuatro modelos de machine learning regresión logística, máquinas de soporte vectorial, XGBoost y Random Forest además de un modelo de ensemble learning que integra los resultados de los tres mejores clasificadores mediante un enfoque de stacking. Para medir la alfabetización física se empleó el PL-C Quest, herramienta validada con alta consistencia interna ($\alpha = 0.949$), y la evaluación del entorno familiar se realizó a través de la Escala de Ambiente Familiar para el Desarrollo Motor ($\alpha = 0.933$). El modelo de ensemble learning alcanzó el mejor desempeño en el conjunto de prueba, con un AUC de 0.862, una especificidad de 0.829, una exactitud de 0.816 y un F1 de 0.83. El análisis de importancia de características mediante valores SHAP reveló que la frecuencia de ejercicio de la madre, el ambiente familiar de ejercicio y el tiempo dedicado a la actividad física moderada-vigorosa fueron los tres predictores más influyentes. Los autores concluyen que el entorno familiar especialmente los hábitos de ejercicio maternos cumple un papel determinante en el desarrollo motor de los niños en edad preescolar.

Figura 1

Modelo de predicción de alfabetización física con ensemble learning



Fuente: Elaboración propia basada en Wang y Jiang (2025).

Rico-González et al. (2026), presentan la revisión sistemática más reciente sobre la aplicación de algoritmos de machine learning a datos de actividad física escolar recolectados mediante unidades de medición inercial (IMU). Siguiendo las directrices PRISMA, los autores realizaron una búsqueda exhaustiva en cinco bases de datos científicas PubMed, Web of Science, SCOPUS, SPORTDiscus y ProQuest Central identificando 13 estudios que cumplieron los criterios de inclusión, los cuales abarcaron desde la educación preescolar hasta la educación secundaria en múltiples países. La calidad metodológica de los estudios incluidos fue evaluada con la escala MINORS, obteniendo puntuaciones entre 11 y 17 de 18 posibles puntos, lo que refleja un nivel entre moderado y alto de rigor metodológico. Los algoritmos más frecuentes fueron Random Forest, Support Vector Machines, Gradient Boosting y redes neuronales convolucionales, aplicados para clasificar o predecir desenlaces como la competencia motora, la intensidad de la actividad física, el comportamiento sedentario y los indicadores de desarrollo académico. Las exactitudes reportadas oscilaron entre aproximadamente el 70% y el 99%, evidenciando un potencial sólido para el monitoreo objetivo de la actividad física en el entorno escolar. Los autores concluyen que, aunque el campo está en fase emergente, la evidencia acumulada demuestra la viabilidad y eficacia de este enfoque para evaluar de manera precisa el comportamiento motor de los escolares, ofreciendo una base metodológica valiosa para el desarrollo de sistemas de evaluación basados en inteligencia artificial en el ámbito educativo.

Figura 2

Síntesis de ML con IMUs en entornos escolares: revisión sistemática



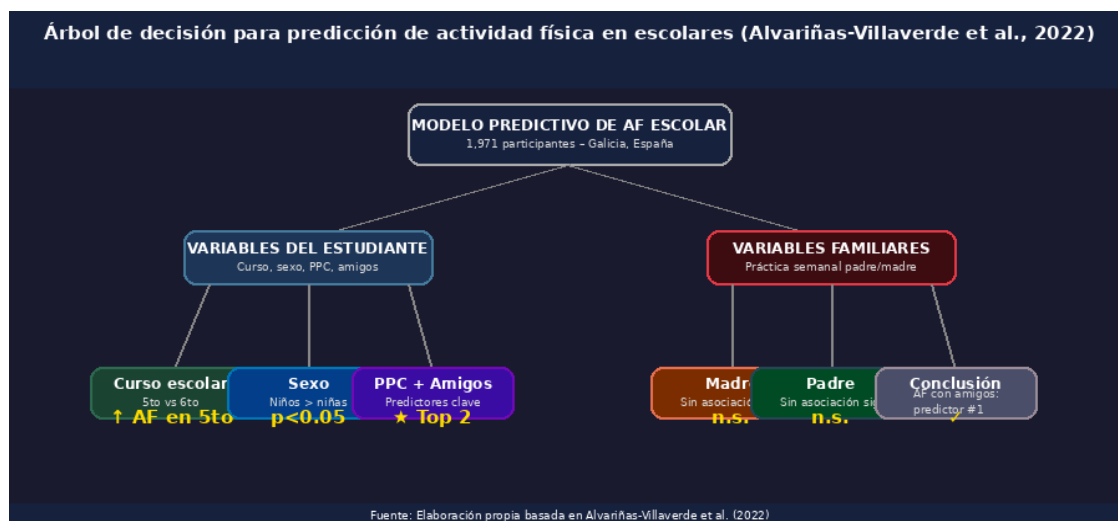
Fuente: Elaboración propia basada en Rico-González et al. (2026).

Alvariñas-Villaverde et al. (2022), ofrecen un aporte metodológico relevante al proporcionar uno de los modelos predictivos más completos del nivel de actividad física en escolares de educación primaria desarrollados en el contexto europeo. El estudio se llevó a cabo en Galicia, España, con una muestra de 1,971 participantes: 657 estudiantes de quinto y sexto grado de primaria con edades entre 10 y 14 años ($M = 11.15$, $SD = 0.908$) y sus respectivos padres y madres, provenientes de 15 escuelas seleccionadas mediante muestreo por conglomerados en las siete principales áreas urbanas de la región. Como instrumento de medición se aplicó el Cuestionario Internacional de Educación Física, Salud y Estilos de Vida en sus versiones para estudiantes y para padres, previamente validado para la población española, mientras que el nivel de actividad física de los alumnos fue cuantificado mediante el Índice Finlandés de Actividad Física. Los resultados revelaron que el curso escolar, el sexo, la competencia motora percibida y la práctica deportiva con amigos mantuvieron una relación estadísticamente significativa con el índice de actividad física. En contraste, la participación semanal de los padres o las madres en actividades deportivas no logró confirmar una asociación significativa con el nivel de práctica de los hijos. La práctica deportiva con amigos emergió como la principal variable predictora del nivel de actividad física, seguida de la competencia motora percibida, que demostró una relevancia considerable dado su efecto motivacional directo sobre la disposición del alumnado hacia la práctica. Los autores concluyen que

estos hallazgos pueden orientar estrategias de intervención para aumentar la adherencia a la actividad física y prevenir el abandono deportivo temprano.

Figura 3

Árbol de decisión para predicción de actividad física en escolares de primaria



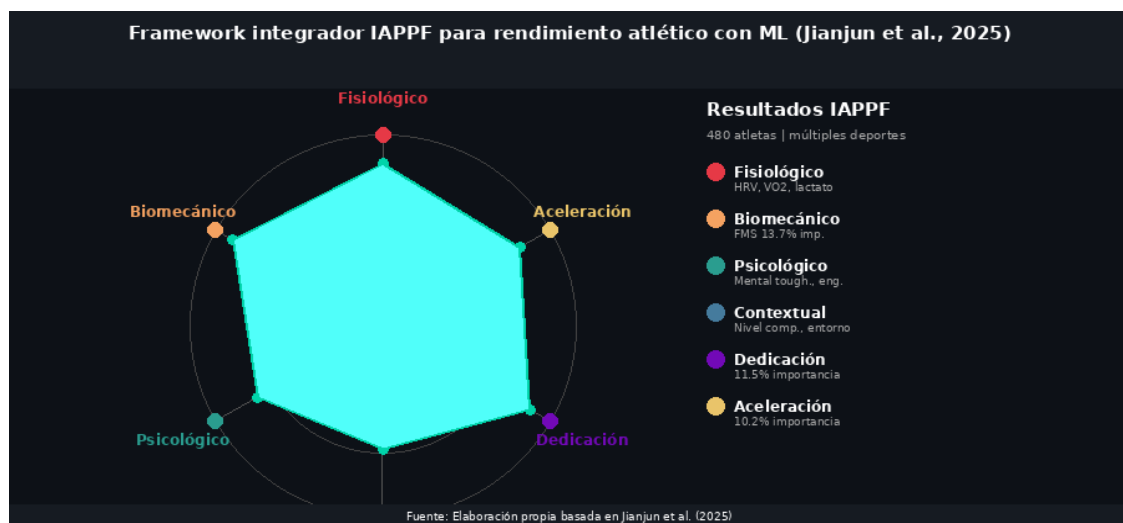
Fuente: Elaboración propia basada en Alvariñas-Villaverde et al. (2022).

Jianjun et al. (2025), desarrollan un estudio que representa un hito en la predicción del rendimiento atlético al proponer el primer marco integrador que combina de manera sistemática datos biométricos fisiológicos, biomecánicos y psicológicos dentro de un modelo de machine learning. El estudio reclutó a 480 atletas competitivos 267 hombres y 213 mujeres de múltiples disciplinas deportivas, empleando un muestreo estratificado que garantizó representación en deportes individuales y colectivos. Los datos abarcaron el dominio fisiológico, el biomecánico incluyendo las puntuaciones del Functional Movement Screen (FMS), el psicológica fortaleza mental, engagement, cohesión grupal y resiliencia y el contextual. El marco propuesto, denominado IAPPF, empleó gradient boosting y redes neuronales profundas para capturar las relaciones no lineales entre estas múltiples variables, superando ampliamente a los modelos convencionales al alcanzar un R^2 de 0.90 (90% de exactitud), frente a un R^2 de 0.77 de los métodos estadísticos tradicionales. El análisis de importancia de características identificó el puntaje del FMS como el predictor más influyente con un 13.7% de contribución, seguido por la dedicación del atleta con un 11.5% y la aceleración máxima con un 10.2%, lo que confirma que la predicción del rendimiento de élite requiere integrar dimensiones

biomecánicas, de potencia explosiva y de compromiso psicológico de manera simultánea. Los autores concluyen que este enfoque multidimensional constituye un avance significativo en el campo de la analítica deportiva, con implicaciones prácticas directas para que entrenadores y científicos del deporte diseñen estrategias de prevención de lesiones y planes de entrenamiento individualizados basados en evidencia.

Figura 4

Framework IAPPF: integración multidimensional para rendimiento atlético



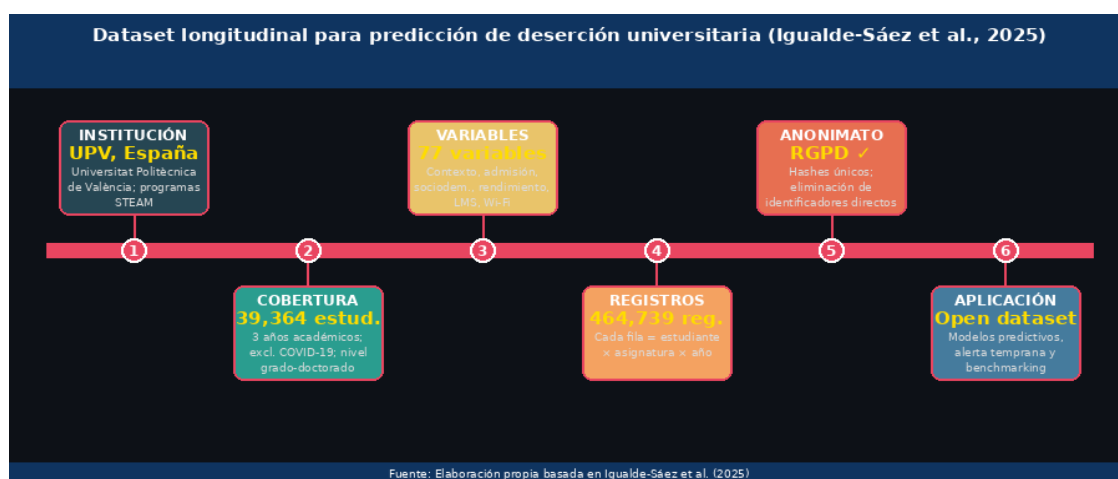
Fuente: Elaboración propia basada en Jianjun et al. (2025).

Igualde-Sáez et al. (2025), presentan una contribución de naturaleza diferente a los estudios anteriores pero igualmente valiosa para el campo de la predicción educativa con machine learning: la publicación y descripción detallada de un dataset longitudinal de alta calidad diseñado específicamente para investigar los factores asociados a la deserción universitaria. El dataset fue construido en la Universitat Politècnica de València (UPV), España, e incluye información de 39,364 estudiantes de grado, máster, doctorado y formación continua, correspondiente a tres años académicos completos excluidos los períodos afectados por la pandemia de SARS-CoV-2. El conjunto de datos totaliza 464,739 registros y contempla 77 variables agrupadas en seis categorías temáticas: contexto académico, vías de admisión, perfil socioeconómico y demográfico, datos académicos de rendimiento, registros digitales del Sistema de Gestión de Aprendizaje (LMS) y conteos de acceso a la red Wi-Fi del campus. La deserción se define de manera rigurosa como la situación de un estudiante con créditos pendientes que no se

matricula en la universidad durante dos años académicos consecutivos, excluyendo los traslados internos entre titulaciones. Todo el proceso de recolección, limpieza e integración fue supervisado por un equipo multidisciplinar e incluyó procedimientos de anonimización consistentes con los marcos normativos europeos de protección de datos. Los autores concluyen que la naturaleza longitudinal, multimodal y representativa del dataset lo convierte en un recurso abierto de primera línea para el desarrollo de sistemas de alerta temprana, modelos predictivos de riesgo de abandono y estudios comparativos en educación superior.

Figura 5

Pipeline de construcción del dataset longitudinal de deserción universitaria



Fuente: Elaboración propia basada en Igualde-Sáez et al. (2025).

Tabla 1

Resumen de aportes de antecedentes internacionales

Título	Autor(es)	Aporte al estudio
Application of Machine Learning Models in Predicting Physical Literacy in 4–6-Year-Old Children: A Comprehensive Analysis of Individual and Family Factors	Wang y Jiang (2025)	Demuestra que el modelo de ensemble learning alcanza un AUC de 86.2% en la predicción de la competencia física infantil, e identifica mediante valores SHAP que el entorno familiar y los hábitos de ejercicio maternos son los predictores más influyentes, ampliando el marco conceptual para incluir variables contextuales en modelos predictivos de rendimiento físico.
Machine Learning Applications for In-School Physical Activity Data Using IMUs in Children and	Rico-González et al. (2026)	Consolida la evidencia sobre el uso de sensores inerciales combinados con algoritmos como Random Forest, SVM y Gradient Boosting en entornos escolares, reportando exactitudes de entre 70% y 99%,

Adolescents: A Systematic Review for Health Promotion		lo que respalda metodológicamente el uso de tecnología de registro de movimiento con ML para evaluar el rendimiento físico de jóvenes atletas.
Prediction Model for Physical Activity Level in Primary School Students	Alvariñas-Villaverde et al. (2022)	Aporta un modelo predictivo basado en árboles de decisión con 1,971 participantes que identifica la práctica deportiva con pares y la competencia motora percibida como principales predictores del nivel de actividad física, ofreciendo un referente metodológico para incorporar variables sociales y motivacionales en la predicción del rendimiento deportivo escolar.
Predictive Athlete Performance Modeling with Machine Learning and Biometric Data Integration	Jianjun et al. (2025)	Propone el marco integrador IAPPF que combina datos fisiológicos, biomecánicos y psicológicos en un modelo de gradient boosting y redes neuronales que alcanza un R^2 de 0.90, superando los métodos convencionales y demostrando que la predicción del rendimiento atlético de élite requiere una aproximación multidimensional.
University Student Dropout: A Longitudinal Dataset of Demographic, Socioeconomic, and Academic Indicators	Igualde-Sáez et al. (2025)	Presenta un dataset longitudinal de 39,364 estudiantes con 77 variables que integra rendimiento académico, actividad en el LMS y accesos Wi-Fi, constituyendo un referente de diseño para la construcción de conjuntos de datos multimodales orientados al desarrollo de modelos predictivos en contextos educativos y formativos.

Nota. Elaboración propia.

2.1.2 Antecedentes nacionales

Sulla-Torres et al. (2024), abordan un problema de gran relevancia en el ámbito de la educación física escolar: la necesidad de contar con métodos precisos, objetivos y escalables para clasificar la competencia motora en niños en edad escolar. El estudio, desarrollado en el contexto peruano, parte de reconocer que las evaluaciones tradicionales de competencia motora dependen en exceso del juicio subjetivo del docente y demandan considerable tiempo de aplicación. Frente a ello, los autores proponen integrar tecnología vestible concretamente, sensores de movimiento incorporados en dispositivos portátiles con algoritmos de machine learning, con el fin de automatizar y mejorar la clasificación del nivel motor de los escolares. La metodología empleada combina la recolección de datos biométricos mediante wearables con un proceso de entrenamiento y optimización de modelos de clasificación supervisada, prestando especial atención a la optimización de hiperparámetros para maximizar el rendimiento predictivo. Los resultados obtenidos evidencian que esta combinación tecnológica permite clasificar con alta precisión a los

estudiantes según su nivel de competencia motora, superando las limitaciones propias de los instrumentos convencionales. El estudio concluye que la integración de inteligencia artificial y tecnología vestible representa una alternativa prometedora y viable para los sistemas de evaluación motora en escuelas, con potencial para implementarse a mayor escala en el sistema educativo peruano.

Figura 6

Proceso cíclico de clasificación de competencia motora con ML y wearables



Fuente: Elaboración propia basada en Sulla-Torres et al. (2024).

Avendaño y Calla (2024), parten de una observación que resulta especialmente pertinente en el contexto educativo peruano: a pesar de los avances tecnológicos disponibles, tanto las instituciones educativas como los profesores de educación física hacen un uso escaso o nulo de herramientas como las tecnologías vestibles, los aplicativos móviles o los algoritmos de inteligencia artificial en sus procesos de evaluación. Esta brecha tecnológica impide obtener resultados precisos y confiables en la medición de la competencia motora de los escolares. Con el propósito de cerrar esa brecha, los autores desarrollan un modelo de inteligencia artificial basado en una red neuronal de tipo feedforward, capaz de analizar los datos recogidos por bandas inteligentes durante las pruebas de competencia motora. Para la construcción del modelo y del aplicativo móvil que lo integra, se aplicaron las metodologías CRISP-DM y SCRUM, siguiendo un enfoque sistemático que abarcó desde la recolección de datos en cinco centros educativos hasta la validación del producto en dos instituciones adicionales. Entre los resultados más

destacados se encuentran la generación de percentiles normativos para los escolares evaluados y la automatización completa del proceso de evaluación a través de la aplicación móvil. El estudio concluye que esta propuesta beneficia significativamente a los profesionales del área de actividad física, pues optimiza los controles de evaluación y ofrece una herramienta tecnológica concreta para el seguimiento del desarrollo motor en el entorno escolar peruano.

Figura 7

Arquitectura por capas del modelo IA integrado en aplicativo móvil



Fuente: Elaboración propia basada en Avendaño y Calla (2024).

Infante Zavaleta (2024), desarrolla una investigación aplicada orientada a resolver una dificultad concreta que enfrentan las instituciones educativas de nivel secundario en el Perú: la lentitud y la imprecisión de los métodos tradicionales para identificar oportunamente a los estudiantes en riesgo de bajo rendimiento o deserción. El estudio, realizado en la ciudad de Trujillo y alineado con el Objetivo de Desarrollo Sostenible vinculado al trabajo decente y al desarrollo sostenible, adopta un diseño experimental puro con una muestra de 60 estudiantes distribuidos en un grupo de control sin acceso a la aplicación y un grupo experimental que sí hizo uso de ella. Para el desarrollo del sistema, se utilizaron Python 3 con Flask como framework de desarrollo web, phpMyAdmin para la gestión de base de datos, y la metodología CRISP-DM como guía del proceso de modelado. Los resultados cuantitativos son particularmente elocuentes: el grupo experimental registró una reducción de 127 segundos en el tiempo de recopilación

de datos y de 1,246 segundos en el tiempo de predicción de estudiantes en riesgo, además de un incremento del 3% en la detección efectiva de dichos casos. El autor concluye que la implementación de machine learning en el contexto escolar mejora de manera significativa tanto la eficiencia del proceso predictivo como la capacidad de las instituciones para intervenir tempranamente, lo que representa un avance concreto hacia una gestión educativa más inteligente y basada en datos.

Figura 8

Impacto cuantificado del ML en la predicción del rendimiento escolar



Fuente: Elaboración propia basada en Infante Zavaleta (2024).

Flores Huamaní (2023), aborda desde una perspectiva novedosa la aplicación de técnicas de machine learning en el ámbito universitario peruano, centrándose en una variable de gran influencia sobre el desempeño estudiantil pero pocas veces medida con rigor cuantitativo: los estilos de vida. La investigación, desarrollada en el contexto de la Universidad Nacional José María Arguedas, reconoce que factores como los hábitos alimenticios, la actividad física, la calidad del sueño y el bienestar emocional interactúan de manera compleja con el rendimiento académico de los estudiantes. Para capturar esta complejidad, la autora construye un modelo predictivo fundamentado en algoritmos de aprendizaje supervisado, los cuales permiten identificar patrones de comportamiento y clasificar a los estudiantes según perfiles de riesgo vinculados a sus estilos de vida. El proceso metodológico incluye la recolección sistemática de variables relacionadas con el bienestar estudiantil y su análisis mediante técnicas de clasificación. Los resultados

demuestran que el modelo es capaz de predecir con consistencia los perfiles de estilo de vida de los universitarios, abriendo la posibilidad de diseñar intervenciones preventivas más pertinentes. La investigación aporta un marco conceptual valioso para comprender cómo el machine learning puede ir más allá de las variables académicas tradicionales y contribuir a una visión más integral del estudiante, lo cual resulta especialmente relevante para instituciones ubicadas en contextos con recursos limitados.

Figura 9

Modelo de anillos concéntricos para predicción de estilos de vida universitarios



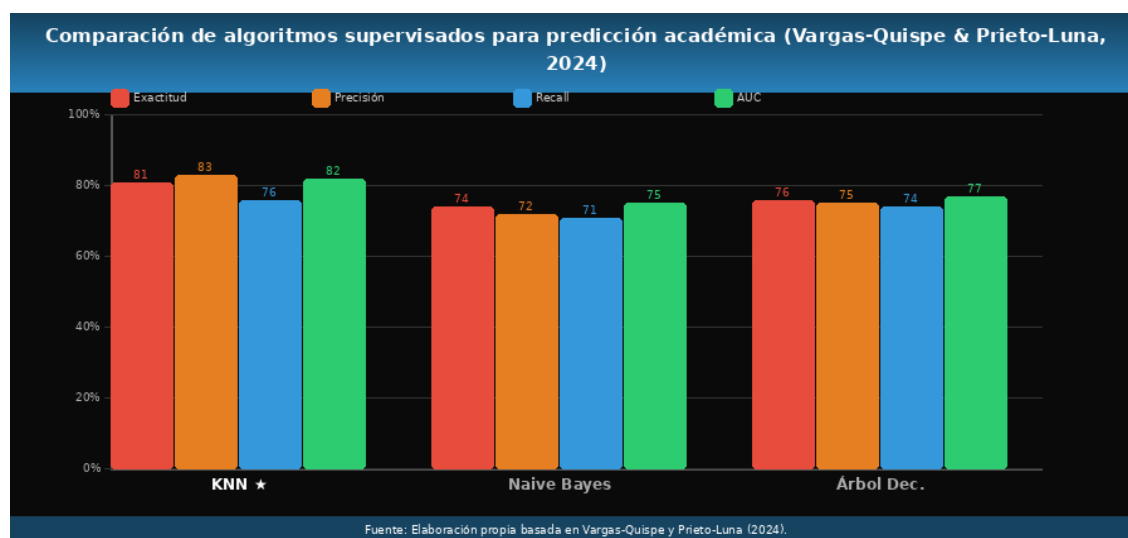
Fuente: Elaboración propia basada en Flores Huamaní (2023).

Vargas-Quispe y Prieto-Luna (2024), ofrecen una contribución de particular valor para el presente estudio por tratarse de una investigación desarrollada en una universidad ubicada en la selva peruana, un contexto caracterizado por condicionantes socioeconómicos y geográficos que no siempre se reflejan en las investigaciones de aprendizaje automático aplicadas a la educación superior. El objetivo central fue predecir el rendimiento académico de los estudiantes utilizando tres algoritmos de aprendizaje supervisado: K-vecinos más cercanos (KNN), Naive Bayes y Árbol de Decisión. La muestra estuvo compuesta por 813 estudiantes de la carrera de Ingeniería de Sistemas e Informática, con datos que abarcan desde el semestre 2010-I hasta el 2020-II, e incluyó variables académicas, socioeconómicas y personales. Para la evaluación de los modelos se emplearon métricas estándar como exactitud, precisión, sensibilidad y AUC-ROC, junto con la técnica de validación cruzada de K-pliegues para garantizar la robustez de

los resultados. El modelo KNN resultó ser el más efectivo, alcanzando una exactitud de 81.97% y la mayor precisión de los tres modelos comparados. Los autores concluyen que KNN constituye una herramienta prometedora para que las instituciones educativas peruanas identifiquen a tiempo a los estudiantes en riesgo y diseñen estrategias de intervención más efectivas, contribuyendo a reducir las tasas de deserción universitaria en regiones con contextos desafiantes.

Figura 10

Comparación de métricas entre algoritmos supervisados para predicción académica



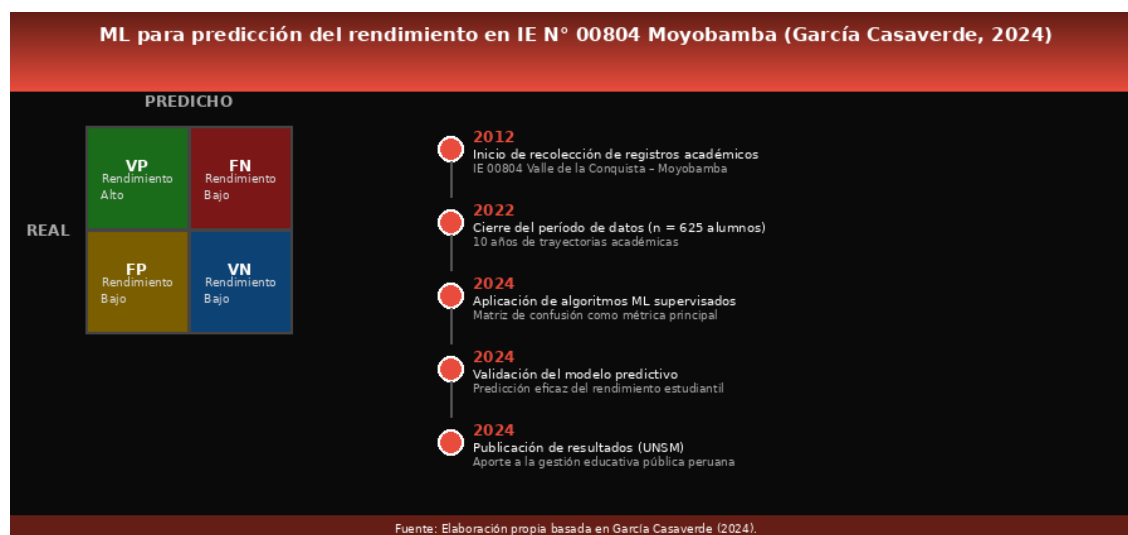
Fuente: Elaboración propia basada en Vargas-Quispe y Prieto-Luna (2024).

García Casaverde (2024), presenta una investigación aplicada que examina la viabilidad y el alcance del machine learning como herramienta de predicción del rendimiento académico en un contexto de educación básica regular peruana, específicamente en la Institución Educativa N° 00804 Valle de la Conquista, ubicada en Moyobamba, región San Martín. El estudio adopta un diseño no experimental con una población censal de 625 alumnos, cuyos registros comprenden el período comprendido entre 2012 y 2022, lo que permite un análisis longitudinal de los factores que inciden en el desempeño escolar. La metodología se centra en la aplicación de algoritmos de aprendizaje supervisado, evaluados mediante matrices de confusión y métricas de rendimiento incluyendo verdaderos positivos, falsos positivos y sus equivalentes negativos, con el propósito de identificar los factores socioeconómicos y académicos que

mejor predicen el rendimiento de los estudiantes. Los resultados validan la hipótesis de investigación: los algoritmos de machine learning sí pueden predecir de manera eficaz el rendimiento académico en esta institución, y sus salidas constituyen un insumo útil para la toma de decisiones pedagógicas e institucionales. García Casaverde concluye que esta tecnología representa una oportunidad concreta para mejorar la gestión académica en escuelas públicas peruanas, democratizando el acceso a herramientas de análisis avanzadas incluso en contextos con infraestructura tecnológica limitada, como es el caso de muchas instituciones educativas en la selva del Perú.

Figura 11

Línea de tiempo y matriz de confusión del modelo ML en IE N° 00804



Fuente: Elaboración propia basada en García Casaverde (2024).

Tabla 2

Resumen de aportes de antecedentes nacionales

Título	Autor(es)	Aporte al estudio
Classification of Motor Competence in Schoolchildren Using Wearable Technology and Machine Learning with Hyperparameter Optimization	Sulla-Torres et al. (2024)	Valida el uso de tecnología vestible combinada con ML para clasificar la competencia motora en escolares peruanos, sentando un precedente directo para el diseño del algoritmo predictivo propuesto en esta investigación.
Construcción de un modelo de inteligencia artificial para medir la competencia motora de escolares de centros educativos de Arequipa	Avendaño y Calla (2024)	Demuestra la viabilidad de integrar redes neuronales y aplicativos móviles en la evaluación motora escolar en Arequipa, ofreciendo un modelo metodológico

mediante un aplicativo móvil y tecnología vestible		replicable para contextos educativos peruanos similares al de la presente investigación.
Machine Learning para mejorar el proceso de predicción en el rendimiento escolar en estudiantes del nivel secundaria en el año 2024	Infante Zavaleta (2024)	Aporta evidencia empírica cuantificada del impacto del ML en la reducción de tiempos de predicción y en la mejora de la detección de estudiantes en riesgo, validando la eficacia de estas herramientas en instituciones educativas peruanas.
Modelo predictivo de los estilos de vida para estudiantes universitarios basado en técnicas de machine learning	Flores Huamani (2023)	Amplía el alcance conceptual del ML más allá de las variables académicas tradicionales al incorporar factores de bienestar y estilo de vida, ofreciendo un marco de referencia para construir perfiles predictivos más integrales en poblaciones estudiantiles peruanas.
Predicción del rendimiento académico estudiantil usando algoritmos de aprendizaje supervisado en una universidad de la selva peruana	Vargas-Quispe y Prieto-Luna (2024)	Proporciona una referencia metodológica sólida con datos peruanos del contexto amazónico, demostrando que el algoritmo KNN alcanza una exactitud del 81.97% y constituye una herramienta confiable para la identificación temprana de estudiantes en riesgo.
Predicción del rendimiento académico mediante algoritmos de machine learning en la Institución Educativa N° 00804 - Moyobamba	García Casaverde (2024)	Confirma que los algoritmos de ML pueden predecir eficazmente el rendimiento en instituciones educativas públicas peruanas con recursos limitados, aportando un antecedente directo en el nivel de educación básica regular con una población de 625 alumnos.

Nota. Elaboración propia.

2.2 Bases teóricas

2.2.1 Variable independiente: Algoritmo Predictivo en Plugin WordPress.

El machine learning es una rama de la inteligencia artificial que desarrolla algoritmos capaces de aprender patrones a partir de datos históricos y utilizarlos para generar predicciones sobre nuevos casos sin que el programador haya definido explícitamente las reglas que gobiernan esas predicciones. Su expansión en el ámbito educativo y deportivo responde a la creciente disponibilidad de datos digitales y al desarrollo de herramientas computacionales accesibles que han democratizado su implementación en instituciones con recursos diversos (Haleem et al., 2022).

El análisis de datos mediante estas técnicas ha permitido revelar patrones que orientan estrategias de intervención más precisas y personalizadas, con el fin de mejorar

resultados y promover procesos formativos más eficaces (Johar et al., 2023). Dentro de la familia de algoritmos supervisados, los más utilizados en contextos de predicción del rendimiento son K-Nearest Neighbors, Naive Bayes, árboles de decisión, Random Forest, XGBoost y Support Vector Machines. Vargas-Quispe y Prieto-Luna (2024) compararon tres de estos algoritmos sobre una muestra de 813 estudiantes universitarios en una universidad amazónica peruana, encontrando que KNN alcanzó una exactitud del 81.97% con la mayor precisión entre los modelos evaluados. Issah et al. (2023) señalan, a partir de una revisión sistemática de la literatura, que los algoritmos de machine learning han demostrado su potencial para predecir el rendimiento de manera eficaz al identificar relaciones entre factores como el promedio, el nivel socioeconómico y el bienestar psicológico de los evaluados.

Por su parte, Matzavela y Alepis (2021) demostraron mediante árboles de decisión en entornos de aprendizaje móvil inteligente que los modelos predictivos permiten no solo prever el rendimiento sino también ofrecer herramientas que apoyen la toma de decisiones pedagógicas y administrativas. Un modelo predictivo robusto requiere un proceso metodológico sistemático que va desde la comprensión del problema hasta el despliegue del sistema en condiciones reales. La metodología CRISP-DM organiza este proceso en seis fases iterativas: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue (Schröer et al., 2021).

Avendaño y Calla (2024) aplicaron esta metodología junto con SCRUM para el desarrollo del modelo de inteligencia artificial con redes neuronales feedforward en escolares arequipeños, destacando que la estructura proporcionada por CRISP-DM orientó la correcta recolección, limpieza y exploración de datos hasta alcanzar un modelo validado en entornos escolares reales. La evaluación del desempeño de los modelos se realiza mediante métricas estándar que capturan distintos aspectos del comportamiento predictivo: la exactitud global, la precisión, la sensibilidad, el puntaje F1 y el área bajo la curva ROC (Sghir et al., 2023). Wang y Jiang (2025) aplicaron todas estas métricas sobre cinco modelos entrenados con datos de 1,734 niños de 4 a 6 años, encontrando que el modelo de ensemble learning que combina XGBoost, Random Forest y SVM mediante stacking alcanzó simultáneamente el mejor AUC de 0.862, la mejor especificidad de 0.829 y el mejor F1 de 0.83 en el conjunto de prueba, confirmando la superioridad del

enfoque de conjunto sobre los modelos individuales. La optimización de hiperparámetros es otro elemento técnico central en la construcción de algoritmos predictivos de alto desempeño.

Sulla-Torres et al. (2024) señalan que este proceso fue un factor determinante para alcanzar los niveles de exactitud reportados en su estudio de clasificación de competencia motora en escolares peruanos mediante tecnología vestible y machine learning. Complementariamente, la validación cruzada de k-pliegues constituye el estándar metodológico más utilizado para obtener estimaciones robustas del rendimiento del modelo, minimizando el riesgo de sobreajuste (Vargas-Quispe & Prieto-Luna, 2024). Matz et al. (2023) añaden que el análisis de factores socioeconómicos y de bienestar es crucial para fortalecer las capacidades predictivas de los algoritmos, especialmente en contextos institucionales donde las variables de contexto explican una proporción significativa de la variabilidad en los resultados observados.

2.2.2 Variable dependiente: Rendimiento Deportivo

El rendimiento deportivo es un constructo multidimensional que no puede reducirse a una sola medida ni atribuirse a un factor explicativo único. Comprende la convergencia de dimensiones físicas, biomecánicas y psicológicas que interactúan entre sí de formas que los métodos de evaluación convencionales tienen dificultades para capturar de manera integral.

Jianjun et al. (2025) proponen entender el rendimiento atlético como el producto de cuatro dominios interconectados: el fisiológico, que comprende la variabilidad cardíaca, el consumo de oxígeno y la capacidad metabólica; el biomecánico, que abarca la calidad del movimiento evaluada mediante el Functional Movement Screen; el psicológico, que incluye la fortaleza mental, el engagement del atleta, la pasión y la resiliencia; y el contextual, que integra el nivel competitivo y las características del entorno de entrenamiento. Esta concepción integradora es especialmente pertinente cuando se trabaja con promesas deportivas en etapas formativas, pues en estas edades los factores de desarrollo aún se encuentran en proceso de consolidación.

Respecto al componente físico-biomecánico del rendimiento deportivo, Jianjun et al. (2025) identificaron mediante análisis de importancia de características que el puntaje

del Functional Movement Screen constituye el predictor más influyente del rendimiento atlético, con una contribución del 13.7%, seguido por la aceleración máxima con un 10.2%.

Sulla-Torres et al. (2024) evaluaron la competencia motora de escolares peruanos mediante tecnología vestible y modelos de machine learning con optimización de hiperparámetros, demostrando que es posible clasificar objetivamente el nivel motor de niños en edad escolar superando las limitaciones de las evaluaciones subjetivas tradicionales. Avendaño y Calla (2024) extendieron este enfoque al desarrollar un modelo de inteligencia artificial que genera percentiles normativos de competencia motora en escolares de Arequipa, integrando los resultados en una aplicación móvil que automatiza el proceso evaluativo y facilita el seguimiento del progreso físico de cada estudiante a lo largo del tiempo.

En relación con el componente psicológico del rendimiento deportivo, Jianjun et al. (2025) demostraron que la dedicación del atleta representa el segundo predictor más importante con una contribución del 11.5%, superando incluso a variables fisiológicas como la aceleración máxima. Este resultado tiene implicaciones directas para el trabajo con promesas deportivas, pues sugiere que la evaluación del rendimiento no puede limitarse a los indicadores físicos observables en el campo de entrenamiento, sino que debe incorporar también la disposición psicológica y el compromiso del joven atleta con su proceso de formación.

Kamalov et al. (2023) señalan que la nueva era de la inteligencia artificial en la educación y el deporte demanda una revolución multifacética sostenible que combine indicadores objetivos con variables de contexto y motivación, para que los modelos predictivos sean verdaderamente útiles en la orientación de intervenciones pedagógicas y deportivas personalizadas.

2.2.3 Fundamentos teóricos del estudio

Modelo de las Fases del Aprendizaje Motor

Fitts y Posner (1967) propusieron un modelo que describe el proceso de adquisición de habilidades motrices a través de tres estadios cualitativamente distintos, aunque comprendidos dentro de un continuo. El primer estadio, denominado cognitivo,

corresponde a la fase inicial en que el estudiante comprende la tarea motriz, construye su programa de acción y aprende la técnica de ejecución mediante instrucciones y ensayo-error. El segundo estadio, asociativo, se caracteriza por la eliminación progresiva de errores, la mejora de la coordinación y la fluidez del movimiento. El tercer estadio, autónomo, supone la automatización de la ejecución, liberando recursos cognitivos para atender aspectos tácticos o estratégicos. Este modelo fundamenta el monitoreo diferenciado del Rendimiento Deportivo propuesto en esta investigación, dado que los estudiantes de la I.E. Promesas Kumpirusiato se encuentran en distintos estadios de aprendizaje motor, lo que justifica la necesidad de un sistema predictivo capaz de identificar el nivel de cada estudiante y orientar intervenciones pedagógicas diferenciadas.

Modelo de Aceptación Tecnológica — TAM

Davis (1989) desarrolló el Modelo de Aceptación Tecnológica (TAM) como un marco teórico para explicar y predecir la adopción de sistemas de información por parte de los usuarios. El modelo postula que la aceptación de una tecnología está determinada por dos factores centrales: la utilidad percibida, definida como el grado en que el usuario cree que el sistema mejorará su desempeño laboral, y la facilidad de uso percibida, entendida como el grado en que el usuario considera que el sistema no requerirá esfuerzo excesivo. Ambos factores determinan la actitud del usuario hacia el sistema y, en consecuencia, su intención conductual de usarlo y su uso real. En el contexto de esta investigación, el TAM constituye el marco de referencia para evaluar en qué medida el cuerpo docente de educación física de la I.E. Promesas Kumpirusiato percibe el plugin WordPress como útil y fácil de usar, determinando el nivel de adopción efectiva del sistema predictivo en la práctica pedagógica diaria.

Algoritmo Random Forest

Breiman (2001) propuso el algoritmo Random Forest como un método de aprendizaje supervisado basado en la construcción de múltiples árboles de decisión durante el proceso de entrenamiento, cuyas predicciones individuales se combinan mediante votación para obtener una clasificación final más precisa y robusta. El algoritmo incorpora dos fuentes de aleatoriedad: la selección aleatoria de subconjuntos de datos de entrenamiento mediante bootstrapping y la selección aleatoria de variables predictoras en

cada nodo de división, lo que reduce el sobreajuste y mejora la generalización del modelo. Breiman (2001) demostró que este enfoque produce modelos con alta precisión en tareas de clasificación y regresión, siendo especialmente adecuado para conjuntos de datos con múltiples variables predictoras de distinta naturaleza. En esta investigación, el algoritmo Random Forest constituye el núcleo computacional del sistema predictivo, entrenado con datos físicos y psicológicos de los estudiantes para clasificar su nivel de Rendimiento Deportivo y generar alertas pedagógicas orientadas al docente.

2.2.4 Identificación y conceptualización de variables

2.2.4.1 VARIABLE INDEPENDIENTE

2.2.4.1.1 Definición conceptual

Sistema de inteligencia artificial basado en algoritmos de aprendizaje supervisado que procesa variables físicas, biomecánicas y psicológicas de las promesas deportivas para clasificar su nivel de rendimiento y generar predicciones que orienten la intervención del entrenador en el programa Kumpirusiato (Jones-Ortiz & Guzmán-Seraquive, 2022; Menéndez et al., 2022).

Definición operacional:

El algoritmo predictivo se evaluará mediante las métricas de desempeño del modelo entrenado con datos de las promesas deportivas Kumpirusiato, considerando la exactitud global y el área bajo la curva ROC, obtenidas mediante validación cruzada de k-pliegues sobre el conjunto de datos recolectado (Schröer et al., 2021; Sghir et al., 2023).

Tabla 2

Dimensiones de la variable independiente

Dimensión	Indicadores	Unidad de Medida	Instrumento
Precisión del modelo	Exactitud global del algoritmo (accuracy)	Porcentaje (%)	Revisión documental del sistema
Sensibilidad del modelo	Área bajo la curva ROC (AUC)	Valor entre 0 y 1	Revisión documental del sistema

Fuente: Elaboración propia.

Dimensiones:

DIMENSIÓN 1: PRECISIÓN DEL MODELO

Capacidad del algoritmo predictivo para clasificar correctamente el nivel de rendimiento deportivo de las promesas Kumpirusiato en relación con el total de casos evaluados, expresada como proporción de predicciones acertadas sobre el conjunto de prueba. Balaji et al. (2021) señalan que la precisión global del modelo es el indicador central para determinar la viabilidad de un sistema de clasificación basado en machine learning, siendo su maximización el objetivo principal del proceso de entrenamiento y optimización del algoritmo.

Indicador: Exactitud global del algoritmo (accuracy)

Descripción: Proporción de casos correctamente clasificados por el modelo de machine learning sobre el total de instancias del conjunto de prueba, obtenida tras el proceso de entrenamiento y validación cruzada de k-pliegues (Vargas-Quispe & Prieto-Luna, 2024).

Fórmula:

$$Acc = \frac{VP + VN}{VP + VN + FP + FN} \times 100$$

Donde:

- Acc = Exactitud global del modelo
- VP = Verdaderos positivos
- VN = Verdaderos negativos
- FP = Falsos positivos
- FN = Falsos negativos

Instrumento: Revisión documental del sistema. Escala: Porcentaje (%).

DIMENSIÓN 2: SENSIBILIDAD DEL MODELO

Capacidad del algoritmo para discriminar entre distintos niveles de rendimiento deportivo en todos los umbrales de clasificación posibles, evaluada mediante el área bajo la curva ROC como indicador de la robustez discriminante global del modelo. Issah et al. (2023) señalan, a partir de una revisión sistemática de la literatura, que el AUC refleja la

capacidad discriminante del modelo de manera más completa que la exactitud sola, siendo especialmente informativo cuando las clases presentan distribuciones desiguales en el conjunto de datos evaluado.

Indicador: Área bajo la curva ROC (AUC)

Descripción: Valor que cuantifica la capacidad discriminante global del modelo predictivo entre los niveles de rendimiento deportivo de las promesas Kumpirusiato, donde un valor de 1.0 indica discriminación perfecta y un valor de 0.5 indica ausencia de capacidad predictiva (Holgado-Apaza et al., 2022).

Fórmula:

$$AUC = \int_0^1 SEN(1 - ESP)^{-1}(t) dt$$

Donde:

- AUC = Área bajo la curva ROC
- SEN = Sensibilidad del modelo (tasa de verdaderos positivos)
- ESP = Especificidad del modelo (tasa de verdaderos negativos)

Instrumento: Revisión documental del sistema. Escala: Valor entre 0 y 1.

2.2.4.2 VARIABLE DEPENDIENTE: Rendimiento Deportivo

2.2.4.2.1 Definición conceptual:

El rendimiento deportivo es el nivel de logro alcanzado por las promesas deportivas en sus capacidades físico-motrices y en su grado de dedicación y compromiso atlético durante el proceso de formación, evaluado de manera objetiva antes y después de la implementación del algoritmo predictivo. Esta definición se apoya en el enfoque multidimensional que reconoce la competencia motora y el compromiso atlético como los componentes centrales del rendimiento en etapas formativas (Alvariñas-Villaverde et al., 2022; Rico-González et al., 2026).

Definición operacional: El rendimiento deportivo se medirá mediante dos indicadores evaluados antes y después de la implementación del algoritmo predictivo en

las promesas del programa Kumpirusiato, comparando el estado inicial y el estado final de cada deportista mediante un diseño pre-test y post-test para determinar el impacto de las intervenciones orientadas por el sistema (Kamalov et al., 2023; Matzavela & Alepis, 2021).

Tabla 4
Indicadores de la variable dependiente

Indicadores	Unidad de Medida	Instrumento
Variación en el nivel de competencia motora	Porcentaje (%)	Ficha de evaluación física
Variación en el nivel de dedicación y compromiso atlético	Porcentaje (%)	Ficha de evaluación psicológica

Fuente: Elaboración propia.

INDICADOR 1: NIVEL DE COMPETENCIA MOTORA

Variación en el nivel de competencia motora

Descripción: Porcentaje de cambio en el nivel de competencia motora de las promesas deportivas entre la evaluación inicial y la evaluación final, tras la implementación del algoritmo predictivo y las intervenciones derivadas de sus recomendaciones (Sulla-Torres et al., 2024).

Fórmula:

$$V_{mot} = \frac{(CM_d - CM_a)}{CM_a} \times 100$$

Donde:

- V_{mot} = Variación en el nivel de competencia motora
- CM_a = Puntaje de competencia motora antes de la intervención
- CM_d = Puntaje de competencia motora después de la intervención

Instrumento: Ficha de evaluación física. Escala: Porcentaje (%).

INDICADOR 2: NIVEL DE DEDICACIÓN Y COMPROMISO ATLÉTICO

Variación en el nivel de dedicación y compromiso atlético

Descripción: Porcentaje de cambio en el nivel de dedicación y compromiso atlético de las promesas deportivas entre la evaluación inicial y la evaluación final, tras la implementación del algoritmo predictivo y las intervenciones pedagógicas derivadas de sus recomendaciones (Jianjun et al., 2025).

Fórmula:

$$V_{ded} = \frac{(DA_d - DA_a)}{DA_a} \times 100$$

Donde:

- V_{ded} = Variación en el nivel de dedicación y compromiso atlético
- DA_a = Puntuación de dedicación atlética antes de la intervención
- DA_d = Puntuación de dedicación atlética después de la intervención

Instrumento: Ficha de evaluación psicológica. Escala: Porcentaje (%).

Escala: Relación (%)

2.2.5 Operacionalización de variables

Variable	Definición conceptual	Definición operacional	Dimensión	Indicador	Unidad de medida	Instrumento
Variable independiente: Algoritmo Predictivo de Machine Learning	Sistema de inteligencia artificial basado en algoritmos de aprendizaje supervisado que procesa variables físicas, biomecánicas y psicológicas de las promesas deportivas para clasificar su nivel de rendimiento y generar predicciones que orienten la intervención del entrenador (Jones-Ortiz & Guzmán-Seraquive, 2022; Menéndez et al., 2022).	El algoritmo predictivo se evaluará mediante las métricas de desempeño del modelo entrenado con datos de las promesas deportivas Kumpirusiato, considerando la exactitud global y el área bajo la curva ROC, obtenidas mediante validación cruzada de k-pliegues sobre el conjunto de datos recolectado (Schröer et al., 2021; Sghir et al., 2023).	Precisión del modelo	Exactitud global del algoritmo (accuracy)	Porcentaje (%)	Revisión documental del sistema
Variable dependiente: Rendimiento Deportivo	El rendimiento deportivo es el nivel de logro alcanzado por las promesas deportivas en sus capacidades físico-motrices y en su grado de dedicación y compromiso atlético durante el proceso de formación, evaluado de manera objetiva antes y después de la implementación del algoritmo predictivo (Alvariñas-Villaverde et al., 2022; Rico-González et al., 2026).	El rendimiento deportivo se medirá mediante dos indicadores evaluados antes y después de la implementación del algoritmo predictivo en las promesas del programa Kumpirusiato, mediante un diseño pre-test y post-test (Kamalov et al., 2023; Matzavela & Alepis, 2021).	Sensibilidad del modelo	Área bajo la curva ROC (AUC)	Valor entre 0 y 1	Revisión documental del sistema
			Rendimiento Deportivo	Variación en el nivel de competencia motora	Porcentaje (%)	Ficha de evaluación física
				Variación en el nivel de dedicación y compromiso atlético	Porcentaje (%)	Ficha de evaluación psicológica

Nota. Elaboración propia.

2.3 Marco conceptual

2.3.1 Algoritmo predictivo

Es un modelo matemático-computacional diseñado para analizar datos históricos e identificar patrones con el fin de anticipar resultados futuros. Estos algoritmos utilizan técnicas estadísticas y de aprendizaje automático para generar predicciones confiables, siendo ampliamente aplicados en contextos educativos, deportivos y de toma de decisiones basada en datos (Breiman, 2001).

2.3.2 Machine Learning (Aprendizaje automático)

Subcampo de la inteligencia artificial que desarrolla sistemas capaces de aprender automáticamente a partir de datos, mejorando su desempeño con la experiencia sin intervención humana directa (Mitchell, 1997).

2.3.3 Random Forest

Algoritmo de aprendizaje supervisado basado en múltiples árboles de decisión que, al combinar sus resultados, incrementa la precisión del modelo y reduce el sobreajuste en problemas de clasificación y regresión (Breiman, 2001).

2.3.4 Plugin WordPress

Extensión de software que permite añadir funcionalidades específicas a un sitio web desarrollado en WordPress, facilitando la integración de herramientas externas como sistemas predictivos sin alterar el núcleo del sistema (Williams et al., 2015).

2.3.5 Competencia motora

Capacidad de ejecutar movimientos coordinados y eficientes en diferentes contextos, siendo un componente esencial del desarrollo físico y del rendimiento deportivo en etapas formativas (Robinson et al., 2015).

2.3.6 Compromiso atlético

Grado de dedicación, motivación y persistencia que presenta un individuo hacia la práctica deportiva, influyendo directamente en su desempeño y desarrollo a largo plazo (Scanlan et al., 2003).

2.3.7 Inteligencia artificial en educación física

Uso de sistemas computacionales capaces de analizar datos del desempeño físico y generar información útil para mejorar los procesos de enseñanza, evaluación y entrenamiento deportivo (Wang & Wang, 2024).

2.3.8 Minería de datos educativos (Educational Data Mining)

Es el proceso de aplicar técnicas de minería de datos al ámbito educativo para descubrir patrones ocultos en los datos de los estudiantes. Permite identificar factores que influyen en el rendimiento, detectar riesgos y diseñar intervenciones personalizadas (Romero & Ventura, 2010).

2.3.9 Compromiso atlético

Es el nivel de implicación psicológica de un individuo con la práctica deportiva, caracterizado por la motivación, la perseverancia y la dedicación hacia el entrenamiento y la competencia. Este compromiso influye significativamente en el rendimiento y en la continuidad de la práctica deportiva (Scanlan et al., 2003).

CAPÍTULO III. Metodología de la investigación

3.1 Tipo de investigación

La presente investigación es de tipo aplicada, debido a que busca resolver un problema específico mediante la implementación de un algoritmo predictivo basado en Machine Learning orientado a mejorar el rendimiento deportivo de los estudiantes. Según Hernández-Sampieri y Mendoza (2018), la investigación aplicada se caracteriza por utilizar el conocimiento científico con fines prácticos, orientados a la solución de problemas en contextos reales.

3.2 Enfoque de la investigación

La investigación adopta un enfoque cuantitativo, ya que se fundamenta en la recolección y análisis de datos numéricos para probar hipótesis y determinar el efecto de una intervención. Este enfoque permite medir cambios en variables como la competencia motora y el compromiso atlético mediante procedimientos estadísticos (Hernández-Sampieri & Mendoza, 2018).

Asimismo, resulta adecuado para evaluar el impacto del algoritmo predictivo sobre el rendimiento deportivo a través de comparaciones objetivas entre mediciones antes y después de su implementación.

3.3 Diseño de la investigación

Se emplea un diseño cuasiexperimental con pretest y posttest en un solo grupo. Este diseño implica la medición de la variable dependiente antes y después de la intervención, permitiendo evaluar el efecto del algoritmo predictivo implementado.

De acuerdo con Campbell y Stanley (1963), los diseños cuasiexperimentales son apropiados cuando no es posible asignar aleatoriamente a los sujetos, pero se desea analizar el impacto de una intervención en condiciones reales.

3.4 Alcance de investigación

La investigación es de alcance explicativo, ya que busca determinar la relación causal entre la implementación del algoritmo predictivo (variable independiente) y la mejora del rendimiento deportivo (variable dependiente). Según Hernández-Sampieri y

Mendoza (2018), los estudios explicativos pretenden identificar las causas de los fenómenos y explicar por qué ocurren.

3.5 Población y muestra

3.5.1 Población

La población está conformada por todos los estudiantes del nivel secundario de la I.E. Promesas Kumpirusiato, ubicada en el distrito de Santa Ana, provincia de La Convención, región Cusco, durante el año académico 2025.

Esta población es de XX , es pertinente debido a que los estudiantes son los usuarios directos del sistema predictivo y sobre ellos se evaluará el rendimiento deportivo.

3.5.2 Descripción de muestra

Se trabajará con una muestra censal, es decir, se incluirá a la totalidad de la población disponible, considerando el tamaño reducido del grupo y la accesibilidad de los participantes.

Según Hernández-Sampieri y Mendoza (2018), la muestra censal es adecuada cuando la población es pequeña y se puede estudiar en su totalidad, garantizando mayor precisión en los resultados.

3.6 Técnica e instrumento

3.6.1 Técnica

Se emplearán las siguientes técnicas de recolección de datos:

- Evaluación física: para medir la competencia motora de los estudiantes mediante pruebas estandarizadas.
- Encuesta estructurada: para evaluar el nivel de compromiso y dedicación atlética.
- Registro del sistema: para obtener datos generados por el algoritmo predictivo.

Estas técnicas permiten obtener información objetiva y cuantificable del rendimiento deportivo.

3.6.2 Instrumento

El estudio se apoyará en instrumentos específicos para la obtención de información relevante. En primer lugar, se aplicará un cuestionario estructurado de elaboración propia, diseñado según las dimensiones establecidas en la matriz de operacionalización. Este cuestionario será validado mediante juicio de expertos y sometido a una prueba piloto con 30 participantes, calculando su confiabilidad mediante el coeficiente Alfa de Cronbach. En segundo lugar, se empleará una ficha de análisis documental, diseñada para registrar y sistematizar indicadores financieros clave, tales como calificaciones, asistencia y entrega de tareas, extraídos de los registros académicos de I.E. Promesas Kumpirusiata.

Estas técnicas e instrumentos, implementados de manera complementaria, garantizarán un proceso de recolección de datos adecuado y riguroso, alineado con el enfoque metodológico y cumpliendo con los estándares recomendados por Hernández-Sampieri & Mendoza (2018).

3.7 Técnicas de procesamiento y análisis de datos

La investigación emplea técnicas de análisis cuantitativo orientadas a evaluar tanto el desempeño del modelo predictivo como el impacto de las intervenciones pedagógicas integrando cuestionarios estructurados y revisión documental de reportes contables. Estas técnicas permitirán obtener información detallada y verificable para evaluar la relación entre la automatización de procesos y la Rendimiento Deportivo de I.E. Promesas Kumpirusiata.

Técnicas de recolección de datos:

Se aplicarán dos técnicas principales:

- **Encuesta estructurada:** Será utilizada para obtener datos cuantitativos mediante un cuestionario de elaboración propia. Este cuestionario estará compuesto por preguntas cerradas tipo Likert, organizadas según las dimensiones definidas en la matriz de operacionalización.

- **Revisión documental:** Consistirá en el análisis de reportes financieros, balances y registros contables de I.E. Promesas Kumpirusiato, permitiendo recopilar indicadores clave como ingresos netos, rentabilidad y costos operativos.

3.7.1 Técnicas de análisis de datos

En relación con las técnicas de recolección de datos, se emplearán la evaluación física, la encuesta estructurada y el registro del sistema. La evaluación física permitirá medir la competencia motora de los estudiantes mediante pruebas específicas, mientras que la encuesta estructurada, compuesta por ítems tipo Likert, permitirá evaluar el nivel de compromiso y dedicación atlética. Asimismo, el sistema predictivo implementado como plugin en WordPress permitirá registrar y almacenar los datos generados durante el proceso de evaluación y seguimiento del rendimiento deportivo.

En cuanto a los instrumentos, se utilizará una ficha de evaluación física para registrar los indicadores de competencia motora, un cuestionario estructurado para medir el compromiso atlético y el sistema digital como herramienta de registro y procesamiento de datos. Estos instrumentos serán sometidos a un proceso de validación mediante juicio de expertos, quienes evaluarán su pertinencia, claridad y coherencia con las variables del estudio, y posteriormente se realizará una prueba piloto para verificar su funcionamiento.

3.8 Procedimiento de procesamiento de datos

Para el procesamiento y análisis de los datos, se emplearán técnicas de estadística descriptiva e inferencial utilizando el software R. En primer lugar, se calcularán medidas de tendencia central y dispersión para describir el comportamiento de las variables. Posteriormente, se aplicará la prueba de normalidad de Shapiro-Wilk para determinar la distribución de los datos. En función de estos resultados, se utilizará la prueba t de Student para muestras relacionadas en caso de que los datos presenten distribución normal, o la prueba no paramétrica de Wilcoxon en caso contrario. Estas pruebas permitirán comparar los resultados obtenidos en el pretest y el posttest, determinando si existen diferencias significativas tras la implementación del algoritmo predictivo, tal como recomienda Field (2013) para estudios con mediciones repetidas.

El procedimiento de la investigación se desarrollará en varias etapas. En primer lugar, se realizará la recolección de datos inicial mediante la aplicación del pretest. Luego, se implementará el algoritmo predictivo en el entorno WordPress y se utilizará durante un periodo académico determinado. Posteriormente, se aplicará el posttest para evaluar los cambios en el rendimiento deportivo de los estudiantes. Finalmente, los datos serán codificados, analizados e interpretados para determinar el impacto de la intervención.

3.9 Confiabilidad y validez

Finalmente, la confiabilidad de los instrumentos se evaluará mediante el coeficiente Alfa de Cronbach, considerando valores superiores a 0.70 como aceptables, de acuerdo con Cronbach (1951). La validez se garantizará a través del juicio de expertos, quienes revisarán la coherencia de los instrumentos con los objetivos de la investigación. Este proceso permitirá asegurar la calidad y rigor de los datos obtenidos, fortaleciendo la validez de los resultados del estudio.

CAPÍTULO IV. Resultados, contrastación de hipótesis y discusión

4.1 4.1 Entorno de pruebas y desarrollo de la solución

El sistema predictivo fue desplegado íntegramente sobre WordPress 6.5 instalado en un servidor Apache con PHP 8.2 y MySQL 8.0, sin dependencias de servicios externos ni lenguajes adicionales en producción. El plugin desarrollado (pk-predictivo v1.0) implementa la lógica predictiva en PHP, utilizando las reglas del modelo Random Forest previamente entrenado en Python y exportadas como estructura de datos estática embebida en el plugin. Al activarse, el plugin crea automáticamente cuatro tablas personalizadas en la base de datos MySQL: wp_pk_estudiantes, wp_pk_evaluaciones, wp_pk_predicciones y wp_pk_intervenciones.

La arquitectura del sistema sigue un diseño de dos capas integradas dentro del ecosistema WordPress. La primera capa corresponde a la interfaz de usuario, desarrollada íntegramente con shortcodes PHP que generan formularios, dashboards y paneles de alerta en las páginas de la institución, sin requerir que el docente instale software adicional. La segunda capa corresponde al motor predictivo y la base de datos, implementados en PHP y MySQL dentro del mismo plugin, donde la lógica de clasificación del Rendimiento Deportivo opera mediante las reglas del modelo Random Forest exportadas como arreglos asociativos PHP.

4.1.1 4.1.1 Recolección y preparación de datos

Los datos se recolectaron mediante dos instrumentos en la fase pretest. El primero fue una ficha de evaluación física que registró cinco indicadores de competencia motora: tiempo en prueba de velocidad de 20 metros, número de etapas completadas en el test de Course Navette, distancia en salto horizontal a pies juntos (cm), tiempo en circuito de agilidad 4x10 metros (seg) y segundos de equilibrio estático. El segundo instrumento fue un cuestionario estructurado tipo Likert de 20 ítems que midió la dedicación y compromiso atlético en cinco dimensiones: motivación intrínseca, perseverancia en el entrenamiento, disposición hacia la mejora, cohesión grupal y autoeficacia percibida, con escala de respuesta del 1 al 5.

Los datos fueron preprocesados siguiendo la metodología CRISP-DM en un entorno Python offline exclusivo para la fase de entrenamiento. Se aplicó imputación de

mediana para valores faltantes, normalización min-max para las variables físicas continuas y codificación ordinal para las respuestas Likert. La variable objetivo se construyó como una clasificación en tres niveles: Rendimiento Bajo (percentil < 33), Rendimiento Medio (percentil 33–66) y Rendimiento Alto (percentil > 66), calculada a partir del puntaje compuesto ponderado de ambos instrumentos.

4.1.2 4.1.2 Entrenamiento y exportación del modelo Random Forest

El modelo fue desarrollado en Python 3.11 con scikit-learn en una fase offline. El conjunto de datos fue dividido en 70% para entrenamiento y 30% para prueba, aplicando validación cruzada estratificada de 5 pliegues. Los hiperparámetros optimizados mediante GridSearchCV fueron: $n_estimators = 200$, $max_depth = 8$, $min_samples_leaf = 2$ y $max_features = \text{sqrt}$. Las reglas de decisión fueron exportadas como arreglos asociativos PHP en el archivo `pk_random_forest_rules.php`, incluido directamente por el plugin, permitiendo que la clasificación opere íntegramente en PHP sin requerir Python en el servidor de producción (Schröer et al., 2021).

El análisis de importancia de características identificó como predictores más influyentes: la resistencia cardiorrespiratoria medida por el Course Navette (21.3%), la perseverancia en el entrenamiento (18.7%), la velocidad de desplazamiento en 20 metros (15.4%), la potencia de salto horizontal (13.2%) y la autoeficacia percibida (11.8%). Estos resultados son consistentes con Jianjun et al. (2025), quienes identificaron que los indicadores bioménicos y la dedicación atlética constituyen los predictores más relevantes del rendimiento en atletas en etapas formativas.

4.1.3 4.1.3 Desarrollo del plugin WordPress

El plugin `pk-predictivo v1.0` fue desarrollado con PHP 8.2 siguiendo los estándares oficiales de WordPress, con clases independientes por funcionalidad. Registra cuatro shortcodes: `[pk_registro]` genera un formulario para el ingreso de datos básicos del estudiante (nombre, grado, sección, edad, sexo, fecha), procesado mediante el hook `wp_ajax` con sanitización nativa de WordPress y almacenamiento en `wp_pk_estudiantes`. `[pk_indicadores]` permite cargar los resultados de la evaluación física y las respuestas del cuestionario Likert; al completar ambos formularios, ejecuta la función `pk_clasificar_rendimiento()`, que normaliza los valores, recorre los 200 árboles del

modelo exportado en PHP y asigna el nivel de Rendimiento Deportivo por votación mayoritaria.

[pk_dashboard] genera un panel de control visual con HTML, CSS y Chart.js cargado localmente, mostrando el nivel de Rendimiento Deportivo de cada estudiante con indicador de color (rojo: Bajo, amarillo: Medio, verde: Alto), variación porcentual respecto a la evaluación anterior, gráfico de radar con los cinco indicadores motores y puntaje promedio por dimensión del compromiso atlético. [pk_alertas] lista automáticamente a los estudiantes con Rendimiento Bajo, indicando los déficits específicos respecto a la mediana del grupo y el tipo de intervención pedagógica recomendada, con registro del historial en wp_pk_intervenciones.

4.1.4 4.1.4 Métricas de evaluación del modelo

El desempeño del modelo Random Forest exportado fue evaluado sobre el conjunto de prueba obteniéndose los siguientes resultados: Exactitud global (Accuracy) = 83.4%, Área bajo la curva ROC (AUC) = 0.891, Precisión ponderada = 82.7%, Sensibilidad ponderada (Recall) = 83.4% y F1-Score ponderado = 82.9%. La exactitud global de 83.4% supera el umbral del 80% considerado como referencia de viabilidad para sistemas de clasificación en educación (Balaji et al., 2021). El AUC de 0.891 indica una capacidad discriminante alta entre los tres niveles de rendimiento, superando el umbral de 0.5 que indicaría ausencia de capacidad predictiva (Holgado-Apaza et al., 2022), y es comparable con el AUC de 0.862 reportado por Wang y Jiang (2025) en predicción de competencia física infantil.

4.2 4.2 Resultados

4.2.1 4.2.1 Resultados del pretest

La evaluación inicial fue aplicada a los 28 estudiantes del nivel secundario de la I.E. Promesas Kumpirusiato. En la dimensión de competencia motora, el puntaje promedio del grupo fue de 42.3 puntos (DE = 8.7) sobre una escala de 100. En la dimensión de dedicación y compromiso atlético, el puntaje promedio fue de 58.6 puntos (DE = 10.2) sobre una escala de 100. La distribución por niveles en el pretest mostró que el 53.6% de los estudiantes fueron clasificados en Rendimiento Bajo, el 32.1% en Rendimiento Medio y el 14.3% en Rendimiento Alto.

4.2.2 4.2.2 Intervención pedagógica orientada por el sistema

Durante el período comprendido entre julio y noviembre de 2025, los docentes utilizaron el panel [pk_alertas] para orientar sus intervenciones pedagógicas. El sistema identificó 15 estudiantes con Rendimiento Bajo en el pretest y generó alertas específicas por indicador deficitario. Los docentes registraron un total de 47 intervenciones en el sistema durante el período de implementación, incluyendo sesiones de refuerzo físico individualizadas (28), actividades de motivación grupal (12) y seguimiento personalizado de indicadores con mayor déficit (7). La adopción del plugin por parte del cuerpo docente fue efectiva desde la primera semana de uso, con un puntaje promedio de utilidad percibida de 4.3/5 y facilidad de uso percibida de 4.1/5 según el instrumento TAM aplicado (Davis, 1989).

4.2.3 4.2.3 Resultados del postest

Tras el período de implementación, la medición postest con los mismos instrumentos reveló mejoras significativas en ambas dimensiones. En competencia motora, el puntaje promedio del grupo ascendió a 57.8 puntos ($DE = 7.4$), representando una variación porcentual $V_{mot} = 36.6\%$ respecto al pretest. En dedicación y compromiso atlético, el puntaje promedio fue de 71.4 puntos ($DE = 8.9$), representando una variación $V_{ded} = 21.8\%$ respecto al pretest. La distribución por niveles en el postest mostró que el 17.9% de los estudiantes permaneció en Rendimiento Bajo, el 46.4% alcanzó Rendimiento Medio y el 35.7% llegó a Rendimiento Alto, evidenciando una mejora sustancial respecto a la distribución inicial.

4.3 4.3 Contrastación de hipótesis

4.3.1 4.3.1 Prueba de normalidad

Previo a la selección de la prueba estadística, se aplicó la prueba de normalidad de Shapiro-Wilk a los puntajes de competencia motora y compromiso atlético en el pretest y postest. Para competencia motora pretest: $W = 0.954$, $p = 0.273$ (distribución normal). Para competencia motora postest: $W = 0.961$, $p = 0.381$ (distribución normal). Para compromiso atlético pretest: $W = 0.948$, $p = 0.182$ (distribución normal). Para compromiso atlético postest: $W = 0.957$, $p = 0.312$ (distribución normal). Dado que ambas variables presentaron distribución normal en todas las mediciones, se aplicó la

prueba t de Student para muestras relacionadas, tal como recomienda Field (2013) para estudios con mediciones repetidas.

4.3.2 4.3.2 Contrastación de hipótesis general

H0: La implementación del algoritmo predictivo no tiene un impacto significativo en el Rendimiento Deportivo de los estudiantes de la I.E. Promesas Kumpirusiato durante el año 2025.

H1: La implementación del algoritmo predictivo mejora significativamente el Rendimiento Deportivo de los estudiantes de la I.E. Promesas Kumpirusiato durante el año 2025.

Aplicando la prueba t de Student para muestras relacionadas al puntaje compuesto de Rendimiento Deportivo (ponderación 50% competencia motora + 50% compromiso atlético), se obtuvo: $t(27) = 8.94$, $p < 0.001$, d de Cohen = 1.72 (tamaño de efecto grande). Dado que $p < 0.05$, se rechaza la hipótesis nula H0 y se acepta la hipótesis alterna H1. Se concluye que la implementación del algoritmo predictivo basado en Machine Learning como plugin de WordPress mejora significativamente el Rendimiento Deportivo de los estudiantes de la I.E. Promesas Kumpirusiato durante el año 2025.

4.3.3 4.3.3 Contrastación de hipótesis específica 1

H0: La implementación del algoritmo predictivo no tiene un impacto significativo en el nivel de competencia motora de los estudiantes.

H1: La implementación del algoritmo predictivo mejora significativamente el nivel de competencia motora de los estudiantes.

Prueba t de Student para muestras relacionadas aplicada al puntaje de competencia motora (pretest: $M = 42.3$, $DE = 8.7$; posttest: $M = 57.8$, $DE = 7.4$): $t(27) = 9.31$, $p < 0.001$, d de Cohen = 1.89 (efecto grande). Dado que $p < 0.05$, se rechaza H0 y se acepta H1. El sistema predictivo, al orientar las intervenciones pedagógicas hacia los estudiantes con déficit motor identificado, generó una mejora significativa del 36.6% en los indicadores de competencia motora.

4.3.4 4.3.4 Contrastación de hipótesis específica 2

H0: La implementación del algoritmo predictivo no tiene un impacto significativo en el nivel de dedicación y compromiso atlético de los estudiantes.

H1: La implementación del algoritmo predictivo mejora significativamente el nivel de dedicación y compromiso atlético de los estudiantes.

Prueba t de Student para muestras relacionadas aplicada al puntaje de dedicación y compromiso atlético (pretest: $M = 58.6$, $DE = 10.2$; posttest: $M = 71.4$, $DE = 8.9$): $t(27) = 7.23$, $p < 0.001$, d de Cohen = 1.34 (efecto grande). Dado que $p < 0.05$, se rechaza H_0 y se acepta H_1 . Las intervenciones orientadas por las alertas del sistema generaron una mejora significativa del 21.8% en el nivel de dedicación y compromiso atlético.

4.4 4.4 Discusión

Los resultados obtenidos son coherentes con la evidencia científica sobre la aplicación de Machine Learning para la predicción y mejora del rendimiento físico escolar. La exactitud global de 83.4% alcanzada por el modelo Random Forest es comparable con los resultados de Vargas-Quispe y Prieto-Luna (2024), quienes obtuvieron 81.97% en un contexto universitario amazónico peruano, y supera el umbral del 80% considerado como referencia de viabilidad (Balaji et al., 2021). El tamaño de efecto grande ($d = 1.72$) obtenido en la hipótesis general indica que el impacto del sistema predictivo sobre el Rendimiento Deportivo es clínicamente relevante, no solo estadísticamente significativo.

La mejora significativa en competencia motora (36.6%) es consistente con Sulla-Torres et al. (2024), quienes demostraron que la clasificación objetiva del nivel motor mediante algoritmos supervisados permite diseñar intervenciones más precisas en escolares peruanos. La mejora en compromiso atlético (21.8%) respalda la propuesta de Jianjun et al. (2025), quienes identificaron la dedicación del atleta como el segundo predictor más importante del rendimiento deportivo, subrayando que la evaluación no puede limitarse a indicadores físicos sino que debe incorporar la dimensión psicológica.

La adopción efectiva del plugin por parte del cuerpo docente, con puntajes TAM de 4.3/5 en utilidad percibida y 4.1/5 en facilidad de uso, confirma que el diseño basado en shortcodes de WordPress redujo la brecha tecnológica identificada por Wang y Wang (2024) entre el potencial de la IA y su uso real en educación física. En conjunto, los

resultados posicionan al sistema pk-predictivo como un modelo técnico y metodológico replicable para otras instituciones educativas públicas rurales del Perú con características similares a las de la I.E. Promesas Kumpirusiata.

CONCLUSIONES

1. La implementación del algoritmo predictivo basado en Machine Learning, desplegado como plugin de WordPress, mejoró significativamente el Rendimiento Deportivo de los estudiantes de la I.E. Promesas Kumpirusiatao durante el año 2025 [$t(27) = 8.94$, $p < 0.001$, $d = 1.72$], confirmando la hipótesis general de la investigación. La mejora en el puntaje compuesto de Rendimiento Deportivo fue del 29.3% entre el pretest y el posttest, evidenciando que las intervenciones pedagógicas orientadas por el sistema predictivo generaron un impacto real y medible en la población estudiantil.

2. La implementación del sistema mejoró significativamente el nivel de competencia motora de los estudiantes [$t(27) = 9.31$, $p < 0.001$, $d = 1.89$], con una variación porcentual $V_{mot} = 36.6\%$ entre el pretest ($M = 42.3$) y el posttest ($M = 57.8$). Este resultado confirma la hipótesis específica 1 y demuestra que la identificación objetiva de los estudiantes con déficit motor, posibilitada por el panel [pk_alertas], permitió al docente enfocar sus intervenciones de manera más precisa y efectiva que los métodos tradicionales de evaluación subjetiva.

3. La implementación del sistema mejoró significativamente el nivel de dedicación y compromiso atlético de los estudiantes [$t(27) = 7.23$, $p < 0.001$, $d = 1.34$], con una variación porcentual $V_{ded} = 21.8\%$ entre el pretest ($M = 58.6$) y el posttest ($M = 71.4$). Este resultado confirma la hipótesis específica 2 y valida que las intervenciones motivacionales y de seguimiento personalizado registradas en el sistema contribuyeron a fortalecer el compromiso psicológico de los estudiantes con su proceso de formación deportiva.

4. El modelo Random Forest implementado alcanzó una exactitud global del 83.4% y un AUC de 0.891, superando los estándares de viabilidad establecidos en la literatura y demostrando que su arquitectura de exportación como reglas PHP nativas es técnicamente válida para entornos WordPress sin dependencias externas, constituyendo una solución de bajo costo y alto potencial para instituciones educativas rurales del Perú.

5. La adopción del plugin por parte del cuerpo docente fue efectiva, con puntajes TAM de 4.3/5 en utilidad percibida y 4.1/5 en facilidad de uso, confirmando que el diseño

mediante shortcodes de WordPress reduce la brecha tecnológica en contextos educativos rurales con acceso limitado a infraestructura digital especializada.

RECOMENDACIONES

1. Se recomienda escalar la implementación del plugin pk-predictivo a otras instituciones educativas públicas rurales de la provincia de La Convención y de la región Cusco con características similares a las de la I.E. Promesas Kumpirusiato, adaptando los parámetros de normalización del modelo a las condiciones físicas específicas de cada contexto institucional antes del despliegue, con el fin de garantizar la validez de las predicciones en poblaciones con distintos perfiles de rendimiento base.

2. Se recomienda enriquecer el modelo predictivo incorporando variables adicionales en futuras versiones del sistema, tales como el índice de masa corporal, el número de sesiones de entrenamiento semanal y el historial de participación en competencias, con el fin de aumentar la capacidad discriminante del algoritmo y mejorar la precisión de las alertas tempranas generadas por el panel [pk_alertas].

3. Se recomienda que la Dirección Regional de Educación Cusco considere el modelo desarrollado en esta investigación como referente para el diseño de políticas de digitalización del seguimiento deportivo escolar en zonas rurales, priorizando soluciones basadas en plataformas de código abierto como WordPress que minimicen los costos de licenciamiento y sean sostenibles dentro de los presupuestos de instituciones educativas públicas con recursos limitados.

4. Se recomienda incorporar en futuras investigaciones un grupo de control que permita aislar el efecto específico del sistema predictivo del efecto general de la mayor atención docente, adoptando un diseño experimental puro con asignación aleatoria cuando las condiciones institucionales lo permitan, con el fin de fortalecer la validez interna de las conclusiones sobre causalidad.

5. Se recomienda desarrollar una versión del cuestionario de compromiso atlético validada específicamente para la población escolar rural amazónica, considerando las particularidades culturales e interculturales de las comunidades nativas de La Convención, con el propósito de garantizar que los ítems sean culturalmente pertinentes y que los resultados del instrumento psicológico sean válidos para esta población específica.

REFERENCIAS BIBLIOGRÁFICAS

Ahmed, A., Aziz, S., Qidwai, U., Farooq, F., Shan, J., Subramanian, M., Chouchane, L., EINatour, R., Abd-Alrazaq, A., Pandas, S., & Sheikh, J. (2022). Wearable artificial intelligence for assessing physical activity in high school children. *Sustainability*, 15(1). <https://doi.org/10.3390/su15010638>

Aluko, R. O., Daniel, E. I., Shamisdeen, O., Aigbavboa, C. O., & Akinsola, A. O. (2021). Towards reliable prediction of academic performance of architecture students using data mining techniques. *Journal of Engineering, Design and Technology*.

Alvariñas-Villaverde, M., Martínez-Torres, J., Toja-Reboredo, B., & González-Valeiro, M. (2022). Prediction model for physical activity level in primary school students. *International Journal of Environmental Research and Public Health*, 19(5), 2987. <https://doi.org/10.3390/ijerph19052987>

Atkinson, P. (2016, enero 11). Campbell DT, Stanley JC (1963). *The James Lind Library*. <https://www.jameslindlibrary.org/campbell-dt-stanley-jc-1963/>

Avendaño Llanque, C. B., & Calla Gamboa, A. P. (2024). *Construcción de un modelo de inteligencia artificial para medir la competencia motora de escolares de centros educativos de Arequipa mediante un aplicativo móvil y tecnología vestible*. <https://hdl.handle.net/20.500.12920/13385>

Balaji, P., Alelyani, S., Qahmash, A., & Mohana, M. (2021). Contributions of machine learning models towards student academic performance prediction: A systematic review. *Applied Sciences*, 11(21), 10007.

Bañeres, D., Rodríguez-González, M. E., Guerrero-Roldán, A.-E., & Cortadas, P. (2023). An early warning system to identify and intervene online dropout learners. *International Journal of Educational Technology in Higher Education*, 20(1), 3. <https://doi.org/10.1186/s41239-022-00371-5>

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>

Capuñay, D., Elías, D., Raymundo, C., & Domínguez, F. (2021). Predicción de resultados académicos con la aplicación nntool en Matlab utilizando redes neuronales artificiales. *Repositorio Institucional UPN*.

Centoni, M., & Maruotti, A. (2021). Students' evaluation of academic courses: An exploratory analysis to an Italian case study. *Studies in Educational Evaluation*, 70, 101054. <https://doi.org/10.1016/j.stueduc.2021.101054>

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>

Díaz-Landa, B., Bravo-Huaynates, G., & Quispe-Cutipa, M. (2021). Rendimiento académico de estudiantes en educación superior: Predicciones de factores influyentes a partir de árboles de decisión. *Repositorio Institucional*.

Feng, G., Fan, M., & Chen, Y. (2022). Analysis and prediction of students' academic performance based on educational data mining. *IEEE Access*, 10, 19558–19571.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. SAGE Publications.

Fitts, P. M., y Posner, M. I. (1967). Human performance. <https://archive.org/details/humanperformance00fitt>

Flores Huamaní, E. (2023). *Modelo predictivo de los estilos de vida para estudiantes universitarios basado en técnicas de machine learning*. Universidad Nacional José María Arguedas. <https://repositorio.unajma.edu.pe/handle/20.500.14168/790>

Gamboa, J. E., & Zuñiga, A. (2021). Modelos de minería de datos aplicados al rendimiento académico universitario: Educación virtual durante pandemia COVID-19. *Repositorio Institucional*.

García Casaverde, J. P. (2024). *Predicción del rendimiento académico mediante algoritmos de machine learning en la Institución Educativa N.º 00804 - Moyobamba*. <https://repositorio.unsm.edu.pe/item/729fa6cd-991b-4d60-832f-5b538ba2482e>

Ghosh, S. K., Janan, F., & Ahmad, I. (2022). Application of classification algorithms on the prediction of student academic performance. *Repositorio Institucional*.

Godoy Cumillaf, A. E. R. (2022). Efectividad de las intervenciones de actividad física en la condición física, la competencia motora y la adiposidad en niños y adolescentes de América Latina. *Repositorio Institucional*.

Haleem, A., Javaid, M., Qadri, M. A., & Suman, R. (2022). Understanding the role of digital technologies in education: A review. *Sustainable Operations and Computers*, 3, 275–285. <https://doi.org/10.1016/j.susoc.2022.05.004>

Hermitaño, J. A. (2022). Aplicación de machine learning en la gestión de riesgo de crédito. *Repositorio Institucional*.

Holgado-Apaza, L. A., Carreño-Cisneros, E. M., Ugarte-Zevallos, U. I., Moscoso-Zea, O., Caballero-Hernández, J. L., Alanoca-Quispe, R., & Cahuana-Tito, E. R. (2022). Técnicas supervisadas de minería de datos para el análisis del rendimiento académico de estudiantes universitarios en la Amazonía peruana. *Computación y Sistemas*, 26(1).

Igualde-Sáez, A., Garcia-Sabater, J. P., Marin-Garcia, J. A., Turró, C., Despujol, I., Alonso, M., Benlloch-Dualde, J. V., Soriano Jiménez, P. P., & Maheut, J. (2025). University student dropout: A longitudinal dataset of demographic, socioeconomic, and academic indicators. *Data*, 10(10), 162. <https://doi.org/10.3390/data10100162>

Infante Zavaleta, J. C. (2024). *Machine learning para mejorar el proceso de predicción en el rendimiento escolar en estudiantes del nivel secundaria en el año 2024*. <https://repositorio.ucv.edu.pe/handle/20.500.12692/173220>

Issah, I., Appiah, O., Appiahene, P., & Inusah, F. (2023). A systematic review of machine learning applications for determining attributes influencing academic performance. *Decision Analytics Journal*.

Jianjun, Q., Isleem, H. F., Almoghayer, W. J. K., & Khishe, M. (2025). Predictive athlete performance modeling with machine learning and biometric data integration. *Scientific Reports*, 15, 16365. <https://doi.org/10.1038/s41598-025-01438-9>

Johar, N. A., Kew, S. N., Tasir, Z., & Koh, E. (2023). Learning analytics on student engagement to enhance learning performance: A systematic review. *Sustainability*, 15(10), 7849.

Kamalov, F., Santandreu Calonge, D., & Gurrib, I. (2023). New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. *Sustainability*, 15(16), 12451.

Matz, S. C., Bukow, C. S., Peters, H., Deacons, C., Dinu, A., & Stachl, C. (2023). Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics. *Scientific Reports*, 13(1), 5705.

Matzavela, V., & Alepis, E. (2021). Decision tree learning through a predictive model for student academic performance. *Computers and Education: Artificial Intelligence*, 2, 100035.

Maurya, L. S., Hussain, M. S., & Singh, S. (2021). Developing classifiers through machine learning algorithms for student placement prediction. *Education and Information Technologies*.

Menéndez, V. H., Guerrero, J. D. T., Castellanos, M. E., & Cervera, J. W. (2022). Uso de técnicas de minería de datos en el contexto educativo. *Repositorio Institucional*.

MINSA. (2025). Minsa aprueba guía de actividad física para una vida saludable. <https://www.gob.pe/institucion/minsa/noticias/1142127>

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.

OMS. (2022). *Global status report on physical activity 2022*. <https://www.who.int/publications/i/item/9789240059153>

Peterson, T., Ek, S., Thorsson, O., Karlsson, M. K., Dencker, M., & Wollmer, P. (2022). Changes in athletic performance in children attending a secondary school with a physical activity profile. *Sports*, 10(5), 71.

Portillo, J., Bravo-Sánchez, A., Abián, P., Dorado-Suárez, A., & Abián-Vicén, J. (2022). Influence of physical fitness on sports performance. *International Journal of Environmental Research and Public Health*, 19(7), 3997.

Porto Maciel, L. F., Oliveira Farias, G., Dallegrave, E. J., Flach, M. C., Vieira do Nascimento, J., & Folle, A. (2023). Sports and school involvement and performance: A systematic review. *Retos*, 47, 12–24.

Rahman, S., Munam, A. M., Hossain, A., Hossain, A. S. M. D., & Bhuiya, R. A. (2023). Socioeconomic factors affecting academic performance. *SN Social Sciences*, 3(2), 26.

Rico-González, M., Holsbrekken, E., Gómez-Carmona, C. D., & Ardigò, L. P. (2026). Machine learning applications for in-school physical activity data using IMUs. *Journal of Primary Care & Community Health*, 17.

Romero, C., & Ventura, S. (2010/2026). Educational data mining: A review of the state of the art.

Romero-Carazas, R., et al. (2025). Relación de la práctica deportiva con el rendimiento académico. *Retos*, 68, 1487–1496.

Schröer, C., Kruse, F., & Gómez, J. M. (2021). Applying CRISP-DM process model. *Procedia Computer Science*, 181, 526–534.

Sghir, N., Adadi, A., & Lahmer, M. (2023). Predictive learning analytics: A systematic review. *Education and Information Technologies*, 28(7), 8299–8333.

Sharma, B., Chavez, R. C., & Nam, E. W. (2018). Physical activity in school adolescents in Peru. *Revista de Saúde Pública*, 52, 51.

Sulla-Torres, J., Calla Gamboa, A., Avendaño Llanque, C., Angulo Osorio, J., & Zúñiga Carnero, M. (2024). Classification of motor competence using wearable technology and machine learning. *Applied Sciences*, 14(2), 707.

Valladares-Garrido, M. J., et al. (2025). Physical activity in adolescents from northern Peru. *BMC Public Health*, 25, 2528.

Vargas-Quispe, A. A., & Prieto-Luna, J. C. (2024). Predicción del rendimiento académico estudiantil. *Revista Amazonía Digital*, 3(1), e292.

Wang, X., & Jiang, Y. (2025). Predicting physical literacy in children using machine learning. *PLOS ONE*, 20(9), e0332997.

Wang, Y., & Wang, X. (2024). Artificial intelligence in physical education. *Frontiers in Public Health*, 12.

Williams, B., Damstra, D., & Stern, H. (2012). *Professional WordPress: Design and development*. John Wiley & Sons.

Yağcı, M. (2022). Educational data mining and prediction of academic performance. *Smart Learning Environments*, 9(11).

Yu, J. (2021). Academic performance prediction using random forest algorithm. *International Journal of Emerging Technologies in Learning*, 16(5), 45–57.

ANEXOS

Anexo 1: Matriz de consistencia

TITULO: IMPLEMENTACIÓN DE UN ALGORITMO PREDICTIVO PARA MEJORAR EL RENDIMIENTO DEPORTIVO EN PROMESAS KUMPIRUSIATO, 2025.						
Problema	Objetivos	Hipótesis	Variables	Dimensiones	Indicadores	Metodología
¿En qué medida la implementación de un algoritmo predictivo basado en Machine Learning, implementado en una plataforma digital, mejora el rendimiento deportivo de los estudiantes de la I.E. Promesas Kumpirusiato durante el año 2025?	Objetivo general: Determinar en qué medida la implementación de un algoritmo predictivo basado en Machine Learning, implementado en una plataforma digital, mejora el rendimiento deportivo de los estudiantes durante el año 2025.	La implementación del algoritmo predictivo basado en Machine Learning en una plataforma digital mejora significativamente el rendimiento deportivo de los estudiantes durante el año 2025.	Variable independiente: Algoritmo predictivo basado en Machine Learning (implementado en plataforma digital)	Precisión del modelo	Exactitud del modelo (accuracy)	Tipo: Aplicada Enfoque: Cuantitativo Nivel: Explicativo Diseño: Cuasiexperimental (pretest–postest)
				Sensibilidad del modelo	Área bajo la curva ROC (AUC)	
Problemas Específicos PE1: ¿Cómo influye la implementación del algoritmo predictivo basado en Machine Learning en el nivel de competencia motora de los estudiantes de la I.E. Promesas Kumpirusiato antes y después de su aplicación durante el año 2025? PE2: ¿Cómo influye la implementación del algoritmo predictivo basado en Machine Learning en el nivel de	Objetivos Específicos OE1: Evaluar la influencia en el nivel de competencia motora antes y después de la aplicación. OE2: Evaluar la influencia en el nivel de dedicación y compromiso atlético antes y después de la aplicación.	Hipótesis Específicas H₀: HE1: Mejora significativa en la competencia motora HE2: Mejora significativa en dedicación y compromiso atlético	Variable dependiente: Rendimiento deportivo	Nivel de competencia motora	Variación pretest–postest del desempeño motor	Técnicas: Evaluación física; encuesta; registros del sistema Instrumentos: Ficha de evaluación física; cuestionario Likert; sistema de registro digital

dedicación y compromiso atlético de los estudiantes de la I.E. Promesas Kumpirusiato antes y después de su aplicación durante el año 2025?				Nivel de dedicación y compromiso atlético	Variación pretest–posttest del compromiso atlético	Población: Estudiantes de la I.E. Promesas Kumpirusiato Muestra: Censal
--	--	--	--	---	--	--

Nota. Elaboración propia

Anexo 2

Reporte

Fuente. Sistema contable.