[1] **FACULTAD DE INGENIERÍA Y ARQUITECTURAeeew**

**CARRERA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

**Implementación de un asistente inteligente basado en Machine Learning para mejorar el aprendizaje del curso de Química Básica en estudiantes de secundaria de escuelas públicas peruanas**

**AUTORES:**

Quilca Urbano Yilmer (ORCID: 0009-0001-8852-6554)

Parra Godoy Jeancarlos Miguel (ORCID:….

**Profesor:**

Dr. Hugo Froilán Vega Huerta

**Lima - Perú**

**2025**

# 

# DEDICATORIA

A mis padres y hermanos,  
por ser el pilar fundamental en mi vida.  
Su apoyo constante, sus valores y enseñanzas  
han sido esenciales para alcanzar esta meta.  
Agradezco profundamente su paciencia,  
comprensión y amor incondicional.

RESUMEN

# AGRADECIMIENTO

Agradezco primeramente a Dios por brindarme la fortaleza y sabiduría necesarias para culminar esta etapa.  
A mi familia, por su amor, paciencia y apoyo incondicional a lo largo de este camino.

Título : Implementación de un asistente inteligente basado en Machine Learning para mejorar el aprendizaje del curso de Química Básica en estudiantes de secundaria de escuelas públicas

Autores : Quilca Urbano Yilmer

Parra Godoy Jeancarlos Miguel

Asesor de tesis : Dr. Hugo Froilán Vega Huerta

Fecha : Abril 2025

El presente proyecto analiza la implementación de un Asistente Inteligente basado en Machine Learning para mejorar el aprendizaje del curso de Química Básica en estudiantes de secundaria de escuelas públicas. Este desarrollo busca optimizar los procesos de enseñanza-aprendizaje mediante la personalización de contenidos y la retroalimentación automatizada, facilitando la comprensión de conceptos clave de manera interactiva y accesible.

Se investigaron plataformas educativas inteligentes enfocadas en el apoyo al aprendizaje en ciencias, evaluando su capacidad de adaptación al ritmo y estilo de cada estudiante. Asimismo, se analizaron técnicas de aprendizaje automático aplicadas al monitoreo del desempeño estudiantil y la generación de recomendaciones personalizadas.

**PALABRAS CLAVE:**Asistente Inteligente, Química Básica, Machine Learning, Aprendizaje Personalizado  
  
  
  
  
  
ABSTRACT

Title : Implementation of an Intelligent Assistant Based on Machine Learning to Improve the Learning of Basic Chemistry in Public High School Students

Authors : Quilca Urbano Yilmer

Parra Godoy Jeancarlos Miguel

Thesis advisor : Dr. Hugo Froilán Vega Huerta

Date : April 2025

This project analyzes the implementation of an intelligent assistant based on Machine Learning to improve the learning of Basic Chemistry in public high school students. This development aims to optimize teaching and learning processes through personalized content and automated feedback, facilitating the understanding of key concepts in an interactive and accessible manner.  
Intelligent educational platforms focused on science learning support were studied, evaluating their ability to adapt to each student's pace and learning style. Additionally, machine learning techniques applied to monitoring student performance and generating personalized recommendations were analyzed.

**KEY WORDS:**

Intelligent Assistant, Basic Chemistry, Machine Learning, Personalized Learning

Contenido

[2](#_Toc201927289)

[DEDICATORIA 2](#_Toc201927290)

[AGRADECIMIENTO 3](#_Toc201927291)

[INTRODUCCION 7](#_Toc201927292)

[CAPÍTULO I: VISIÓN DEL PROYECTO 8](#_Toc201927293)

[1.1 Antecedentes del Problema 1.1.1 El Negocio 8](#_Toc201927294)

[Figura 1.1. Porcentaje de estudiantes con logro en Ciencia y Tecnología Fuente (MINEDU, 2024) 9](#_Toc201927295)

[1.1.2 Proceso del Negocio 10](#_Toc201927296)

[Fuente: (Elaboración propia 2025 ) 10](#_Toc201927297)

[**1.2.2 Descripción del problema** 1.2.2.1 Problema principal 13](#_Toc201927298)

[1.2.2.2 Problema secundarios 13](#_Toc201927299)

[1.3. Objetivos del Proyecto 14](#_Toc201927300)

[1.3.1. Marco Lógico 14](#_Toc201927301)

[Figura 1 15](#_Toc201927302)

[Figura 2 15](#_Toc201927303)

[1.3.2. Objetivo general 17](#_Toc201927304)

[**1.3.3. Objetivos específicos** 18](#_Toc201927305)

[1.4 Justificación del Proyecto 19](#_Toc201927306)

[1.4.1 Justificación Académica 19](#_Toc201927307)

[1.4.2 Beneficios Tangibles 19](#_Toc201927308)

[1.4.3 Beneficios Intangibles 20](#_Toc201927309)

[**1.5 Alcance del Proyecto** 20](#_Toc201927310)

[CAPÍTULO III: ESTADO DEL ARTE 3.1 Artículos 21](#_Toc201927311)

[**3.1.12.** **Artificial Intelligence Applications in K-12 Education: A Review of Recent Trends** Biagini (2025) **3.1.12.** **Aplicaciones de la inteligencia artificial en la educación K-12: una revisión de tendencias recientes** Biagini (2025) Citado 0, fuente: Scopus 56](#_Toc201927312)

[**DOI:** 10.1109/ACCESS.2022.3179356 56](#_Toc201927313)

[CAPÍTULO IV: MODELADO DEL NEGOCIO 65](#_Toc201927314)

[**4.1 Reglas del Negocio** 65](#_Toc201927315)

[**4.2 Casos de uso** 65](#_Toc201927316)

[CAPÍTULO V: REQUERIMIENTOS DEL PROYECTO 69](#_Toc201927317)

[**5.1.1 Requerimientos Funcionales** 69](#_Toc201927318)

[**5.1.2 Requerimientos No Funcionales** 70](#_Toc201927319)

INTRODUCCIONLa enseñanza de las ciencias, y en especial de la Química Básica, representa un reto considerable en el entorno de las escuelas públicas de nivel secundario. Factores como la escasez de recursos, el elevado número de alumnos por aula y la variabilidad en los ritmos de aprendizaje dificultan una instrucción personalizada y eficiente. Esta problemática puede generar falta de interés y escasa comprensión de los contenidos, afectando negativamente el desempeño académico del estudiante.

Frente a este panorama, la integración de tecnologías emergentes como la inteligencia artificial, y en particular el aprendizaje automático (Machine Learning), abre nuevas posibilidades para mejorar los métodos educativos. Una de las soluciones más innovadoras es la creación de asistentes inteligentes que puedan ajustarse al nivel de conocimiento del estudiante, detectar sus dificultades específicas y proporcionarle contenidos personalizados que fortalezcan su proceso de aprendizaje.

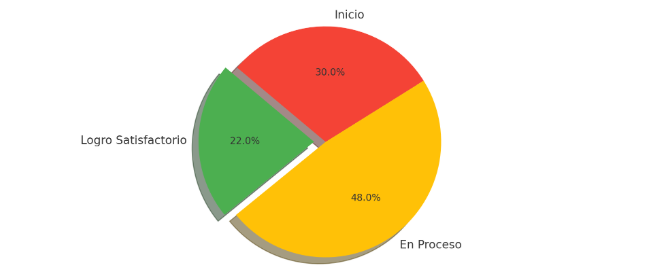
En esta línea, el proyecto plantea la implementación de un asistente inteligente fundamentado en técnicas de Machine Learning, orientado a apoyar la enseñanza de la Química Básica. El sistema propuesto tiene como finalidad analizar el rendimiento de los estudiantes, sugerir materiales acordes a su nivel de comprensión y ofrecer retroalimentación inmediata. Con ello, se espera mejorar la comprensión de los conceptos esenciales del curso, incrementar la motivación y lograr un impacto positivo en los resultados académicos.

Esta investigación se enmarca en la búsqueda de innovación educativa mediante herramientas tecnológicas accesibles, con el fin de reducir las brechas en la calidad de la educación, especialmente en contextos de escuelas públicas.

# CAPÍTULO I: VISIÓN DEL PROYECTO

# 1.1 Antecedentes del Problema 1.1.1 El Negocio

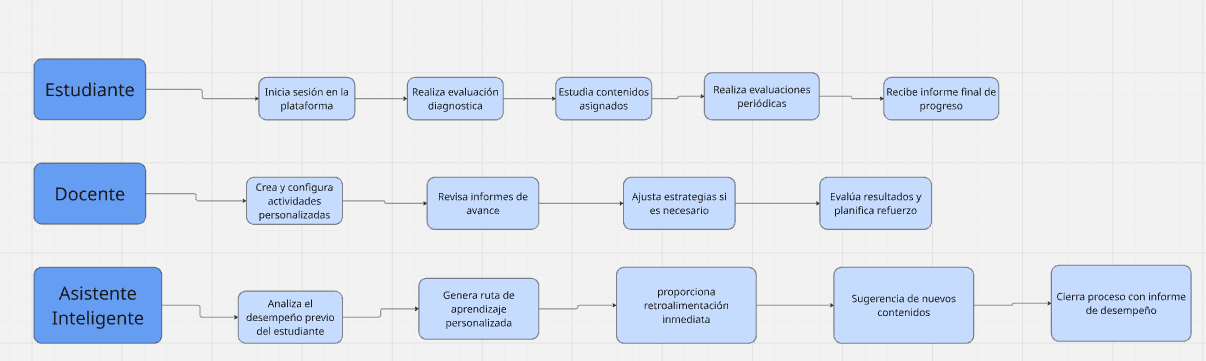
La educación pública en el Perú, especialmente en el nivel secundario, enfrenta múltiples desafíos en la enseñanza de ciencias, y particularmente en el curso de Química Básica. Según datos del Ministerio de Educación (MINEDU), los resultados de las evaluaciones censales muestran bajos niveles de logro en el área de Ciencia y Tecnología, revelando que un gran porcentaje de estudiantes no logra comprender conceptos fundamentales de la materia.

Estudios recientes indican que las escuelas públicas carecen de recursos tecnológicos adecuados, materiales didácticos suficientes y estrategias pedagógicas adaptativas. A ello se suma la alta cantidad de estudiantes por aula, que limita la atención personalizada y dificulta la identificación oportuna de las dificultades de aprendizaje en cada estudiante.

*Figura 1.1. Porcentaje de estudiantes con logro en Ciencia y Tecnología  
Fuente (MINEDU, 2024)*

1.1.2 Proceso del Negocio

Este diagrama ilustra una plataforma de aprendizaje donde el **Estudiante** estudia y es evaluado, el **Docente** personaliza las actividades y supervisa el progreso, y un **Asistente Inteligente** individualiza la ruta de aprendizaje y provee retroalimentación constante.

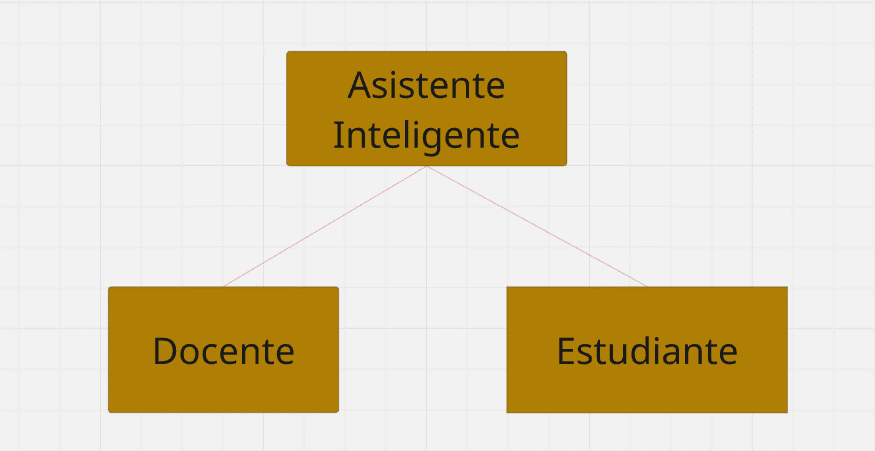
  
  
  
 Figura 1.2 Proceso de interacción

Fuente: (Elaboración propia 2025 )  
  
  
  
1.1.3 Modelo de proceso (AS IS)

El diagrama ilustra cómo un asistente inteligente basado en Machine Learning mejora el aprendizaje de Química Básica en estudiantes de secundaria peruanos. El proceso inicia con el docente diagnosticando el contexto. Luego, el sistema identifica brechas y evalúa el nivel del estudiante para entregar contenido personalizado y retroalimentación inmediata, fomentando la interacción que lleva a una mejora en la comprensión y, finalmente, al incremento del rendimiento académico.

Diagrama

El contenido generado por IA puede ser incorrecto.

  
  
  
1.1.4 Organigrama  
  
  
  
 Figura 1.3 Organigrama de escuela pública   
 Fuente ( Elaboración propia, 2025)  
  
  
1.2 Formulación del problema  
 1.2.1 Realidad problemática

En el ámbito de la educación pública peruana, persiste una preocupación significativa respecto al bajo nivel de logro académico en las áreas de ciencia, siendo la asignatura de Química Básica una de las más afectadas en el nivel secundario. Los informes de rendimiento educativo reflejan que un gran número de estudiantes no logra desarrollar adecuadamente las competencias científicas esperadas, lo cual repercute negativamente en su formación integral y limita su acceso a oportunidades en campos científicos y tecnológicos.

Esta problemática se intensifica debido a condiciones estructurales como el exceso de estudiantes por aula, la carencia de recursos pedagógicos actualizados, las limitaciones tecnológicas presentes en las instituciones educativas y las dificultades del docente para aplicar metodologías innovadoras que se ajusten a las necesidades individuales de los alumnos. El enfoque tradicional de enseñanza, en muchos casos uniforme y poco interactivo, no responde eficazmente a la diversidad de estilos y ritmos de aprendizaje existentes, generando desmotivación y bajo compromiso con la materia.

Ante esta situación, las tecnologías emergentes, como la inteligencia artificial y el aprendizaje automático (Machine Learning), surgen como herramientas con gran potencial para mejorar los procesos educativos. No obstante, su incorporación en las escuelas públicas aún es limitada, y existen pocas iniciativas orientadas a su aplicación práctica y contextualizada en el aula.

La ausencia de sistemas inteligentes capaces de personalizar la enseñanza y brindar apoyo efectivo tanto a estudiantes como a docentes constituye una barrera en la búsqueda de una educación más equitativa y de calidad. Por ello, se vuelve imprescindible diseñar e implementar soluciones tecnológicas adaptativas que permitan mejorar la comprensión de los contenidos de Química, incrementar el interés por la ciencia y elevar el rendimiento académico mediante el uso de asistentes inteligentes accesibles para las instituciones públicas.

## 1.2.2 Descripción del problema 1.2.2.1 Problema principal

En el sistema educativo peruano, existen diversos factores estructurales y pedagógicos que han generado impactos significativos en la enseñanza de las ciencias. Entre ellos, se destacan la escasez de laboratorios y materiales didácticos (65%), la falta de capacitación actualizada en química por parte de los docentes (52%) y el predominio de estrategias de enseñanza poco dinámicas y tradicionales (60%). Estas condiciones afectan directamente la calidad del proceso formativo, impidiendo que los estudiantes desarrollen una comprensión sólida y aplicada de los contenidos científicos. Todo ello ha contribuido a generar un problema central en el país, evidenciando las Dificultades en el aprendizaje del curso de Química Básica en estudiantes de secundaria de escuelas públicas peruanas, lo cual representa un obstáculo crítico para el progreso académico y profesional de los adolescentes. Esta situación no solo reduce el nivel de comprensión científica, sino que además se manifiesta en un bajo rendimiento académico en ciencias naturales (77%), con consecuencias a largo plazo como el menor acceso a carreras científicas y tecnológicas (70%) y dificultades para resolver problemas científicos en contextos reales (68%). Así, el problema se perpetúa en un ciclo que limita las oportunidades de formación científica, innovación y desarrollo social en el Perú.

Variable 1: Estudiantes que logran nivel satisfactorio en ciencia 17% según (Evaluación Censal de Estudiantes - MINEDU)

Variable 2: Estudiantes en Perú que superan el nivel básico en ciencias 30% según (PISA - OCDE)

1.2.2.2 Problema secundarios

* Insuficiente acceso a laboratorios y materiales didácticos de ciencias, lo cual limita la aplicación práctica de los conceptos.
* Docentes sin formación continua en estrategias innovadoras de enseñanza de química, lo que reduce la calidad pedagógica del curso.
* Bajo nivel de motivación e interés de los estudiantes hacia las ciencias, relacionado con metodologías poco participativas.

## 1.3. Objetivos del Proyecto

## 1.3.1. Marco Lógico

1.3.1.1. Árbol de problemas:

## Figura 1

*Árbol de problemas determinados*

Dificultades para resolver problemas científicos en contextos reales (68%)

Dificultades en el aprendizaje del curso de Química Básica en estudiantes de secundaria de escuelas públicas peruanas

Variable 1: Estudiantes que logran nivel satisfactorio en ciencia 17% según (Evaluación Censal de Estudiantes - MINEDU)

Variable 2: Estudiantes en Perú que superan el nivel básico en ciencias 30% según (PISA - OCDE)

Bajo rendimiento académico en ciencias naturales (solo 77%)

Menor acceso a carreras científicas y tecnológicas (70%)



Escasez de laboratorios y materiales didácticos (65%)

falta de capacitación actualizada en química por parte de los docentes (52%)

Estrategias de enseñanza poco dinámicas y tradicionales (60%)

## Figura 2

*Árbol de objetivos determinados*

Fortalecer la capacidad para aplicar conocimientos científicos en la vida real (incremento 35%)

mejorar el aprendizaje del curso de Química Básica en estudiantes de secundaria de escuelas públicas peruanas

Variable 1: Elevar el porcentaje de estudiantes con desempeño satisfactorio en ciencias más del 40% – MINEDU)

variable 2: Alcanzar un nivel igual o superior al 50% de estudiantes que superan el nivel básico en ciencias (PISA – OCDE)

Mejorar el rendimiento académico en ciencias naturales (objetivo: 40%)

Aumentar la elección de carreras científicas y tecnológicas (incremento 30%)



implementación de laboratorios escolares y la entrega de materiales didácticos modernos (65%),

Capacitación continua y actualizada de docentes en química (52%)

Aplicación de estrategias dinámicas e innovadoras en la enseñanza (60%)

### 1.3.2. Objetivo general

Para fortalecer la formación científica en el nivel secundario, se propone una serie de acciones enfocadas en optimizar los recursos y métodos de enseñanza en las escuelas públicas del país. Una de las principales medidas es la implementación de laboratorios escolares y la entrega de materiales didácticos modernos (65%), lo cual permitirá que los estudiantes desarrollen experiencias prácticas y significativas. También se contempla la capacitación continua y actualizada de docentes en química (52%), con el propósito de reforzar sus habilidades profesionales y mejorar su desempeño pedagógico. En este marco de intervención educativa, el objetivo principal es mejorar el aprendizaje del curso de Química Básica en estudiantes de secundaria de escuelas públicas peruanas, como eje transformador del proceso de enseñanza. A fin de lograrlo, se fomentará además la aplicación de estrategias dinámicas e innovadoras en la enseñanza (60%), orientadas a despertar el interés de los estudiantes por la ciencia. Con estas acciones articuladas, se espera como resultado mejorar el rendimiento académico en ciencias naturales (40%), aumentar la elección de carreras científicas y tecnológicas 30%), y fortalecer la capacidad para aplicar conocimientos científicos en la vida real (35%), contribuyendo así al desarrollo de competencias clave para el siglo XXI.

Variable 1: Elevar el porcentaje de estudiantes con desempeño satisfactorio en ciencias más del 40% – MINEDU)

variable 2: Alcanzar un nivel igual o superior al 50% de estudiantes que superan el nivel básico en ciencias (PISA – OCDE)

1.3.2.1 Modelo de proceso (TO-BE)

El diagrama muestra un asistente inteligente educativo que, tras la configuración del docente, analiza datos del estudiante para identificar brechas de aprendizaje. Luego, genera rutas personalizadas y retroalimentación, permitiendo al estudiante interactuar con contenido adaptado y aplicar conocimientos en simulaciones, lo que finalmente busca mejorar su rendimiento académico.

Diagrama

El contenido generado por IA puede ser incorrecto.

## **1.3.3. Objetivos específicos**

* Implementar laboratorios escolares y dotar de materiales didácticos modernos en al menos el 65% de las instituciones públicas seleccionadas, para fortalecer el aprendizaje práctico y experimental de la Química Básica.
* Desarrollar programas de capacitación continua y actualizada en química para el 52% del personal docente, con el fin de mejorar su preparación y promover el uso de enfoques pedagógicos más efectivos.
* Fomentar el uso de estrategias de enseñanza dinámicas, activas e innovadoras en un 60% de las sesiones del curso de química, a través de talleres metodológicos y acompañamiento docente.

# 1.4 Justificación del Proyecto

## 1.4.1 Justificación Académica

El presente proyecto busca atender una problemática educativa urgente: el bajo nivel de aprendizaje en el curso de Química Básica entre estudiantes de secundaria de instituciones públicas peruanas. Esta situación es consecuencia de factores estructurales y pedagógicos como la falta de laboratorios, escasos materiales didácticos, docentes con capacitación desactualizada y metodologías de enseñanza poco dinámicas. Desde una perspectiva académica, abordar esta situación es fundamental para garantizar una formación científica sólida, equitativa y de calidad. El proyecto no solo contribuirá al mejoramiento del rendimiento académico en ciencias naturales, sino que también estimulará la curiosidad científica y el pensamiento crítico, habilidades esenciales en la formación escolar moderna.

## 1.4.2 Beneficios Tangibles

Entre los beneficios concretos del proyecto se encuentran la implementación de laboratorios funcionales en las escuelas públicas beneficiarias, la distribución de materiales didácticos de química, la capacitación técnica y pedagógica de docentes en contenidos actualizados y estrategias innovadoras, así como la mejora estadística en los niveles de rendimiento académico de los estudiantes. Estas acciones permitirán mejorar las condiciones materiales y metodológicas del proceso educativo en ciencias, lo cual puede medirse mediante evaluaciones periódicas de desempeño y seguimiento institucional.

## 1.4.3 Beneficios Intangibles

Los beneficios intangibles están relacionados con el desarrollo de una actitud positiva hacia la ciencia, el fortalecimiento de la autoestima académica de los estudiantes, el compromiso docente con la innovación educativa y la motivación vocacional hacia carreras científicas y tecnológicas. Asimismo, el proyecto contribuirá a generar un clima escolar más participativo, colaborativo y estimulante, lo cual repercute en la calidad de vida escolar y en la relación entre docentes y alumnos.

## **1.5 Alcance del Proyecto**

El proyecto se implementará en instituciones públicas de nivel secundario, con especial énfasis en aquellas que presentan mayores dificultades en el área de ciencias. Incluirá intervenciones pedagógicas, infraestructurales y formativas. Su enfoque será regional o interprovincial, dependiendo del diagnóstico de necesidades educativas en cada jurisdicción. Las actividades estarán orientadas a mejorar directamente el aprendizaje en química, pero también servirán como modelo replicable para otras áreas de ciencias. El alcance temporal será de uno a dos años escolares, incluyendo fases de diagnóstico, implementación, monitoreo y evaluación. El impacto esperado trasciende el aula, promoviendo una cultura científica que favorezca el desarrollo personal, académico y social de los estudiantes involucrados.

# 

# 

# **CAPÍTULO III: ESTADO DEL ARTE** 3.1 Artículos

**3.1.1.** **How Peruvian Secondary Students View Scientists and their Works: Ready, Set, and Draw** [2]

**Cómo los estudiantes peruanos de secundaria ven a los científicos y sus trabajos: Listos, preparados y listos** [2]

**DOI:** [**https://doi.org/10.46328/ijemst.1099**](https://doi.org/10.46328/ijemst.1099)

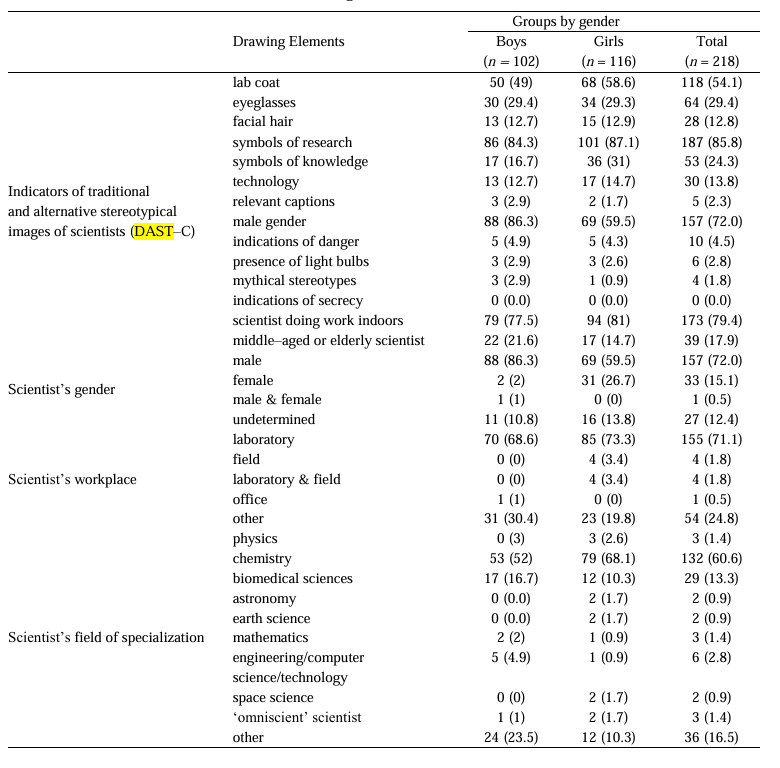
4 citaciones, fuente: Web Scopus

El artículo se propone investigar a fondo y analizar las percepciones y representaciones mentales que los estudiantes de secundaria de escuelas públicas peruanas tienen sobre los científicos y la naturaleza de su trabajo.

Esto implica ir más allá de una simple descripción para: identificar la prevalencia de estereotipos tradicionales (como el "científico loco" o el genio solitario en un laboratorio) entre estos estudiantes; comprender la influencia de factores socioculturales y educativos específicos del contexto peruano en la formación de estas imágenes; y determinar las implicaciones que estas percepciones tienen para el interés de los estudiantes en seguir carreras en Ciencia, Tecnología, Ingeniería y Matemáticas (STEM). En última instancia, la investigación busca proporcionar una base empírica sólida para el desarrollo de estrategias pedagógicas más efectivas que desafíen los estereotipos y fomenten una imagen más realista, inclusiva y atractiva de la ciencia.

**Metodologías:**

El estudio empleó un diseño de investigación descriptivo-exploratorio, combinando enfoques cualitativos y cuantitativos para obtener una comprensión integral de las percepciones de los estudiantes. Participaron 106 estudiantes de secundaria (de aproximadamente 13 a 16 años) de escuelas públicas en zonas rurales y urbanas de la región de Lima, Perú. El principal instrumento de recolección de datos fue el "Draw-A-Scientist Test" (DAST), una prueba ampliamente reconocida donde se pidió a los estudiantes dibujar a un científico trabajando y luego describir su dibujo. Los datos se analizaron cuantitativamente para identificar la prevalencia de estereotipos específicos (gafas, batas de laboratorio, equipo científico, etc.) y cualitativamente a través del análisis de las descripciones escritas, para comprender el significado subyacente de sus representaciones.

**Frecuencias (y Porcentajes) de Indicadores del DAST–C y Otros Elementos en los Dibujos de Estudiantes, según su Género.**

**Importancias:**

Alta Prevalencia de Estereotipos: El estudio reveló una marcada tendencia entre los estudiantes peruanos a dibujar científicos bajo estereotipos tradicionales: predominantemente masculinos, de raza blanca, mayores, con gafas, en laboratorios y rodeados de símbolos de investigación y, en ocasiones, de peligro.

Impacto Negativo en el Interés STEM: Estas percepciones estereotipadas son cruciales porque pueden actuar como barreras significativas, desalentando el interés de los estudiantes por las disciplinas científicas y tecnológicas. La imagen del "científico" como un personaje inalcanzable o poco diverso puede limitar la auto identificación con la ciencia, especialmente para las niñas y grupos minoritarios.

Necesidad de Intervención Educativa: El artículo enfatiza la urgencia de que los educadores aborden activamente estos estereotipos en el aula. Se sugiere la promoción de una imagen más realista y diversa de los científicos, la exposición a experiencias prácticas y a científicos reales de diferentes géneros y orígenes, y la integración de la historia de la ciencia para mostrar su evolución y variedad.

Contexto Peruano Específico: El valor añadido de este estudio radica en su enfoque en el contexto peruano, aportando una visión específica sobre las percepciones en una región donde la investigación de este tipo es menos común, lo que lo hace particularmente relevante para proyectos educativos en Perú.

**Utilidad del artículo para el proyecto de tesis**

Gracias al aporte de este artículo, al revelar que los estudiantes de secundaria peruanos tienen percepciones estereotipadas sobre los científicos y su trabajo, lo que a menudo lleva a que consideren la ciencia como "difícil", "aburrida" o "solo para genios", justifica profundamente la necesidad de un asistente inteligente. Para que el asistente inteligente sea realmente adaptativo, **implementaremos** el Test de Dibujo de Científicos (DAST) como herramienta para hallar cómo los estudiantes perciben la ciencia y a los científicos. A través de esta técnica, hallaremos estereotipos, preconcepciones y errores comunes en Química Básica. Con estos hallazgos, implementaremos estrategias de personalización que adapten los contenidos al contexto educativo peruano y promuevan una visión más cercana, inclusiva y motivadora de la química en estudiantes de secundaria de escuelas públicas.

**3.1.2.** **The Exploration of Predictors for Peruvian Teachers’ Life Satisfaction through an Ensemble of Feature Selection Methods and Machine Learning** [3]

**Exploración de predictores de la satisfacción vital de docentes peruanos mediante un conjunto de métodos de selección de características y aprendizaje automático** [3]

**DOI:** [**https://doi.org/10.3390/su16177532**](https://doi.org/10.3390/su16177532)

4 citaciones, fuente: Web Scopus

Este artículo se enfoca en una problemática crucial para el desarrollo educativo y social de Perú: la satisfacción con la vida de sus docentes. El bienestar de los maestros no es solo una cuestión de justicia social, sino que tiene un impacto directo y significativo en la calidad de la enseñanza y, por ende, en el aprendizaje de los estudiantes. En un país como Perú, con desafíos estructurales en su sistema educativo, comprender qué factores contribuyen a la felicidad y realización de sus docentes es fundamental para diseñar políticas y programas de apoyo efectivos.

El estudio parte de la premisa de que identificar y comprender los determinantes de la satisfacción con la vida en el ámbito docente puede llevar a intervenciones más precisas y eficientes. Tradicionalmente, se ha reconocido la importancia del entorno laboral, las condiciones salariales, el apoyo institucional y las relaciones interpersonales, entre otros, como factores influyentes. Sin embargo, este artículo busca ir más allá, utilizando herramientas avanzadas para desentrañar las relaciones complejas entre múltiples variables y la satisfacción docente en el contexto específico peruano. La investigación se propone no solo señalar qué factores son importantes, sino también cómo interactúan y cuáles tienen un peso predictivo más significativo, ofreciendo una visión más granular y accionable para los tomadores de decisiones.

**Metodologías:**

Para abordar esta compleja tarea, el estudio empleó una combinación rigurosa de técnicas de análisis de datos:

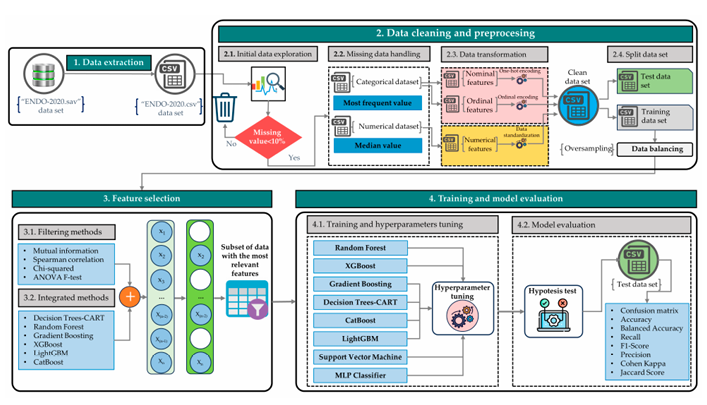
Recopilación de datos: Se inició con la obtención de datos a través de encuestas dirigidas a una muestra representativa de docentes de diversas regiones de Perú. Estas encuestas habrían incluido una amplia gama de preguntas relacionadas con su vida profesional y personal, percepción de su bienestar, condiciones de trabajo, apoyo recibido, entre otros aspectos que podrían influir en su satisfacción.

Selección de características (Feature Selection): Un paso crítico en el análisis de datos complejos es la identificación de las variables más relevantes, es decir, aquellas que tienen el mayor poder predictivo y no son redundantes. El estudio destacó por utilizar un enfoque de "ensamble" o combinación de diversos métodos de selección de características. Esto implica que, en lugar de depender de un único algoritmo que podría tener sesgos o limitaciones, se consolidaron los resultados de múltiples técnicas de selección (por ejemplo, métodos basados en filtros como la correlación, en envolturas como la eliminación recursiva de características, o incrustados como la regresión Lasso). Esta aproximación combinada permitió obtener una selección de variables más robusta, confiable y menos propensa a errores, optimizando la interpretabilidad y eficiencia de los modelos posteriores.

Aprendizaje Automático (Machine Learning): Una vez que se identificaron las características más influyentes a través del proceso de selección, se aplicaron diversos algoritmos de aprendizaje automático. Estos algoritmos fueron entrenados para modelar la relación entre las características seleccionadas y la satisfacción con la vida de los docentes, permitiendo predecir y comprender mejor cómo estas variables se asocian con el nivel de bienestar reportado. Si bien el resumen general no detalla los algoritmos específicos, este tipo de estudios suelen emplear modelos como regresión lineal, árboles de decisión, Random Forest, o modelos de boosting, entre otros, para construir un modelo predictivo preciso.

Análisis estadístico: Complementariamente a los métodos de aprendizaje automático, se realizaron análisis estadísticos tradicionales para validar las relaciones encontradas, determinar la significancia estadística de los predictores identificados y proporcionar una base sólida para las conclusiones del estudio.

***Propuesta metodológica para la exploración de predictores de la satisfacción vital en maestros de primaria en Perú***



**Importancias:**

Identificación de factores clave: La principal importancia radica en la capacidad del estudio para identificar de manera precisa los factores específicos que tienen una influencia significativa en la satisfacción con la vida de los docentes peruanos. Esta información es crucial para el diseño de políticas públicas y programas de bienestar que sean verdaderamente pertinentes y efectivos.

Bienestar docente como pilar educativo: El artículo subraya la trascendencia del bienestar docente, no solo como un derecho de los profesionales de la enseñanza, sino como un elemento fundamental para la calidad del sistema educativo. Docentes satisfechos y motivados son más propensos a ser efectivos en el aula y a permanecer en la profesión, contribuyendo a la estabilidad y mejora continua de la educación.

**Utilidad del artículo para el proyecto de tesis**

Gracias al artículo la implementación del asistente inteligente busca apoyar a los maestros y mejorar la calidad de la enseñanza en Química Básica. Su objetivo es potenciar el aprendizaje de los estudiantes mediante inteligencia artificial. El estudio destacó por utilizar un enfoque integral y aplicar resultados de múltiples técnicas de selección para personalizar la enseñanza en escuelas públicas peruanas.Mediante técnicas de selección de características y algoritmos de Machine Learning**, hallaremos** los factores que más influyen en el aprendizaje de Química Básica. Para ello, se recopilarán datos relevantes de los estudiantes, que luego serán preprocesados cuidadosamente —tratando datos faltantes y balanceando clases si es necesario—. Al aplicar modelos como LightGBM o Random Forest, el asistente inteligente podrá identificar patrones clave y generar intervenciones personalizadas que mejoren el rendimiento académico de los estudiantes de secundaria en escuelas públicas.

Preguntar a ChatGPT

**3.1.3.** **Modeling Job Satisfaction of Peruvian Basic Education Teachers Using Machine Learning Techniques** [4]

**Modelado de la satisfacción laboral de docentes de educación básica peruana mediante técnicas de aprendizaje automático** [4]

5 citaciones, fuente: Web Scopus

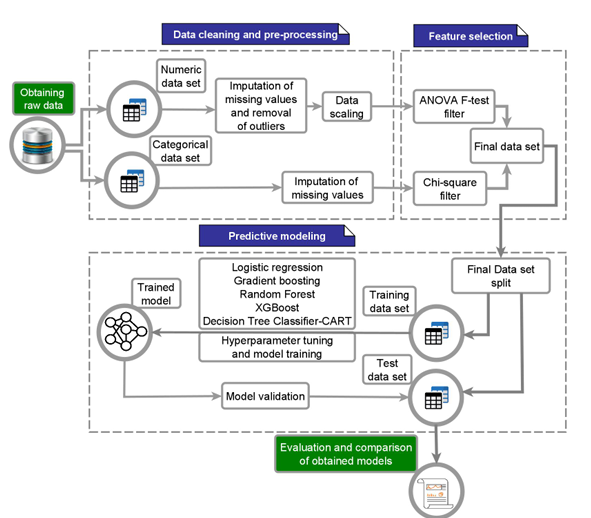
**DOI: https://doi.org/10.3390/app13063945**

El artículo titulado "Modeling Job Satisfaction of Peruvian Basic Education Teachers Using Machine Learning Techniques" presenta un estudio que busca modelar y predecir la satisfacción laboral de los docentes de educación básica en Perú mediante técnicas de aprendizaje automático. Utilizando un conjunto de datos provenientes de la Encuesta Nacional a Docentes (ENDO-2018), los investigadores llevaron a cabo un proceso de limpieza y pre procesamiento de la información, seguido de una selección de variables relevantes mediante el análisis estadístico ANOVA F-test y Chi-Square. Posteriormente, implementaron diferentes modelos predictivos, incluyendo regresión logística, Decision Trees-CART, Random Forest, Gradient Boosting y XGBoost, logrando una precisión en torno al 74%.

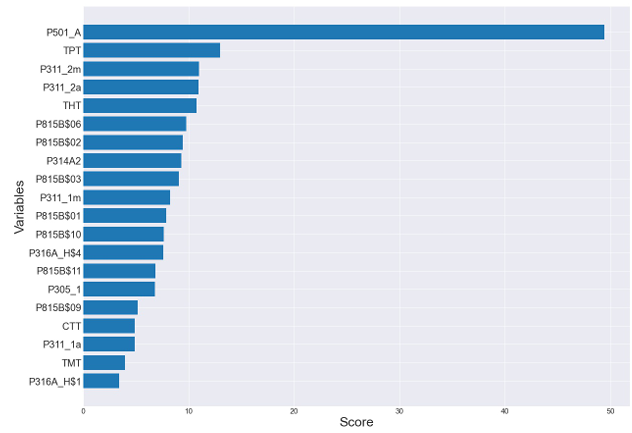
**Metodologías utilizadas:**

* **Selección de variables relevantes:** El estudio empleó técnicas como el análisis de varianza (ANOVA F-test) y la prueba de Chi-Cuadrado para filtrar y seleccionar los atributos más influyentes en la satisfacción laboral. Esto permite reducir la complejidad del conjunto de datos (de 942 atributos a 11 variables relevantes) y enfocar el análisis en factores clave, facilitando interpretaciones claras y acciones concretas.
* **Preprocesamiento de datos:** Se realizó un trabajo exhaustivo en limpiar y transformar la información, diferenciando entre variables categóricas y numéricas, lo cual es esencial para obtener modelos predictivos precisos y confiables en contextos reales y complejos.
* **Modelado con algoritmos de machine learning:** La implementación y comparación de diferentes modelos como regresión logística, Decision Trees, Random Forest, Gradient Boosting y XGBoost permitieron determinar qué herramientas ofrecen mayor desempeño en predicciones de satisfacción laboral. La obtención de métricas como precisión del 74%, sensibilidad igual, y alto valor predictivo negativo demuestra que estos modelos pueden identificar con confiabilidad quiénes están satisfechos o insatisfechos en su trabajo.

*Metodología propuesta para el modelado de la satisfacción laboral docente*



*Puntuaciones ANOVA en la predicción de la satisfacción laboral docente*



**Importancia:**

Este trabajo es pionero en aplicar técnicas avanzadas de machine learning para estudiar un fenómeno complejo en el contexto educativo peruano, en un escenario donde tradicionalmente predominaban metodologías estadísticas descriptivas o modelos estructurales. La capacidad de predecir la satisfacción laboral con buen grado de exactitud abre nuevas posibilidades para que las instituciones educativas puedan diseñar intervenciones específicas, dirigidas a mejorar las condiciones laborales y motivar a los docentes.

**Utilidad del artículo para el proyecto de tesis**

Mediante el artículo se demuestra cómo la implementación del asistente inteligente en Química Básica aplica técnicas de aprendizaje automático y busca modelar y predecir el rendimiento estudiantil a partir de sus interacciones. Se emplearán modelos como la regresión logística, cuya precisión se evaluará mediante la comparación con otros algoritmos. El estudio brinda un marco metodológico valioso para implementar un asistente inteligente en Química Básica. **Implementaremos** enfoques como Random Forest o XGBoost para hallar patrones que permitan predecir el rendimiento, detectar dificultades y personalizar el aprendizaje. Para ello, recopilaremos y preprocesaremos datos detallados —mediante codificación, escalado y tratamiento de valores perdidos— y evaluaremos modelos usando validación cruzada. El análisis de importancia de características nos permitirá hallar qué aspectos inciden más en el aprendizaje y generar intervenciones específicas que fortalezcan el dominio de la materia en estudiantes de secundaria.

**3.1.4. Application of Artificial Neural Networks to Predict the Use of Mobile Learning by University Student** [5]

**Aplicación de redes neuronales artificiales para predecir el uso del aprendizaje móvil por parte de estudiantes universitarios** [5]

1 citaciones, fuente: Web Scopus

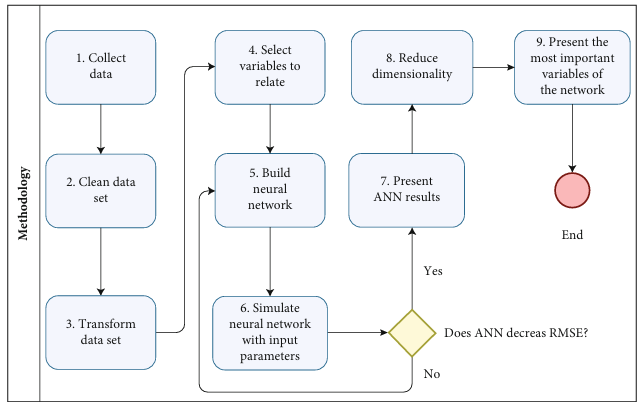
**DOI:** [**https://doi.org/10.1155/hbe2/1518987**](https://doi.org/10.1155/hbe2/1518987)

El artículo de investigación, "Aplicación de Redes Neuronales Artificiales para Predecir el Uso del Aprendizaje Móvil por Estudiantes Universitarios", se enfoca en determinar la intención de estudiantes de educación superior peruanos de usar dispositivos móviles para el aprendizaje. Para lograr esto, el estudio propone un modelo predictivo basado en Redes Neuronales Artificiales (ANNs). Las ANNs son técnicas de aprendizaje automático supervisado que imitan la organización y operación del cerebro humano para procesar datos y tomar decisiones. Son sistemas informáticos que pueden aprender de la observación y la experiencia, y luego usar ese conocimiento para reconocer patrones y hacer predicciones.

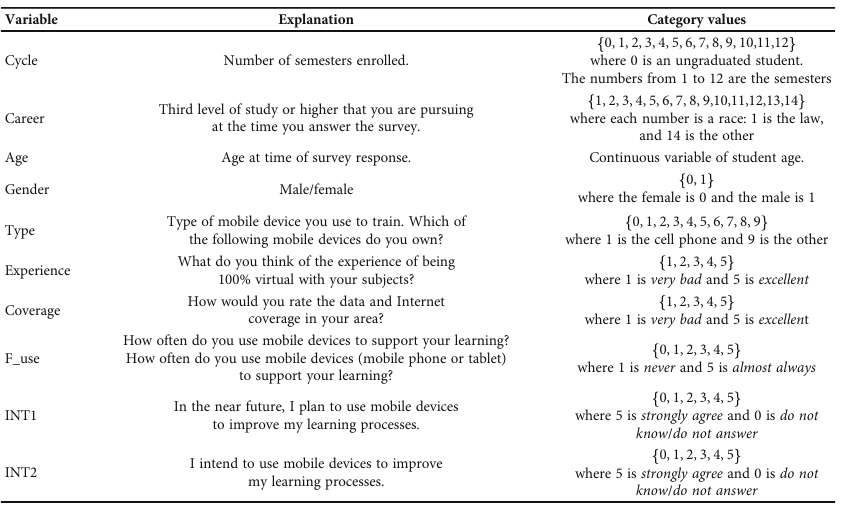
**Metodología**

La metodología del estudio implica la recolección de 6520 puntos de datos a través de una encuesta sobre la intención de usar dispositivos móviles para el aprendizaje en el sector de educación terciaria peruana. Los datos se limpiaron, transformaron y luego se utilizaron para construir la red neuronal.

La arquitectura de la ANN considera una capa de entrada con variables como ciclo, carrera, edad, género, tipo de dispositivo, experiencia con materias virtuales, cobertura de internet y frecuencia de uso. La red también incluye una capa oculta con un número definido de neuronas y una capa de salida con variables dependientes que indican la intención de aprender (INT1 y INT2). Se probaron cuatro configuraciones de ANN con 10, 15, 20 y 25 neuronas en la capa oculta para simular el rendimiento óptimo. La evaluación del rendimiento de la ANN se realizó utilizando métricas como el error cuadrático medio (MSE) y el error cuadrático medio de la raíz (RMSE).

*Metodología*

*Variables y categorías*



**Importancia y Resultados Resaltantes**

Entre los hallazgos principales, se encontró que la ANN con 10 neuronas en su capa oculta mostró el rendimiento óptimo, con valores de RMSE más bajos. Esto indica una mejor alineación entre las predicciones del modelo y los valores reales. Factores como la experiencia con asignaturas virtuales, la frecuencia de uso de dispositivos móviles para el aprendizaje y la cobertura de internet son cruciales para las dos variables de intención (INT1 y INT2). Para la variable INT1, la carrera también se consideró importante.

El estudio resalta la capacidad de las ANNs para modelar relaciones no lineales complejas entre variables, lo que las hace superiores a los métodos estadísticos tradicionales como la regresión lineal múltiple y el modelado de ecuaciones estructurales para predecir el uso del aprendizaje móvil. Además, la técnica de línea de regresión multivariable adaptativa (MARS) se empleó para reducir la dimensionalidad y medir la relevancia de las variables de entrada

**Utilidad del artículo para el proyecto de tesis**

Gracias al aporte del artículo, se verifica la iniciativa se enfoca en determinar cómo un asistente inteligente puede mejorar el aprendizaje en Química Básica mediante el uso de dispositivos móviles y tecnologías de aprendizaje automático. Este asistente interactúa con materias virtuales, adaptando los contenidos según el desempeño del estudiante. Para ello, se aplicarán Redes Neuronales Artificiales (ANNs), con el fin de personalizar la experiencia educativa y fomentar un aprendizaje más efectivo en contextos escolares públicos peruanos, La investigación nos brinda una base útil para implementar un asistente inteligente en Química Básica. A partir de datos sobre la interacción y el progreso del estudiante, **implementaremos** redes neuronales que, tras un preprocesamiento riguroso y un diseño iterativo, hallarán los factores que más influyen en el aprendizaje. Este enfoque permitirá ofrecer una experiencia personalizada, adaptada al ritmo y perfil de cada estudiante de secundaria en escuelas públicas.



**3.1.5. Density-Based Unsupervised Learning Algorithm to Categorize College Students into Dropout Risk Levels** [6]

**Algoritmo de aprendizaje no supervisado basado en densidad para categorizar a estudiantes universitarios según niveles de riesgo de abandono escolar**

[6]

11 citaciones, fuente: Web Scopus

**DOI:** [**https://doi.org/10.3390/data7110165**](https://doi.org/10.3390/data7110165)

El artículo de investigación titulado "Density-Based Unsupervised Learning Algorithm to Categorize College Students into Dropout Risk Levels" aborda una problemática crítica y persistente a nivel global: la deserción universitaria. Este fenómeno no solo representa una pérdida significativa para el estudiante, afectando su futuro profesional y personal, sino que también impone cargas económicas y de recursos considerables a las instituciones educativas y a la sociedad en general. La investigación se enfoca en desarrollar una solución innovadora para este desafío, proponiendo un **algoritmo de aprendizaje no supervisado basado en densidad** con el objetivo de **categorizar a los estudiantes universitarios en diferentes niveles de riesgo de deserción.** Esto permite identificar proactivamente a aquellos estudiantes que necesitan apoyo, incluso antes de que muestren signos evidentes de abandono, lo cual es fundamental para implementar estrategias de retención efectivas.

**Metodología Detallada:**

La metodología central del estudio se articula en torno al uso de **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**, un algoritmo de clustering no supervisado. A diferencia de otros métodos que requieren definir el número de clústeres de antemano (como K-Means), DBSCAN tiene la ventaja de poder identificar grupos de forma arbitraria y detectar el "ruido" o valores atípicos, lo que lo hace ideal para explorar patrones ocultos en datos complejos sin suposiciones previas.

El proceso metodológico se estructuró en varias etapas clave:

**Recolección y Caracterización de Datos:** Se recopiló un conjunto de datos robusto de estudiantes de ingeniería de una universidad peruana. Lo distintivo de este conjunto de datos es su amplitud, que abarca una variedad de variables cruciales para entender el comportamiento del estudiante:

**Variables Sociodemográficas:** Incluyen datos básicos como edad, género, lugar de origen, que pueden influir en la adaptación a la vida universitaria.

**Variables Socioeconómicas:** Aspectos como el nivel de ingresos familiares y el tipo de escuela secundaria de procedencia (pública vs. privada) son importantes, ya que el contexto socioeconómico puede impactar el acceso a recursos y la estabilidad del estudiante.

**Variables Psicométricas:** Se incorporaron indicadores de salud mental, como los niveles de ansiedad y depresión. Esta es una adición importante, ya que reconoce que el bienestar emocional juega un papel crucial en la permanencia académica.

**Variables de Gestión Académica:** Comprenden el rendimiento académico directo, como las notas promedio, el número de cursos reprobados y la asistencia a clases, que son indicadores directos del progreso del estudiante.

**Variables de Contexto Virtual:** Consideran la interacción del estudiante con plataformas de aprendizaje en línea o herramientas digitales, un aspecto cada vez más relevante en la educación moderna.

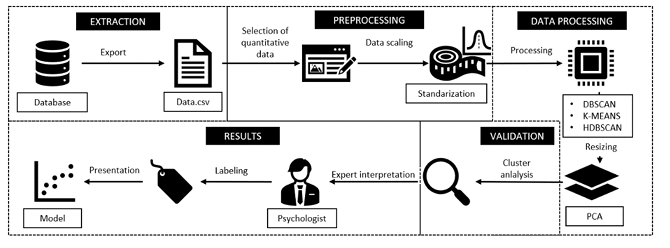
**Preprocesamiento de Datos Riguroso:** Para asegurar la calidad y la efectividad del modelo, se llevaron a cabo pasos de preprocesamiento esenciales:

**Importancia y Resultados Resaltantes:**

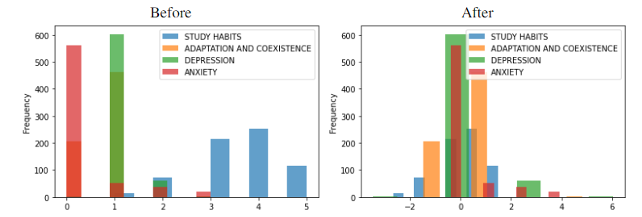
La principal importancia de este artículo radica en su contribución a la predicción y prevención de la deserción universitaria mediante un enfoque innovador de aprendizaje no supervisado. Esto es fundamental porque no requiere etiquetas de "abandono" previas (es decir, no necesita saber de antemano quién abandonó y quién no), lo que lo hace aplicable en contextos donde estos datos son escasos o difíciles de obtener.

**Los resultados más resaltantes incluyen:**

* La exitosa categorización de los estudiantes en distintos niveles de riesgo de deserción (bajo, medio y alto). Esta capacidad de segmentación permite a las instituciones diseñar intervenciones diferenciadas y dirigidas.
* La identificación de patrones y características específicas asociadas a cada grupo de riesgo. Por ejemplo, el estudio encontró que factores socioeconómicos, psicométricos (como la ansiedad y la depresión) y de rendimiento académico son cruciales para determinar el nivel de riesgo de deserción. Esto proporciona información valiosa sobre las causas subyacentes de la deserción.
* La demostración de que la utilización de DBSCAN es efectiva para descubrir estructuras de datos complejas que otros algoritmos de clustering supervisado o no supervisado tradicionales podrían pasar por alto. Esto lleva a una comprensión más profunda de los factores que contribuyen a la deserción estudiantil, permitiendo intervenciones más precisas.



*Modelo propuesto*



*Contraste de distribuciones de datos antes y después de la normalización*

**Utilidad del artículo para el proyecto de tesis**

El siguiente artículo se enfoca en ofrecer una solución innovadora para mejorar el aprendizaje, adaptando estrategias previamente aplicadas en estudiantes universitarios al contexto de secundaria en escuelas públicas peruanas. En estudios anteriores, se recopiló un conjunto de datos educativos que permitió analizar factores como el rendimiento académico y la estabilidad del estudiante. Uno de los principales aportes del enfoque basado en Machine Learning es su amplitud, ya que permite integrar múltiples variables y personalizar la experiencia de aprendizaje según las necesidades individuales. El uso de algoritmos de aprendizaje no supervisado, como DBSCAN, presentado en el estudio sobre categorización de estudiantes según riesgo de deserción, ofrece una técnica aplicable al asistente inteligente en Química Básica. Esta metodología permite agrupar automáticamente a los estudiantes según sus patrones de interacción, errores frecuentes o estilos de aprendizaje, revelando perfiles que no serían evidentes mediante enfoques tradicionales. Al identificar estos grupos, el asistente puede adaptar el contenido y la retroalimentación —como ofrecer explicaciones visuales a algunos estudiantes o más ejercicios prácticos a otros e incluso detectar casos atípicos que requieran intervenciones diferenciadas.

**3.1.6. An analysis of diverse computational models for predicting student achievement on e-learning platforms using machine learning**. (Koti Mani Kumar Tirumanadham et al., 2024)

**3.1.6. Análisis de diversos modelos computacionales para predecir el rendimiento estudiantil en plataformas de aprendizaje electrónico mediante aprendizaje automático.** [7]

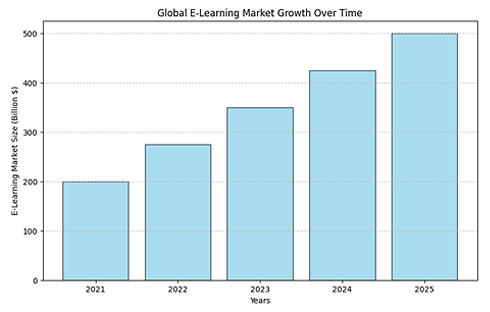
6 citaciones, fuente: Web Scopus

**DOI:** 10.11591/ijece.v14i6.pp7013-7021

El estudio aborda el desafío de evaluar el desarrollo estudiantil en entornos de aprendizaje electrónico (e-learning), especialmente con la gran cantidad de datos en las bases de datos educativas. Dada la creciente integración de las plataformas de e-learning en la educación, comprender y mejorar el rendimiento de los estudiantes en estos entornos es crucial para la efectividad de las actividades educativas en línea.

**Metodologías clave:**

* **Manejo de datos desequilibrados:** Se utiliza la técnica de sobre muestreó de minorías sintéticas (SMOTE) para equilibrar los conjuntos de datos, que a menudo están desequilibrados en los entornos educativos. Esto es fundamental porque los conjuntos de datos desequilibrados pueden sesgar los modelos predictivos hacia las clases mayoritarias.
* **Selección híbrida de características (R³FE):** Esta es una contribución importante del estudio. Combina la regularización L2 (Ridge) y la eliminación recursiva de características (RFE).
  + La regularización L2 (Ridge) reduce los coeficientes de las características menos importantes hacia cero.
  + RFE elimina iterativamente las características menos importantes, construyendo modelos sobre las restantes para encontrar el subconjunto óptimo.
  + La combinación de estos métodos, denominada R³FE, genera un conjunto de características óptimo. Este enfoque no solo mejora la precisión de la predicción, sino que también reduce el sobreajuste y aumenta la interpretabilidad del modelo.
* **Modelo de conjunto (Ensemble Model):** El estudio propone un modelo de conjunto que integra tres algoritmos robustos: Random Forest, Gradient Boosting y AdaBoost.
  + **Random Forest:** Genera múltiples árboles de decisión y agrega sus resultados para aumentar la precisión y la robustez.
  + **Gradient Boosting:** Construye modelos secuencialmente para corregir los errores de los anteriores.
  + **AdaBoost:** También construye modelos secuencialmente, pero se centra en los eventos mal clasificados al alterar sus pesos, lo que mejora la resiliencia y la precisión, especialmente en conjuntos de datos ruidosos. La combinación de estos tres enfoques aprovecha sus fortalezas individuales para lograr un rendimiento predictivo superior, mayor precisión, robustez e interpretabilidad en la predicción del rendimiento estudiantil.



*Crecimiento del mercado mundial del aprendizaje electrónico a lo largo del tiempo*

**Importancia y resultados sobresalientes:**

* **Mejora de la precisión de la predicción:** El estudio enfatiza que estas técnicas, especialmente la selección de características híbridas y el manejo de datos desequilibrados, mejoran significativamente la precisión de la predicción y abordan los desafíos de los grandes y complejos conjuntos de datos educativos.
* **Rendimiento del modelo propuesto:** El modelo híbrido R³FE + Ensemble mostró un rendimiento "sobresaliente" con las siguientes métricas:
  + Precisión (Accuracy): 97%
  + Exactitud (Precision): 96%
  + Sensibilidad (Recall): 98%
  + Puntuación F1 (F1-Score): 97% Estos resultados demuestran la solidez y consistencia del modelo para predecir el rendimiento académico.
* **Características clave identificadas:** El proceso de selección de características identificó 10 de las 16 características principales con el mayor valor predictivo. Estas incluyen género, stage\_id, grade\_id, subject, related, raisedhands, visited\_resources, announcements\_view, discussion y student absence days. Estas características ofrecen una visión completa del perfil del estudiante y su participación en el sistema de e-learning.

**Utilidad del artículo para el proyecto de tesis**

Gracias a este artículo se verifica que la creciente integración de las plataformas de e-learning en contextos educativos en línea ha impulsado el uso de técnicas de Machine Learning. El enfoque enfatiza que estas técnicas permiten trabajar con datos desequilibrados y aplicar selección de características para identificar variables con mayor valor predictivo, mejorando así el aprendizaje en Química Básica en escuelas públicas. En síntesis, el estudio sobre modelos computacionales para predecir el rendimiento en plataformas e-learning aporta técnicas valiosas que pueden integrarse en el asistente inteligente de Química Básica. Entre ellas, la selección de características híbrida (R³FE) permite identificar variables estudiantiles clave, mientras que técnicas como SMOTE ayudan a equilibrar los conjuntos de datos. Combinado con modelos de aprendizaje de conjunto, como Random Forest o Gradient Boosting, se pueden generar predicciones precisas del rendimiento. Esto permitirá al asistente ofrecer intervenciones personalizadas y efectivas que optimicen el proceso de aprendizaje de la química.

****

**3.1.7. La Inteligencia Artificial y los asistentes virtuales: uso e incidencia en el aprendizaje y en el desarrollo de proyectos de estudiantes de pregrado y posgrado en una Facultad del sureste de México**[8]

**3.1.7. Artificial Intelligence and virtual assistants: use and impact on learning and project development of undergraduate and graduate students in a faculty in southeastern Mexico** (Silva Payró et al., 2025)

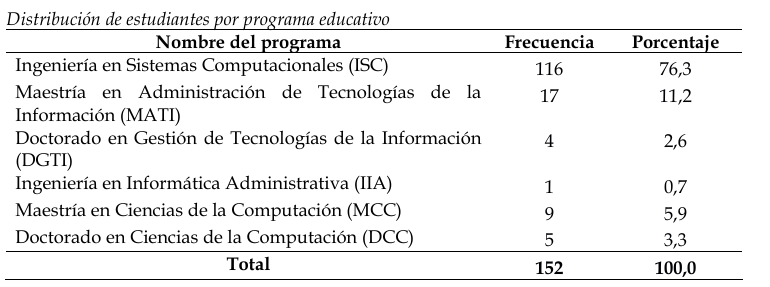
1 citaciones, fuente: Web Scopus

**DOI:**[**https://doi.org/10.31637/epsir-2025-1378**](https://doi.org/10.31637/epsir-2025-1378)

El artículo "La Inteligencia Artificial y los asistentes virtuales: uso e incidencia en el aprendizaje y en el desarrollo de proyectos de estudiantes de pregrado y posgrado en una Facultad del sureste de México" tiene como objetivo principal comprender el uso y la percepción de las herramientas de Inteligencia Artificial (IA) entre estudiantes de educación superior.

**Metodologías:**

* **Enfoque Cuantitativo Descriptivo, No Experimental y Transversal:** El estudio se llevó a cabo con una muestra de 152 estudiantes de Tecnologías de la Información. Esto implica que se recolectaron datos numéricos para describir el fenómeno de estudio, sin manipular variables ni establecer relaciones de causa y efecto, en un único momento en el tiempo.
* **Muestreo no probabilístico por conveniencia:** La selección de los participantes se realizó de forma no aleatoria, incluyendo a estudiantes matriculados en programas de pregrado y posgrado de una Facultad de Tecnologías de la Información en una universidad pública del sureste de México. La muestra estuvo conformada por 117 hombres (77%) y 35 mujeres (23%), con edades entre 18 y 55 años (media de 22.73 años).
* **Recolección de Información:** Se utilizó un cuestionario digital distribuido a través de Google Forms de enero a mayo de 2024. El cuestionario se dividió en dos secciones: la primera para datos sociodemográficos (edad, género, programa educativo, porcentaje de avance curricular, tiempo de estudio) y la segunda con 10 preguntas sobre el uso de asistentes virtuales con IA, adaptadas de estudios previos.
* **Análisis de la Información:** Se generó una base de datos con IBM SPSS Statistic 21, utilizando medidas de tendencia central (media) y dispersión (desviación estándar), así como estadísticas descriptivas y distribución de frecuencias.

****

**Importancias y Aspectos Más Resaltantes:**

* **Percepción Positiva a pesar de la Familiaridad Limitada:** A pesar de que los estudiantes no están muy familiarizados con chatbots o bocinas inteligentes para el aprendizaje autodirigido, consideran que el uso de asistentes virtuales inteligentes es claro y sencillo. Además, creen que la implementación de IA en sus asignaturas aumentaría su rendimiento académico.
* **Potencial para el Rendimiento Académico:** Se sugiere que el uso de herramientas de IA en los programas de estudio de nivel superior puede fortalecer las técnicas de estudio de los alumnos y mejorar su formación.
* **Necesidad de Formación y Supervisión:** Los estudiantes no se sienten capaces de diseñar un proyecto de IA sin ayuda. Por ello, se propone que los proyectos de IA que desarrollen estén vinculados a sus disciplinas, adaptados a los planes de estudio y cuenten con la supervisión del profesorado para un uso provechoso.
* **Oportunidad para la Innovación Educativa:** El estudio concluye que, aunque la adopción actual es limitada, existe un potencial significativo para la mejora del rendimiento académico, el desarrollo de habilidades, y la innovación educativa a través de la IA, siempre y cuando se aborde la seguridad y la confianza en su uso, así como la formación adecuada.
* **IA como Catalizador de Innovación:** La Inteligencia Artificial es vista como un potente catalizador de innovación y desarrollo tecnológico en todos los sectores, incluido el educativo, donde es fundamental para una educación moderna y de vanguardia. La UNESCO estima que la IA en educación alcanzará un valor de 6.000 millones de dólares para 2024 y puede ayudar a enfrentar desafíos educativos y acelerar el progreso hacia el Objetivo de Desarrollo Sostenible (ODS) 4.

**Utilidad del artículo para el proyecto de tesis**

El articulo presenta un estudio con una muestra de 152 estudiantes de Tecnologías de una Facultad de Tecnologías, se analizaron variables como edad, género, programa educativo, porcentaje de avance curricular y tiempo de estudio. Una segunda sección con 10 preguntas sobre el uso de asistentes virtuales con IA mostró que muchos no están muy familiarizados con chatbots, lo que refuerza la necesidad de desarrollar asistentes educativos con un diseño claro y sencillo. En el contexto del asistente inteligente para Química Básica, esto implica incorporar funcionalidades de chatbot que permitan resolver dudas en tiempo real, ofrecer rutas de aprendizaje personalizadas y mantener una interfaz intuitiva y accesible. Para garantizar su efectividad, es esencial asegurar su integración curricular y su uso bajo supervisión docente, de modo que complemente y potencie la enseñanza tradicional de la química.

****

**3.1.8.** **Prediction of Students’ Academic Performance in the Programming Fundamentals Course Using Long Short-Term Memory Neural Networks** [9]

**3.1.8.** **Predicción del rendimiento académico de los estudiantes en el curso Fundamentos de Programación mediante redes neuronales de memoria a largo plazo y a corto plazo** [9]

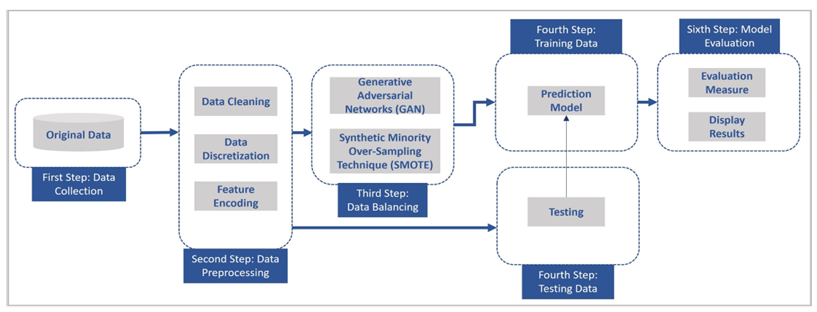
16 citaciones, fuente: Web Scopus

**DOI:**10.1109/ACCESS.2024.3350169

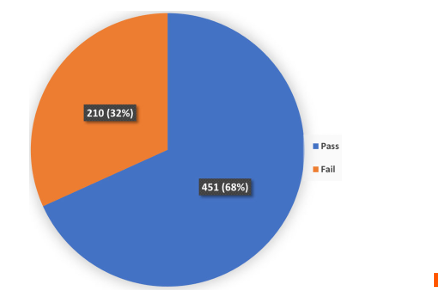
El artículo "Prediction of Students' Academic Performance in the Programming Fundamentals Course Using Long Short-Term Memory Neural Networks" se enfoca en la predicción del rendimiento académico de estudiantes universitarios peruanos en el curso de "Fundamentos de Programación" para identificar a aquellos en riesgo de reprobar. Esto es crucial para que las universidades implementen estrategias que mejoren la calidad de la enseñanza y reduzcan la deserción y el fracaso académico, desafíos que se han incrementado en Perú, especialmente después de la pandemia de COVID-19.

**Metodologías:**

* **Minería de Datos Educativos (EDM) y Machine Learning:** La investigación se basa en la Minería de Datos Educativos (EDM), que aplica estadísticas y Machine Learning para extraer patrones de conjuntos de datos educativos. Se utilizan diversas técnicas de Machine Learning para generar modelos predictivos.
* **Modelos Predictivos Comparados:** El estudio compara la eficiencia de las Redes Neuronales de Memoria Larga a Corto Plazo (LSTM) con otros modelos predictivos como Redes Neuronales Profundas (DNN), Árbol de Decisión (DT), Random Forest (RF), Regresión Logística (LR), Clasificador de Máquinas de Soporte Vectorial (SVM) y K-Nearest Neighbor (KNN).
* **Manejo de Datos Desequilibrados:** Un desafío importante en los algoritmos de Machine Learning es el desequilibrio de clases en los conjuntos de datos, lo que puede llevar a un sobreajuste y baja precisión. Para abordar esto, la investigación utiliza dos técnicas de remuestreo: Redes Generativas Antagónicas (GAN) y la Técnica de Sobremuestreo Sintético de Minorías (SMOTE) para balancear los datos.
* **Fases del Marco Propuesto:** La metodología se divide en 5 fases: Recolección de Datos, Balanceo de Datos, Entrenamiento de Datos, Pruebas de Datos y Evaluación del Modelo.
* **Evaluación del Modelo:** Se utilizaron métricas de rendimiento cuantitativas como precisión (accuracy), exactitud (precision), exhaustividad (recall) y puntuación F1 (F1-Score), error de clasificación, sensibilidad y especificidad para verificar los resultados de los modelos. Se empleó validación cruzada estratificada K-fold con k=5 para todos los modelos evaluados.
* **Arquitectura LSTM:** El diseño de la red LSTM consta de un codificador LSTM con 64 neuronas y una función de activación ReLu, un decodificador LSTM con una capa LSTM de 32 neuronas y una función ReLu, y una capa densa con dos neuronas y función SoftMax para predecir si el estudiante aprueba o reprueba.



*Diagrama de flujo del enfoque propuesto*



*Distribución desequilibrada de la etiqueta de clase*

**Importancias y Aspectos Más Resaltantes:**

* **Alta Precisión de Predicción:** El modelo propuesto demostró una superioridad en la clasificación. El modelo LSTM-GAN obtuvo una precisión del 98.3% en la semana 8, seguido por DNN-GAN con 98.1% de precisión. Esto es crucial para la identificación temprana de estudiantes en riesgo.
* **Contexto Peruano:** El estudio se enfoca en universidades peruanas, abordando las altas tasas de deserción y fracaso en carreras de ingeniería como Ciencias de la Computación, Ingeniería de Software e Ingeniería de Sistemas de Información, que oscilan entre el 15% y 20% para deserción y 25% y 30% para fracaso.
* **Abordaje del Desequilibrio de Clases:** La inclusión de GAN y SMOTE para balancear los datos es un punto clave, ya que este desequilibrio es un desafío importante para los algoritmos de Machine Learning y puede causar inestabilidad en la precisión de los resultados.

**Utilidad del artículo para el proyecto de tesis**

El articulo nos facilita una propuesta que se basa en la Minería de Datos Educativos (EDM) para optimizar el aprendizaje en Química Básica mediante un asistente inteligente. El proceso inicia con la recolección de datos sobre el desempeño de los estudiantes, lo cual es importante para los algoritmos de Machine Learning al momento de entrenar modelos precisos. Se considera el diseño de la red LSTM para analizar secuencias de interacción y el uso combinado de GAN y SMOTE para equilibrar los datos y mejorar la calidad del entrenamiento en contextos educativos con información limitada o desequilibrada. El uso de redes neuronales LSTM, propuesto en el estudio sobre predicción del rendimiento en cursos de programación, resulta especialmente adecuado para el asistente inteligente en Química Básica, ya que permite analizar secuencias de aprendizaje a lo largo del tiempo. Esta técnica puede modelar la trayectoria educativa de cada estudiante, prediciendo su rendimiento futuro o identificando posibles dificultades en temas específicos a partir de su historial de interacciones, aciertos y errores. Con esta información, el asistente podrá activar intervenciones personalizadas de manera anticipada, como refuerzos o recursos adicionales, optimizando así el apoyo al estudiante y mejorando su desempeño en el curso.

****

**3.1.9.** **Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study** [10]

**3.1.9.** **Algoritmos de aprendizaje automático supervisado para predecir la deserción escolar y el éxito académico: un estudio comparativo** [10]

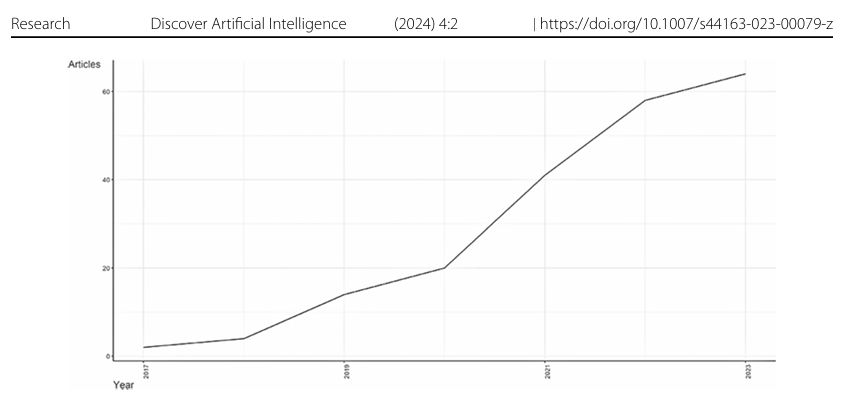
21 citaciones, fuente: Web Scopus

**DOI:**10.1007/s44163-023-00079-z

El artículo "Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study" tiene como objetivo evaluar la eficacia de diversos algoritmos de Machine Learning para predecir el abandono estudiantil y el éxito académico, utilizando un conjunto de datos de una institución de educación superior. El estudio se enfoca en algoritmos capaces de manejar eficazmente datos desequilibrados.

**Metodologías:**

* **Manejo de Datos Desequilibrados:** Para abordar el desequilibrio de clases, se empleó la técnica de remuestreo SMOTE (Synthetic Minority Over-sampling Technique).
* **Algoritmos de Machine Learning:** Se aplicó una variedad de algoritmos, incluyendo Árbol de Decisión (DT), Máquina de Soporte Vectorial (SVM), Random Forest (RF) , así como algoritmos de *boosting* como Gradient Boosting (GB), Extreme Gradient Boosting (XGBoost), CatBoost (CB) y Light Gradient Boosting Machine (LightGBM) (LB).
* **Optimización de Hiperparámetros:** Para mejorar el rendimiento de los modelos, se realizó la sintonización de hiperparámetros utilizando Optuna.
* **Detección de Valores Atípicos:** Se empleó el método Isolation Forest (IF) para identificar valores atípicos o anomalías dentro del conjunto de datos.
* **Análisis Comparativo:** El estudio es un análisis comparativo que explora varios métodos para predecir el éxito y el abandono estudiantil en instituciones de educación superior.



*Producción científica anual*

**Importancias y Aspectos Más Resaltantes:**

* **Rendimiento Superior de Algoritmos de *Boosting*:** Los hallazgos indican que los algoritmos de *boosting*, particularmente LightGBM y CatBoost con Optuna, superaron a los métodos de clasificación tradicionales. Por ejemplo, el modelo CatBoost inicial obtuvo un F1-Score de 0.86 para la clase "Graduado", 0.87 para "Abandono" y 0.82 para "Matriculado". Con Optuna, los F1-Scores para "Abandono" y "Matriculado" aumentaron ligeramente a 0.88 y 0.84 respectivamente.
* **Abordaje del Desequilibrio de Clases:** La investigación resalta que el desequilibrio de clases es un problema común en la minería de datos educativos, donde el número de estudiantes que abandonan es significativamente menor que el de aquellos que permanecen o tienen un buen rendimiento académico. Este desequilibrio afecta negativamente la precisión de los modelos predictivos. El uso de SMOTE es crucial para mitigar este problema.
* **Relevancia para Intervenciones Académicas:** Predecir el comportamiento de los estudiantes es una tarea crucial para las instituciones educativas, ya que puede ayudar a mejorar el diseño del currículo y planificar intervenciones de apoyo académico oportunas y personalizadas. Las técnicas de Machine Learning pueden mejorar la retención estudiantil al permitir a los educadores reconocer las debilidades de los estudiantes y proporcionar métricas de aprendizaje en cualquier etapa del progreso educativo.
* **Guía para Investigadores y Profesionales:** El estudio proporciona información valiosa sobre la efectividad de varios algoritmos de Machine Learning para predecir el abandono estudiantil y el éxito académico. Al comparar estos algoritmos, el proyecto ofrece orientación a investigadores y profesionales en la elección de enfoques adecuados para tareas predictivas similares.
* **Limitaciones:** La generalización del estudio a otros contextos está limitada debido a su dependencia de un único conjunto de datos.

**Utilidad del artículo para el proyecto de tesis**

El articulo nos indica que se aplicaron varios algoritmos de Machine Learning con enfoque en el buen rendimiento académico. Se utilizó el método Isolation Forest para detectar datos atípicos y se integraron algoritmos de boosting, particularmente LightGBM y CatBoost, por su alta precisión. Los principales hallazgos permitieron personalizar el apoyo educativo a cada estudiante. La investigación no ayuda a utilizar algoritmos de aprendizaje supervisado, especialmente los de **boosting como LightGBM y CatBoost (optimizados con Optuna)**, para **predecir el rendimiento estudiantil** y el riesgo de abandono. Además, propone técnicas como **SMOTE para manejar datos desequilibrados** y **SHAP para interpretar las predicciones** del modelo, permitiendo al asistente identificar proactivamente a los estudiantes en riesgo y personalizar las intervenciones de aprendizaje en química.

**3.1.10.** **Visual-Predictive Data Analysis Approach for the Academic Performance of Students from a Peruvian University** [11]

**3.1.9.** **Enfoque de análisis de datos visual-predictivo para el rendimiento académico de estudiantes de una universidad peruana** [11]

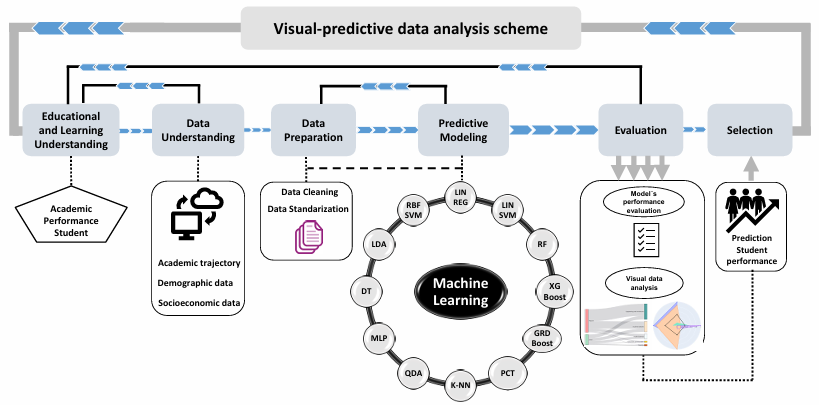
18 citaciones, fuente: Web Scopus

**DOI:**10.3390/app122111251

El artículo "Visual-Predictive Data Analysis Approach for the Academic Performance of Students from a Peruvian University" propone un enfoque integral para analizar y predecir el rendimiento académico de estudiantes universitarios peruanos. Reconoce que el éxito académico depende de múltiples factores y requiere de un análisis que combine métodos cuantitativos con otras formas de análisis.

**Metodologías:**

* **Enfoque Visual-Predictivo de Análisis de Datos:** El estudio emplea un enfoque novedoso que integra la comprensión del dominio, el análisis de visualización de datos y la construcción de modelos de Machine Learning.
* **Minería de Datos Educativos (EDM) y Machine Learning:** La investigación se basa en la Minería de Datos Educativos (EDM), una disciplina que aplica técnicas de minería de datos para analizar el contexto de los estudiantes y los factores relacionados con su rendimiento. El Machine Learning se utiliza para evaluar la predicción del rendimiento estudiantil en la educación superior.
* **Modelo de Machine Learning (XGBoost):** Se utilizó y entrenó un modelo de Machine Learning XGBoost para la predicción.
* **Análisis Visual de Datos:** Se combinaron los resultados del modelo de Machine Learning con un análisis visual de datos para identificar las variables relevantes asociadas al rendimiento académico de los estudiantes.
* **Identificación de Variables Relevantes:** El enfoque permitió identificar las variables clave que influyen en el éxito académico de los estudiantes.

****

*Esquema de análisis de datos visual-predictivo para el rendimiento académico de los estudiantes universitarios.*

**Importancias y Aspectos Más Resaltantes:**

* **Contexto Peruano:** El estudio se enfoca en una universidad peruana, lo que es altamente relevante dado el interés del país en mejorar su sistema educativo, que en 2019 ocupaba el puesto 127 de 137 en calidad educativa. Las universidades peruanas buscan activamente métodos para identificar a los estudiantes en riesgo académico y mejorar su rendimiento.
* **Precisión del Modelo Predictivo:** El modelo XGBoost entrenado alcanzó una precisión de hasta el 91.5%. Esta alta precisión es crucial para identificar eficazmente a los estudiantes con bajo rendimiento.
* **Herramienta para Políticas de Apoyo:** El enfoque visual-predictivo se considera una herramienta valiosa para desarrollar e implementar políticas dirigidas a apoyar a estudiantes con bajo rendimiento o estimular a aquellos con rendimiento avanzado. También proporciona información sobre la situación académica de las diferentes carreras universitarias.
* **Enfoque Integral:** La combinación de la comprensión del dominio, la visualización de datos y el Machine Learning permite un análisis integral de un problema complejo como el rendimiento académico multifactorial.

**Utilidad del artículo para el proyecto de tesis**

Por medio del articulo se verifica que el Machine Learning se utilizó como herramienta central para la predicción del rendimiento estudiantil. A través del desarrollo de un modelo de Machine Learning, fue posible identificar variables relevantes que influyen directamente en el desempeño académico, lo que permite personalizar estrategias de enseñanza según las necesidades de cada estudiante. El estudio sobre análisis visual-predictivo del rendimiento académico propone un enfoque útil para el asistente inteligente en Química Básica, combinando minería de datos educativos y algoritmos como Random Forest o Gradient Boosting. Su valor diferencial es integrar predicción y visualización, permitiendo mostrar gráficamente el progreso y las dificultades del estudiante. Esto requiere recopilar datos, entrenar modelos e interpretar resultados con SHAP, facilitando intervenciones tempranas y personalizadas por parte del asistente y del docente.

**3.1.11.** **Utilization of Artificial Intelligence and Machine Learning in Chemistry Education: A Critical Review** [12]

**3.1.11. Utilización de la inteligencia artificial y el aprendizaje automático en la educación en química: una revisión crítica** [12]

Citado por: 5 veces, fuente: Web of Science y Scopus

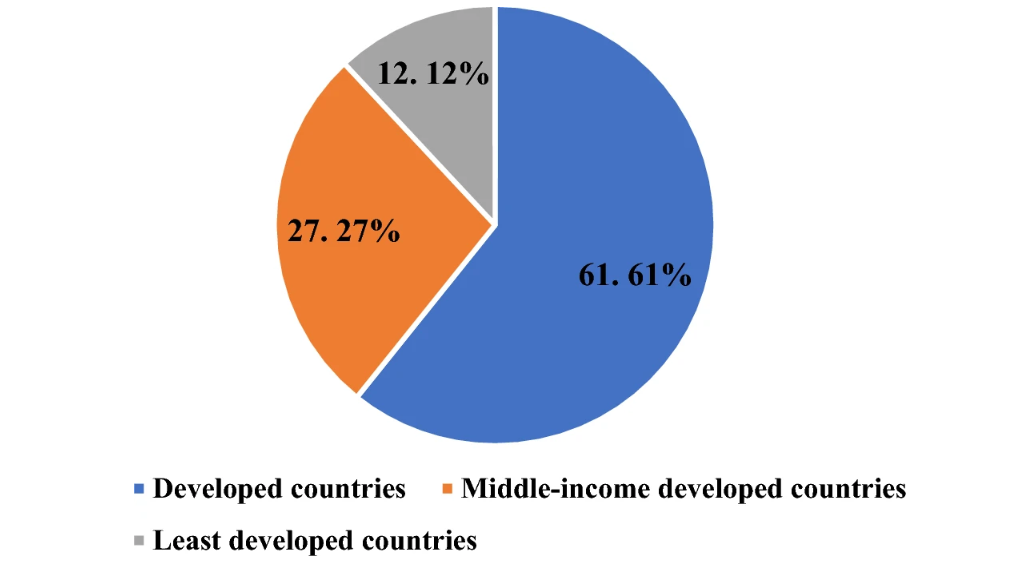
**DOI:**<https://doi.org/10.1007/s44217-024-00197-5>  
  
Este artículo ofrece una revisión integral sobre la aplicación de la inteligencia artificial (IA) y el aprendizaje automático (ML) en la enseñanza de la química, especialmente en contextos educativos. Reconoce la necesidad de métodos más personalizados y efectivos para mejorar la comprensión de conceptos químicos, dado que esta es una de las materias con mayor tasa de dificultad y abandono escolar.

El texto destaca cómo las tecnologías basadas en IA permiten adaptar el proceso de aprendizaje al ritmo y estilo individual del estudiante, mediante sistemas de tutoría inteligente, plataformas de aprendizaje adaptativo y análisis predictivo de rendimiento. Asimismo, enfatiza la transformación de los roles del docente y del estudiante en entornos donde el aprendizaje está mediado por tecnologías inteligentes.

La revisión también analiza estudios empíricos que muestran mejoras significativas en la motivación, participación y resultados de aprendizaje al aplicar herramientas de IA en la enseñanza de la química básica. Además, subraya la importancia de formar a los profesores para que sean capaces de integrar estas tecnologías de manera efectiva.

**Metodologías:**

**Revisión Sistemática:**

  
El estudio se desarrolló a partir de una revisión de literatura sistemática, abarcando más de 60 artículos indexados entre los años 2000 y 2020. Las fuentes principales incluyeron Scopus, Web of Science y Google Scholar. Se emplearon criterios de inclusión rigurosos para seleccionar artículos relacionados con la aplicación de IA y ML en la educación en química, excluyendo aquellos que no incluían resultados empíricos.

**Utilidad del artículo para el proyecto de tesis**

Por medio del articulo nos muestra que la inteligencia artificial y el aprendizaje automático permiten crear asistentes educativos con aprendizaje adaptativo, ajustando el contenido según el progreso del estudiante. Esto es clave en entornos donde el aprendizaje es desigual, como en escuelas públicas. El estudio sobre el uso de IA y Machine Learning en la enseñanza de la química es especialmente pertinente para el asistente inteligente en Química Básica. Integrar Sistemas como Tutores Inteligentes (ITS) para personalizar la instrucción, predecir el rendimiento y adaptar rutas de aprendizaje. También nos ayuda a implementar evaluaciones automatizadas y, a futuro, laboratorios virtuales con IA. Para su aplicación, es clave asegurar la calidad de los datos y considerar aspectos éticos y de privacidad en el diseño del sistema.

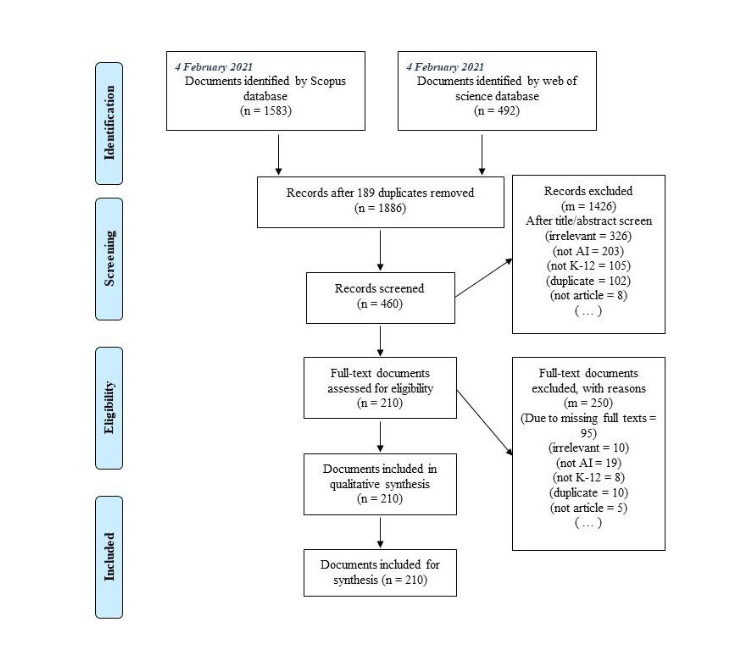
**3.1.12.** **Artificial Intelligence Applications in K-12 Education: A Review of Recent Trends** [NO\_PRINTED\_FORM] [13]  
**3.1.12.** **Aplicaciones de la inteligencia artificial en la educación K-12: una revisión de tendencias recientes** [NO\_PRINTED\_FORM] [13]  
Citado 1, fuente: Scopus

**DOI:** 10.1109/ACCESS.2022.3179356

Este artículo proporciona una revisión exhaustiva de cómo se está utilizando la inteligencia artificial (IA) en el ámbito de la educación básica (K-12), analizando las aplicaciones recientes y emergentes. Considera cómo estas tecnologías están siendo integradas en diferentes aspectos del proceso educativo, desde la enseñanza personalizada hasta el análisis predictivo del comportamiento estudiantil.

El documento destaca la rápida expansión de herramientas basadas en IA para apoyar tanto a docentes como a estudiantes. Entre los desarrollos más mencionados se encuentran los sistemas de tutoría inteligente, la analítica del aprendizaje, los agentes conversacionales y los asistentes virtuales. Estas herramientas no solo permiten una atención más personalizada, sino que también recopilan datos que ayudan a mejorar la toma de decisiones pedagógicas.

Se concluye que, a pesar del gran potencial de la IA en la educación, su integración en contextos escolares requiere un enfoque ético, formación docente y políticas claras que aseguren la equidad y privacidad de los datos.

  
  
  
*PRISMA flow diagram of the study*

**Metodología**

**Revisión de literatura estructurada:**

El estudio siguió un protocolo PRISMA para identificar, seleccionar y analizar los artículos publicados entre 2010 y 2020 sobre la aplicación de la IA en la educación K-12. Se consultaron bases de datos como Scopus, Web of Science y ERIC.

**Utilidad del artículo para el proyecto de tesis**

El articulo nos indica que actualmente, se viene analizando las aplicaciones más recientes y emergentes de la analítica del aprendizaje como una herramienta clave para brindar una atención más personalizada. Estas tecnologías permiten identificar patrones en el desempeño estudiantil y adaptar los contenidos y estrategias pedagógicas a las necesidades individuales, favoreciendo así un entorno educativo más efectivo y centrado en el estudiante. La investigación es altamente relevante para el desarrollo del asistente inteligente en Química Básica. Propone diseñarlo como un sistema de tutoría adaptativo que personalice el aprendizaje mediante Machine Learning, prediciendo el rendimiento, automatizando la evaluación y sugiriendo recursos a través de sistemas de recomendación. También nos ayuda a explorar la generación de contenido inteligente y el análisis de sentimientos. Para su implementación, es esencial considerar la privacidad de los menores, las implicaciones éticas y la capacitación docente en el uso pedagógico de esta tecnología en escuelas públicas peruanas.

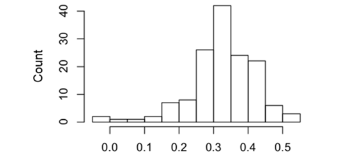
**3.1.13.** **Incorporating an Intelligent Tutoring System into the DiscoverOChem Learning Platform.** [14]  
  
**3.1.13.** **Incorporación de un sistema de tutoría inteligente en la plataforma de aprendizaje DiscoverOChem.** [14]

4 citaciones, fuente; Web Scopus

**DOI:** 10.1021/acs.jchemed.3c00117

En este artículo, se presenta un marco para desarrollar sistemas de tutoría inteligentes e incorporarlos a plataformas de aprendizaje en línea. Este enfoque fue aplicado para mejorar la plataforma DiscoverOChem y se describen los resultados preliminares obtenidos mediante técnicas de minería de datos. El análisis reveló correlaciones entre el rendimiento de los usuarios en páginas anteriores de la plataforma y su desempeño futuro. A partir de estos datos, se construyeron modelos predictivos, encontrando que los modelos de bosque aleatorio ofrecían los mejores resultados.

Con estos modelos se desarrolló un módulo de instrucciones que permite un sistema de inteligencia artificial tipo recomendado. Este sistema alerta a los usuarios sobre páginas que se espera sea difícil para ellos, y les proporciona enlaces a materiales de revisión recomendados, mejorando así su preparación para las siguientes secciones. Además, se creó una función de búsqueda basada en temas que sirve como base para este sistema de recomendaciones. Los autores sostienen que este marco es aplicable a otras plataformas de aprendizaje, especialmente en cursos con contenido acumulativo y respuestas definidas.



*Distribución de constante de correlon*

**Metodología:**

Para llevar a cabo esta investigación, se diseñó un marco estructurado destinado a desarrollar e integrar un sistema de tutoría inteligente dentro de una plataforma de aprendizaje en línea. El proceso comenzó con la recolección y análisis de datos de interacción de los usuarios con la plataforma DiscoverOChem, permitiendo aplicar técnicas de minería de datos para identificar patrones de comportamiento. A través de este análisis, se observaron correlaciones entre el rendimiento de los estudiantes en módulos anteriores y su desempeño en secciones futuras del curso.

Posteriormente, se construyeron modelos predictivos utilizando diversos algoritmos de aprendizaje automático, con el objetivo de anticipar el rendimiento de los usuarios en contenidos por venir. Entre los modelos evaluados, el algoritmo de bosque aleatorio (Random Forest) destacó por su capacidad predictiva y precisión, por lo cual fue seleccionado como base del sistema. A partir de estos modelos, se desarrolló un módulo de instrucciones inteligente que permite generar recomendaciones personalizadas para cada usuario.

Este sistema recomendado funciona identificando las secciones del curso que podrían representar una mayor dificultad para el estudiante, generando advertencias previas y proporcionando enlaces a materiales de repaso adaptados a sus necesidades. Para lograr esto, se implementó también un motor de búsqueda por temas dentro de la plataforma, facilitando el acceso a contenido específico y relevante según el perfil del usuario.

En conjunto, la metodología combina el análisis de datos educativos, técnicas de aprendizaje automático y diseño instruccional adaptativo, con el fin de mejorar la experiencia de aprendizaje y apoyar el desarrollo autónomo del estudiante dentro de un entorno digital.

**Importancia:**

El trabajo presentado en este documento permite a la plataforma de aprendizaje DiscoverOChem tomar decisiones significativas de tutoría inteligente para usuarios individuales. Antes de las mejoras actuales, la plataforma contenía algunos aspectos similares a los de ITS, pero no funcionaba realmente como un sistema de tutoría inteligente. Contenía un módulo estudiantil en forma de base de datos que mantiene registros de las páginas que se han cargado y las respuestas que ha enviado cada usuario. Las páginas de visualización podían presentar resúmenes de estos resultados al instructor, pero el sistema en sí no utilizaba este conocimiento para tomar decisiones que promovieran el aprendizaje individual de los estudiantes.

**Utilidad del artículo para el proyecto de tesis:**

Este artículo proporciona la implementación de un asistente inteligente basado en Machine Learning permite identificar correlaciones entre distintas variables educativas y el rendimiento de los usuarios, lo que facilita el diseño de modelos predictivos capaces de anticipar el rendimiento académico. Gracias a estos modelos, el sistema puede generar notificaciones o recomendaciones que alerta a los usuarios sobre posibles dificultades, promoviendo intervenciones oportunas para mejorar los resultados educativos. El estudio ofrece un modelo útil para el asistente inteligente en Química Básica. Propone adaptarlo al progreso del estudiante mediante la recopilación de datos de rendimiento y el uso de modelos predictivos como Random Forest para anticipar dificultades. Al analizar relaciones entre conceptos, el asistente podrá ajustar la dificultad, recomendar materiales y personalizar rutas de aprendizaje, mejorando así la comprensión y el desempeño en la asignatura.

**3.1.14.** **Developing a Curated Chatbot as an Exploratory Communication Tool for Chemistry Learning.** [15]

**3.1.14.** **Desarrollo de un chatbot curado como herramienta de comunicación exploratoria para el aprendizaje de la química.** [15]

Citado por: 16, Fuente: web ScopusDOI: 10.1021/acs.jchemed.3c00520

Este artículo apoya directamente el objetivo de la tesis al explorar cómo los chatbots basados en IA (ChatGPT y Bard) pueden asistir en el aprendizaje de la química orgánica. Evalúa su capacidad para interpretar estructuras químicas (como SMILES e InChI) y responder preguntas académicas, lo que respalda la idea de que estos asistentes pueden ser entrenados para apoyar el aprendizaje autónomo y personalizado en química básica. Aunque el artículo se enfoca en química orgánica, los fundamentos y metodología son aplicables a un curso de química básica en secundaria.

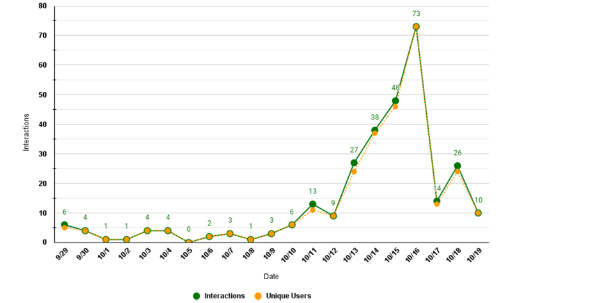


Diagrama de dispersión de las interacciones registradas del chatbot (verde) y de los usuarios únicos (naranja) durante los 20 días que el chatbot estuvo disponible para uso estudiantil. El último día mostrado fue la fecha límite para la tarea escrita que el chatbot debía proporcionar.

**Metodología:**

La investigación emplea un enfoque experimental comparativo:

* Se utilizaron dos chatbots (ChatGPT 3.5/4 y Google Bard).
* Se evaluó su precisión al resolver tareas relacionadas con la química orgánica como:
* Nomenclatura IUPAC.
* Identificación de grupos funcionales.
* Cálculo del grado de insaturación.
* Traducción entre formatos químicos (SMILES, InChi, estructuras condensadas).
* Diseño del estudio: Las mismas preguntas se aplicaron en dos rondas con distintos usuarios para evaluar reproducibilidad y exactitud.
* Los resultados fueron cuantificados en porcentajes de precisión por tipo de tarea.

**Utilidad del artículo para el proyecto de tesis:**

El articulo nos muestra que los chatbots impulsan el aprendizaje autónomo al ofrecer respuestas personalizadas sobre temas clave como Nomenclatura IUPAC y Cálculo del grado de insaturación, facilitando la comprensión de contenidos en Química Básica mediante interacciones adaptativas. La investigación sugiere un enfoque centrado en un chatbot interactivo como medio principal de comunicación. Para su implementación, se debe diseñar una lógica conversacional basada en árboles de decisión que guíe al estudiante por conceptos de Química Básica, curando contenido relevante desde una base de datos. El sistema debe desarrollarse como una aplicación web accesible y someterse a pruebas piloto, con el fin de recopilar datos de uso y ajustar su desempeño, optimizando así su impacto en el aprendizaje de estudiantes de secundaria peruanos.

****

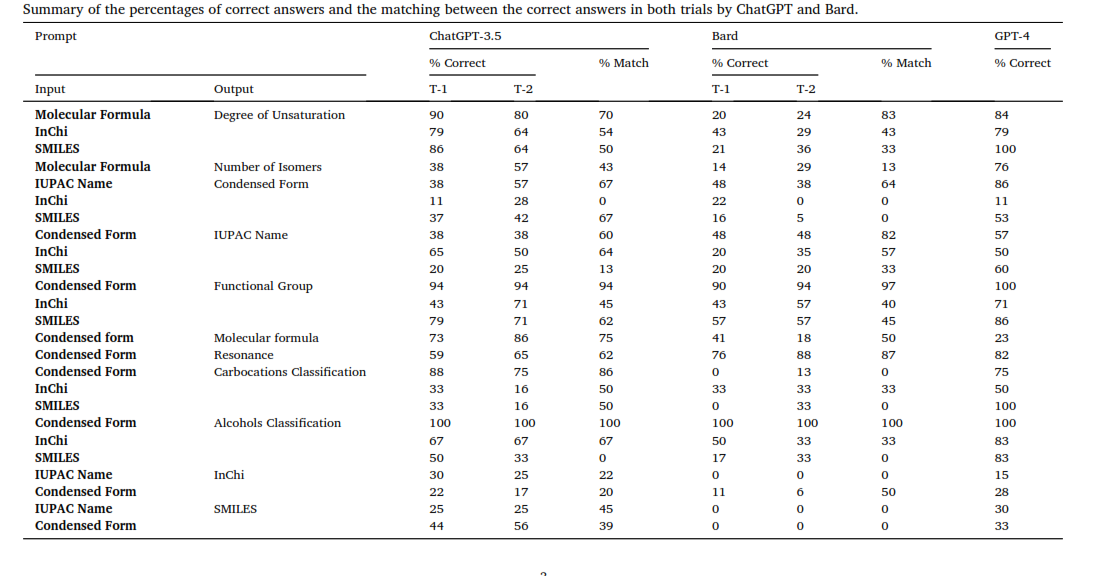
**3.1.15. Exploring the potential of AI-Chatbots in organic chemistry: An assessment of ChatGPT and Bard.** [1]

**3.1.15.** **Explorando el potencial de los chatbots de IA en la química orgánica: una evaluación de ChatGPT y Bard.** [1]

Citado 3, fuente web scopus

DOI: <https://doi.org/10.1016/j.caeai.2023.100170>

Este artículo se relaciona estrechamente con el objetivo del proyecto de tesis al evaluar el desempeño de chatbots basados en inteligencia artificial (ChatGPT y Bard) en la enseñanza de conceptos de química, particularmente orgánica. El estudio explora su capacidad para interpretar estructuras químicas en distintos formatos (como SMILES e InChi) y resolver preguntas relacionadas. A pesar de estar centrado en química orgánica y nivel universitario, sus conclusiones sobre el potencial educativo de estos asistentes, sus limitaciones, y su interacción con estudiantes son transferibles a un contexto de química básica en secundaria. Ofrece evidencia relevante para fundamentar el desarrollo e implementación de un asistente inteligente como herramienta pedagógica.



*Tabla 1GG*

**Metodologia:**

• Tipo de estudio: Experimental comparativo.

• Herramientas evaluadas: ChatGPT 3.5, ChatGPT 4 y Bard.

• Técnicas:

* Se usaron prompts químicos estandarizados sobre temas como estructuras condensadas, nomenclatura IUPAC, grupos funcionales, fórmulas moleculares, etc.
* Se realizaron pruebas con dos cuentas distintas para evaluar precisión y reproducibilidad de las respuestas.
* Las respuestas fueron cuantificadas en porcentaje de aciertos y coincidencias entre ensayos.
* Se incluyó una comparación entre versiones (3.5 vs 4) de ChatGPT.

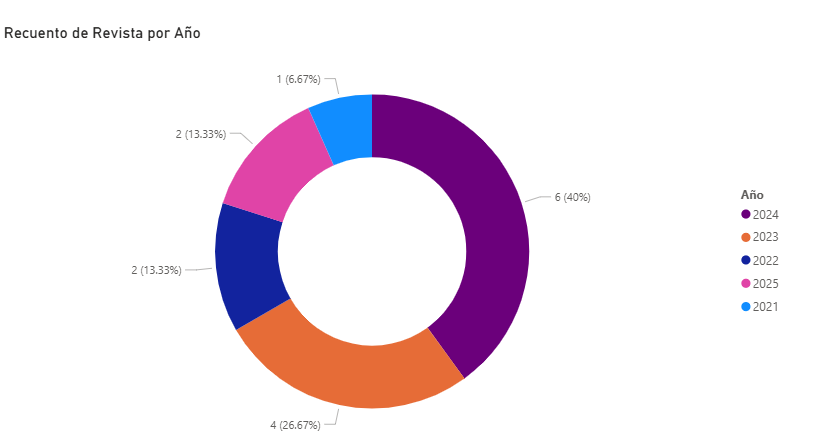
• Análisis: Se examinaron errores comunes y se discutieron implicancias educativas.

**Utilidad del artículo para el proyecto de tesis**

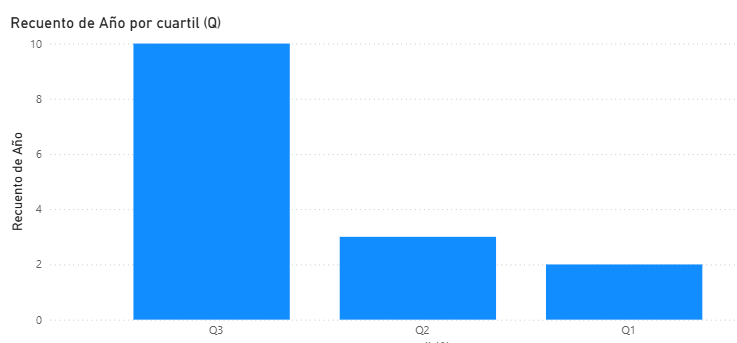
El articulo propone que inteligencia artificial ha demostrado un gran potencial educativo al integrarse en herramientas como ChatGPT, capaces de generar explicaciones claras y contextualizadas. Mediante el uso de prompts químicos estandarizados, es posible ofrecer respuestas precisas y coherentes que apoyen la comprensión de contenidos científicos de forma accesible y ajustada a las necesidades del usuario. Implementaremos IA-Chatbots basados en Grandes Modelos de Lenguaje (LLM) como componente central del asistente inteligente de Química Básica. Los estudiantes podrán formular preguntas en lenguaje natural y hallar respuestas claras, incluso sobre contenidos complejos como la notación química. Hallaremos posibles imprecisiones o desajustes con el currículo peruano mediante revisión experta y ajustes iterativos, asegurando que el chatbot funcione como una herramienta educativa confiable y contextualizada.

Preguntar a ChatGPT

* **Gráficos de las revistas revisadas**



La imagen muestra el porcentaje de años revisadas por articulo



La imagen muestra la cantidad de cuartiles revisadas en los artículos

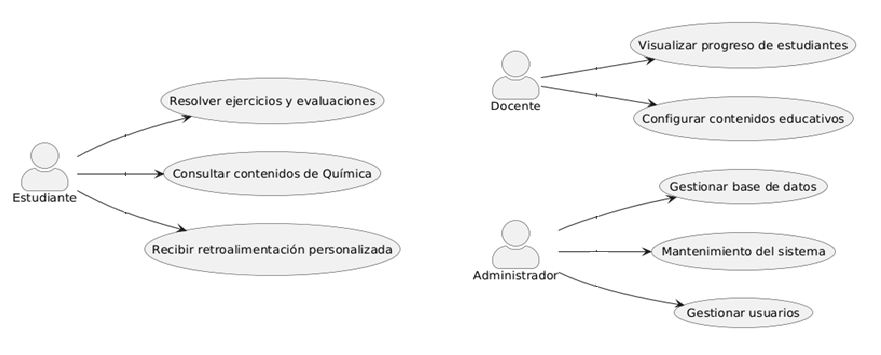
CAPÍTULO IV: MODELADO DEL NEGOCIO

**4.1 Reglas del Negocio**

|  |  |
| --- | --- |
| **Código** | **Regla del Negocio** |
| RN1 | El asistente inteligente estará disponible 24/7 para todos los estudiantes registrados. |
| RN2 | Los contenidos del sistema estarán alineados con el currículo nacional de Química Básica del MINEDU. |
| RN3 | El sistema ajustará la dificultad de los ejercicios en función del rendimiento del estudiante usando algoritmos de aprendizaje automático. |
| RN4 | Cada sesión de aprendizaje generará un reporte automático para el docente asignado. |
| RN5 | Los docentes tendrán acceso a un panel para personalizar contenidos, revisar el avance y brindar refuerzo individual. |
| RN6 | El sistema almacenará las interacciones del estudiante para análisis de rendimiento y mejora del modelo de recomendación. |
| RN7 | Los estudiantes recibirán retroalimentación inmediata después de cada interacción, incluyendo explicaciones conceptuales. |

## **4.2 Casos de uso**

4.2.1 Diagrama de casos de uso del negocio



**4.2.1.1 Actores del negocio**

**Roles de Usuarios en el Asistente Inteligente de Química Básica**

El sistema del asistente inteligente de Química Básica interactuará con los siguientes roles de usuario clave, cada uno con responsabilidades y funcionalidades específicas:

* **Estudiante:** Es el usuario final y principal beneficiario del asistente. Su objetivo es reforzar y profundizar sus conocimientos en Química Básica a través de la interacción personalizada con los contenidos y ejercicios adaptativos proporcionados por el sistema.
* **Docente:** Actúa como supervisor y facilitador del proceso de aprendizaje. Monitorea el progreso individual y grupal de los estudiantes a través de reportes detallados y puede configurar parámetros específicos del asistente o reforzar contenidos de Química Básica en función de las necesidades detectadas.
* **Administrador:** responsable del mantenimiento general del sistema. Esto incluye la gestión de usuarios (altas, bajas, modificaciones), la supervisión de la base de datos, la gestión de contenidos a nivel global y el aseguramiento del correcto funcionamiento técnico de la plataforma.

**4.2.1.2 Casos de Uso del Negocio (CUN)**

**CUN** **- Iniciar sesión**

Actor Principal: Estudiante

Precondición: Estudiante registrado.

Flujo principal:

* El estudiante accede a la plataforma.
* Introduce usuario y contraseña.
* El sistema valida credenciales y concede acceso.

**CUN - Realizar diagnóstico inicial**

Actor Principal: Estudiante

Precondición: Sesión iniciada correctamente.

Flujo principal:

* El estudiante accede al módulo de diagnóstico.
* El sistema presenta preguntas iniciales.
* El sistema evalúa respuestas y determina el nivel del estudiante.

**CUN - Resolver ejercicios personalizados**

Actor Principal: Estudiante

Precondición: Nivel identificado.

Flujo principal:

* El asistente genera ejercicios adaptativos.
* El estudiante responde.
* El sistema ofrece retroalimentación inmediata.

**CUN - Revisar progreso de estudiantes**

Actor Principal: Docente

Precondición: Docente autorizado.

Flujo principal:

* El docente accede al panel de control.
* Consulta reportes de estudiantes individuales o por grupo.

**CUN - Generar retroalimentación automática**

Actor Principal: Asistente IA

Precondición: Ejercicio resuelto.

Flujo principal:

* El sistema analiza la respuesta del estudiante.
* Emite feedback adaptado al tipo de error o acierto.

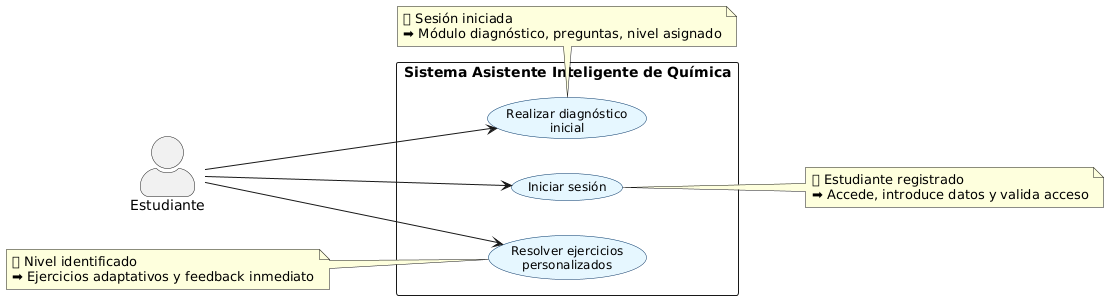
**CUN- Registrar estudiante**

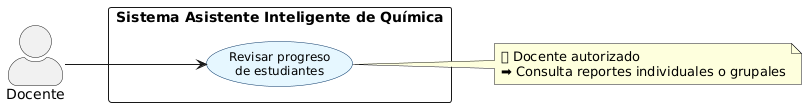
Actor Principal: Administrador

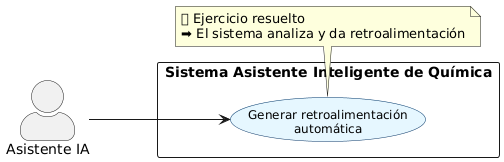
Precondición: Datos del estudiante validados.

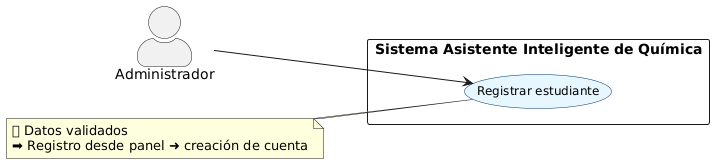
Flujo principal:

* El administrador accede al panel de registro.
* Introduce los datos del estudiante.
* El sistema crea la cuenta correspondiente.









CAPÍTULO V: REQUERIMIENTOS DEL PROYECTO

**5.1 Requerimientos del Software:**

Los requerimientos del software del asistente inteligente para el aprendizaje de Química Básica se dividen en dos categorías: requerimientos funcionales y no funcionales, los cuales aseguran el correcto desempeño del sistema dentro del entorno educativo de las escuelas públicas peruanas.

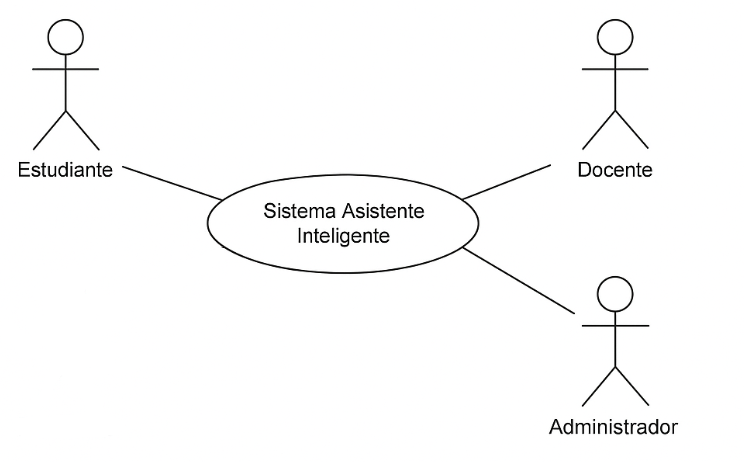
###### **5.1.1 Requerimientos Funcionales**

|  |  |  |
| --- | --- | --- |
| **Código** | **Requerimiento funcional** | **Descripción** |
| RF-01 | Registro de usuarios | El sistema deberá permitir que los estudiantes y docentes se registren con sus datos personales e institucionales. |
| RF-02 | Inicio de sesión | El sistema deberá permitir que los usuarios accedan mediante credenciales válidas. |
| RF-03 | Diagnóstico de nivel | El asistente deberá evaluar el conocimiento inicial del estudiante mediante una prueba diagnóstica. |
| RF-04 | Recomendación de contenidos | El sistema deberá generar sugerencias de material educativo personalizado en función del nivel del estudiante. |
| RF-05 | Retroalimentación automatizada | El asistente deberá proporcionar comentarios automáticos sobre el desempeño del estudiante. |
| RF-06 | Visualización de progreso | El estudiante podrá ver su avance a lo largo del curso mediante gráficos o tablas. |
| RF-07 | Gestión docente | El docente podrá acceder a reportes individuales y grupales del rendimiento de sus alumnos. |

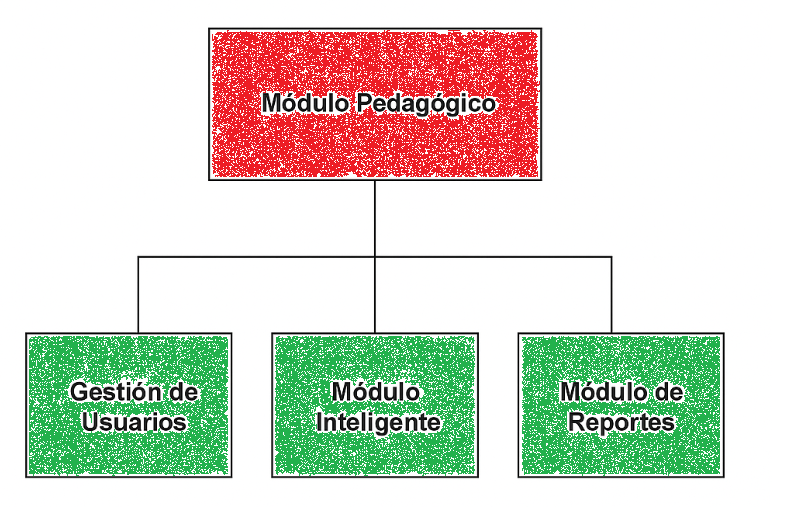
### **5.1.2 Requerimientos No Funcionales**

|  |  |  |
| --- | --- | --- |
| **Código** | **Requerimiento no funcional** | **Descripción** |
| RNF-01 | Usabilidad | El sistema deberá contar con una interfaz intuitiva y accesible para estudiantes de secundaria. |
| RNF-02 | Rendimiento | El asistente deberá procesar solicitudes y mostrar respuestas en menos de 3 segundos. |
| RNF-03 | Portabilidad | El sistema deberá ser accesible desde navegadores web modernos y dispositivos móviles. |
| RNF-04 | Seguridad | Las credenciales de usuario deben cifrarse y almacenarse de manera segura. |
| RNF-05 | Mantenibilidad | El código deberá seguir buenas prácticas de desarrollo que permitan su actualización. |
| RNF-06 | Escalabilidad | El sistema deberá soportar un número creciente de usuarios sin degradación del servicio. |
| RNF-07 | Accesibilidad | Deberá cumplir con criterios de accesibilidad para estudiantes con discapacidades visuales o auditivas. |

**5.2 Casos de Uso del Sistema**

****

**5.2.2 Diagrama de Paquetes**



**5.3. Especificación de los Actores del Sistema**

|  |  |
| --- | --- |
| **Actor** | **Descripción** |
| **Estudiante** | Usuario principal del sistema. Interactúa con el asistente inteligente para acceder a contenidos educativos, resolver ejercicios y recibir retroalimentación. El sistema registra su progreso para personalizar el aprendizaje. |
| **Docente** | Responsable de configurar criterios pedagógicos, monitorear el desempeño de los estudiantes y validar contenidos sugeridos por el asistente inteligente. Su rol es fundamental en la supervisión del proceso educativo. |
| **Administrador** | Encargado del mantenimiento técnico del sistema. Gestiona usuarios, asegura el correcto funcionamiento de la plataforma y realiza tareas de configuración avanzada. |
| **Sistema Asistente Inteligente** | |  | | --- | |  |  |  | | --- | | Módulo automatizado que aplica algoritmos de Machine Learning para analizar el rendimiento estudiantil, identificar brechas de aprendizaje y recomendar rutas de estudio personalizadas. Interactúa con estudiantes y docentes mediante interfaces adaptadas. | |

**5.3.1 Especificaciones de CUS**CUS01 - Interacción con el Asistente Inteligente

|  |  |
| --- | --- |
| **Campo** | **Descripción** |
| **Código** | CUS01 |
| **Nombre** | Interacción con el Asistente Inteligente |
| **Descripción** | El estudiante accede al sistema, recibe contenidos personalizados, realiza ejercicios y obtiene retroalimentación inmediata. |
| **Actor Principal** | |  | | --- | |  |  |  | | --- | | Estudiante | |
| **Actores Secundarios** | |  | | --- | |  |  |  | | --- | | Sistema Asistente Inteligente | |
| **Precondiciones** | - El estudiante debe estar registrado y autenticado en el sistema. - Dispositivo con acceso a Internet. |
| **Flujo Principal** | • El estudiante inicia sesión. • Accede a su ruta de aprendizaje personalizada. • Interactúa con el contenido. • Recibe retroalimentación inmediata. • El sistema registra el desempeño. |
| **Postcondiciones** | - El avance queda registrado en la base de datos. - Se actualiza la ruta de aprendizaje según el rendimiento. |

CUS02 - Monitoreo del Progreso Académico

|  |  |
| --- | --- |
| **Campo** | **Descripción** |
| **Código** | CUS02 |
| **Nombre** | Monitoreo del Progreso Académico |
| **Descripción** | El docente visualiza reportes individuales y grupales del desempeño de los estudiantes. |
| **Actor Principal** | Docente |
| **Actores Secundarios** | Sistema Asistente Inteligente |
| **Precondiciones** | - El docente debe estar registrado y tener permisos de acceso. - Datos del estudiante almacenados previamente. |
| **Flujo Principal** | • El docente inicia sesión. • Selecciona grupo o estudiante. • Visualiza reporte. • Toma decisiones pedagógicas basadas en el rendimiento. |
| **Postcondiciones** | - Información consultada queda registrada en el historial de monitoreo. |

**BIBLIOGRAFÍA**

[1] K. Hallal, R. Hamdan, and S. Tlais, “Exploring the potential of AI-Chatbots in organic chemistry: An assessment of ChatGPT and Bard,” *Computers and Education: Artificial Intelligence*, vol. 5, Jan. 2023, doi: 10.1016/j.caeai.2023.100170.

[2] G. Chionas and A. Emvalotis, “How peruvian secondary students view scientists and their works: Ready, set, and draw!,” *International Journal of Education in Mathematics, Science and Technology*, vol. 9, no. 1, pp. 116–137, 2021, doi: 10.46328/ijemst.1099.

[3] L. A. Holgado-Apaza *et al.*, “The Exploration of Predictors for Peruvian Teachers’ Life Satisfaction through an Ensemble of Feature Selection Methods and Machine Learning,” *Sustainability*, vol. 16, no. 17, p. 7532, Aug. 2024, doi: 10.3390/su16177532.

[4] L. A. Holgado-Apaza *et al.*, “Modeling Job Satisfaction of Peruvian Basic Education Teachers Using Machine Learning Techniques,” *Applied Sciences (Switzerland)*, vol. 13, no. 6, Mar. 2023, doi: 10.3390/app13063945.

[5] A. Valencia-Arias, J. A. Uribe-Gómez, E. Flores-Siapo, L. Palacios-Moya, A. Gallegos, and E. M. Rojas, “Application of Artificial Neural Networks to Predict the Use of Mobile Learning by University Students,” *Hum Behav Emerg Technol*, vol. 2024, 2024, doi: 10.1155/hbe2/1518987.

[6] M. A. Valles-Coral *et al.*, “Density-Based Unsupervised Learning Algorithm to Categorize College Students into Dropout Risk Levels,” *Data (Basel)*, vol. 7, no. 11, Nov. 2022, doi: 10.3390/data7110165.

[7] N. S. Koti Mani Kumar Tirumanadham, T. Sekhar, and S. Muthal, “An analysis of diverse computational models for predicting student achievement on e-learning platforms using machine learning,” *International Journal of Electrical and Computer Engineering*, vol. 14, no. 6, pp. 7013–7021, Dec. 2024, doi: 10.11591/ijece.v14i6.pp7013-7021.

[8] M. P. Silva Payró, R. Mena de la Rosa, and R. Cruz Romero, “Artificial Intelligence and virtual assistants: use and impact on learning and project development of undergraduate and graduate students in a faculty in southeastern Mexico,” *European Public and Social Innovation Review*, vol. 10, Jan. 2025, doi: 10.31637/epsir-2025-1378.

[9] L. Vives *et al.*, “Prediction of Students’ Academic Performance in the Programming Fundamentals Course Using Long Short-Term Memory Neural Networks,” *IEEE Access*, vol. 12, pp. 5882–5898, 2024, doi: 10.1109/ACCESS.2024.3350169.

[10] A. Villar and C. R. V. de Andrade, “Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study,” *Discover Artificial Intelligence*, vol. 4, no. 1, Dec. 2024, doi: 10.1007/s44163-023-00079-z.

[11] D. Orrego Granados, J. Ugalde, R. Salas, R. Torres, and J. L. López-Gonzales, “Visual-Predictive Data Analysis Approach for the Academic Performance of Students from a Peruvian University,” *Applied Sciences (Switzerland)*, vol. 12, no. 21, Nov. 2022, doi: 10.3390/app122111251.

[12] A. Iyamuremye *et al.*, “Utilization of artificial intelligence and machine learning in chemistry education: a critical review,” *Discover Education*, vol. 3, no. 1, Jul. 2024, doi: 10.1007/s44217-024-00197-5.

[13] G. Biagini, “Towards an AI-Literate Future: A Systematic Literature Review Exploring Education, Ethics, and Applications,” 2025, *Springer*. doi: 10.1007/s40593-025-00466-w.

[14] C. E. Jakobsche, P. Kongsomjit, C. Milson, W. Wang, and C. K. Ngan, “Incorporating an Intelligent Tutoring System into the DiscoverOChem Learning Platform,” *J Chem Educ*, vol. 100, no. 8, pp. 3081–3088, Aug. 2023, doi: 10.1021/acs.jchemed.3c00117.

[15] A. T. Lolinco and T. A. Holme, “Developing a Curated Chatbot as an Exploratory Communication Tool for Chemistry Learning,” *J Chem Educ*, vol. 100, no. 10, pp. 4092–4098, Oct. 2023, doi: 10.1021/acs.jchemed.3c00520.