 **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

INDICE

[2](#_Toc197633620)

[DEDICATORIA 2](file:///C:\Users\Admin\Downloads\seman%205%20(2).docx#_Toc197633621)

[AGRADECIMIENTO 3](file:///C:\Users\Admin\Downloads\seman%205%20(2).docx#_Toc197633622)

[INTRODUCCION 7](#_Toc197633623)

[CAPÍTULO I: VISIÓN DEL PROYECTO 8](#_Toc197633624)

[1.1 Antecedentes del Problema 1.1.1 El Negocio 8](#_Toc197633625)

[Figura 1.1. Porcentaje de estudiantes con logro en Ciencia y Tecnología Fuente (MINEDU, 2024) 9](#_Toc197633626)

[1.1.2 Proceso del Negocio 10](#_Toc197633627)

[**1.2.2 Descripción del problema** 1.2.2.1 Problema principal 13](#_Toc197633628)

[1.3. Objetivos del Proyecto 14](#_Toc197633629)

[1.3.1. Marco Lógico 14](#_Toc197633630)

[Figura 1 14](#_Toc197633631)

[Figura 2 15](#_Toc197633632)

[1.3.2. Objetivo general 16](#_Toc197633633)

[1.3.3.Objetivos específico 17](#_Toc197633634)

[1.4 Justificación del Proyecto 17](#_Toc197633635)

[1.4.1 Justificación Académica 17](#_Toc197633636)

[1.4.2 Beneficios Tangibles 17](#_Toc197633637)

[1.4.3 Beneficios Intangibles 18](#_Toc197633638)

[1.5 Alcance del Proyecto 18](#_Toc197633639)

[CAPÍTULO III: ESTADO DEL ARTE   
 3.1 Artículos 19](#_Toc197633640)

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.

## 

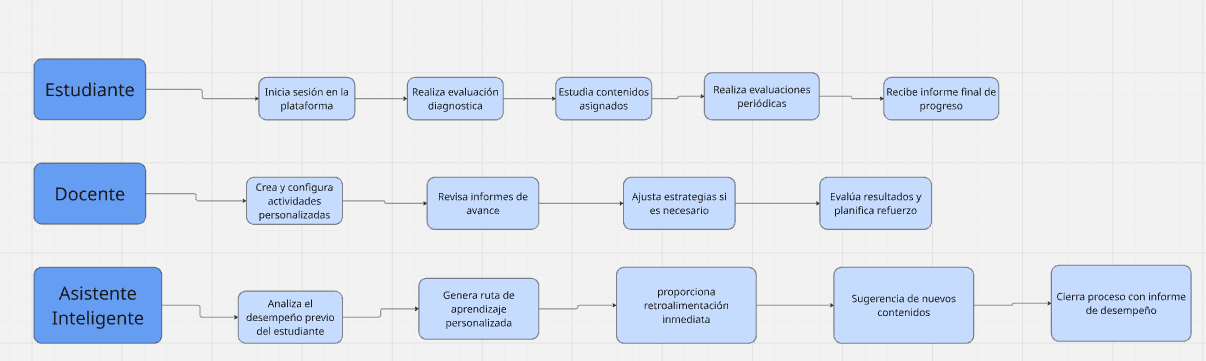
# CAPÍTULO I: VISIÓN DEL PROYECTO

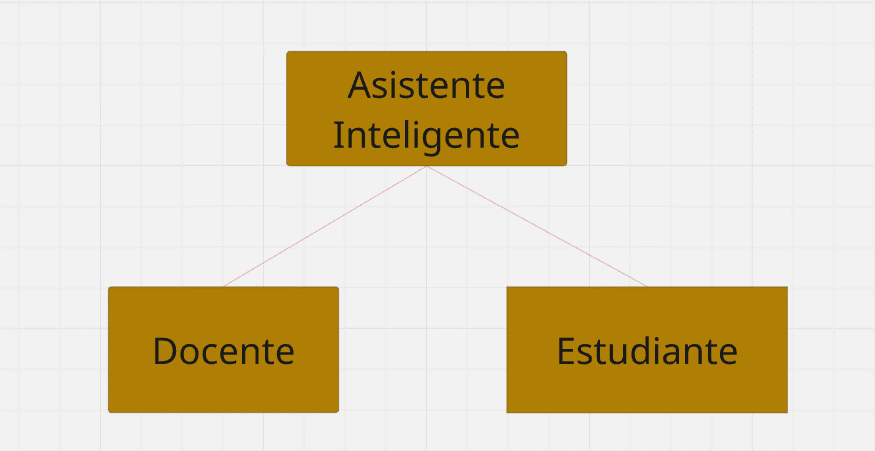
# 1.1 Antecedentes del Problema 1.1.1 El Negocio

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  
  
  
  
  
  
  
 Figura 1.2 Proceso de interacción  
 Fuente: (Elaboración propia 2025 )

1.1.3 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)

## **1.3.3.Objetivos específico**

* 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** (Chionas & Emvalotis, 2021)

**Cómo los estudiantes peruanos de secundaria ven a los científicos y sus trabajos: Listos, preparados y listos** (Chionas & Emvalotis, 2021)

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

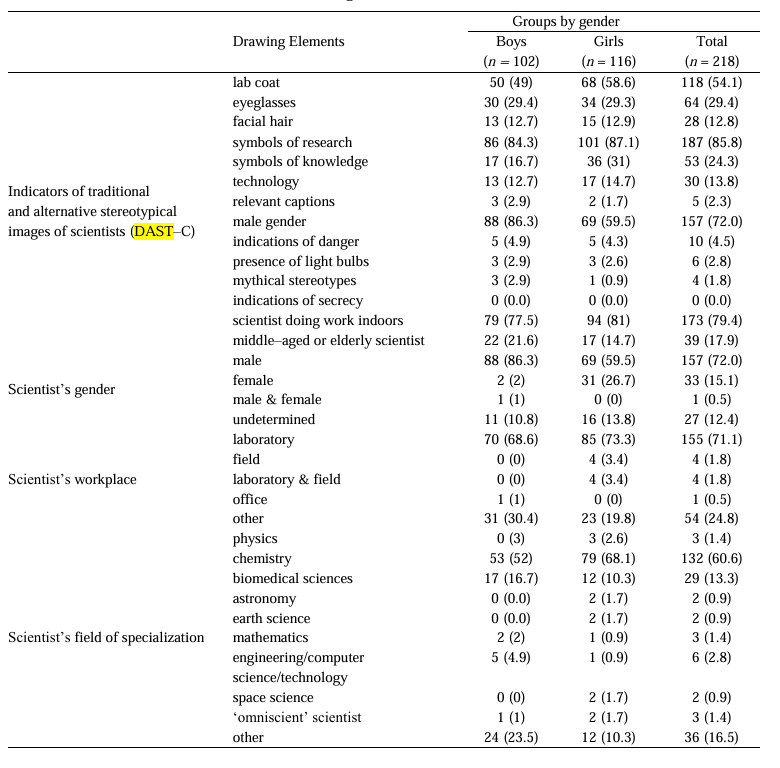
4 citaciones, fuente: Web Scopus

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. El artículo proporciona una comprensión valiosa sobre las percepciones y estereotipos que tienen los estudiantes peruanos de secundaria acerca de los científicos y su trabajo. Esta información resulta útil para el diseño de un asistente inteligente basado en Machine Learning que, además de facilitar el aprendizaje de Química Básica, pueda abordar y modificar dichas ideas preconcebidas, promoviendo una imagen más realista, inclusiva y motivadora de la ciencia. De esta forma, se busca incrementar el interés, la motivación y la participación activa de los estudiantes en el curso.

.

**3.1.2.** **The Exploration of Predictors for Peruvian Teachers’ Life Satisfaction through an Ensemble of Feature Selection Methods and Machine Learning** (Holgado-Apaza et al., 2024)

**Exploración de predictores de la satisfacción vital de docentes peruanos mediante un conjunto de métodos de selección de características y aprendizaje automático** (Holgado-Apaza et al., 2024)

**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:

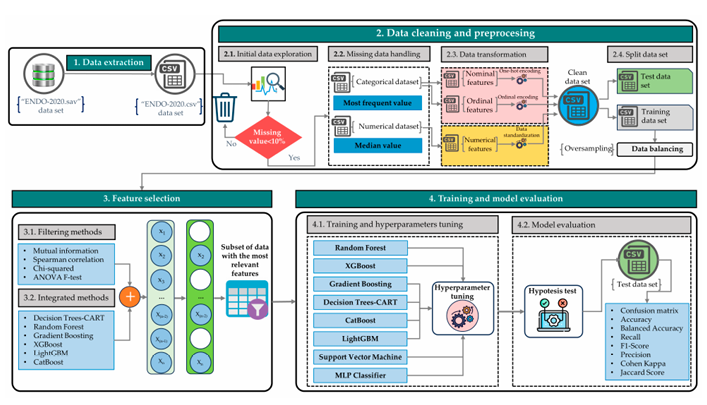
Recopilación de datos: Se inició con la obtención de datos a través de encuestas dirigidas a una muestra representativa de docentes de diversas regiones de Perú. Estas encuestas habrían incluido una amplia gama de preguntas relacionadas con su vida profesional y personal, percepción de su bienestar, 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**

Por medio de este artículo de investigación se tomarán en cuenta que 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 del estudiante. 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. Se utilizan técnicas de inteligencia artificial y machine learning para identificar los factores que más influyen en la satisfacción laboral de los profesores en Perú. Mediante el análisis de datos reales y diferentes algoritmos, se determinan las variables más relevantes. Esto puede servir para desarrollar asistentes inteligentes que, analizando información, ayuden a mejorar el aprendizaje de los estudiantes en Química Básica, haciéndolo más personalizado y efectivo. En general, estas herramientas muestran cómo la inteligencia artificial puede apoyar en decisiones para mejorar la educación.

**3.1.3.** **Modeling Job Satisfaction of Peruvian Basic Education Teachers Using Machine Learning Techniques** (Holgado-Apaza et al., 2023)

**Modelado de la satisfacción laboral de docentes de educación básica peruana mediante técnicas de aprendizaje automático** (Holgado-Apaza et al., 2023)

5 citaciones, fuente: Web Scopus

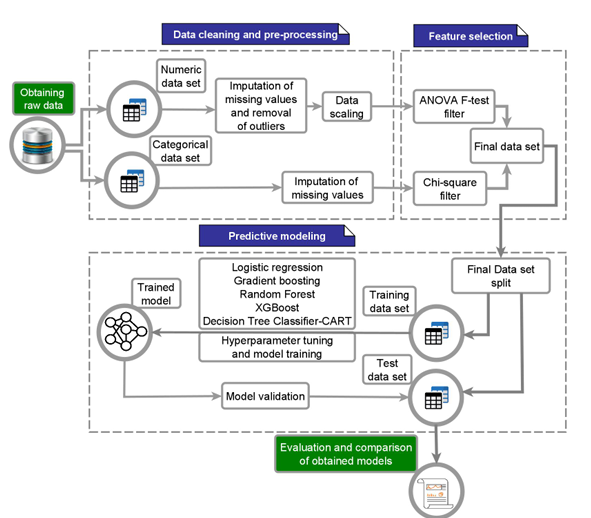
**DOI: https://doi.org/10.3390/app13063945**

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%.

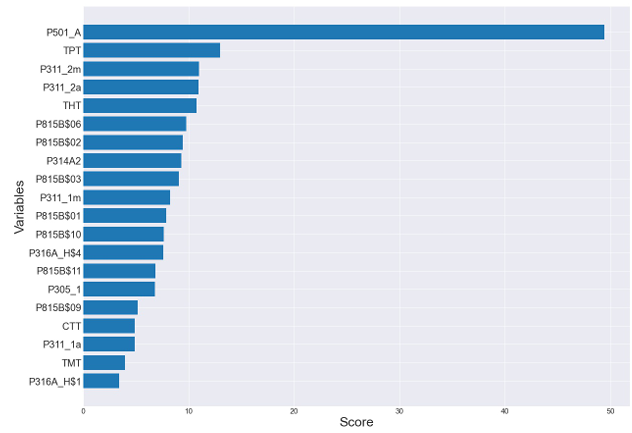
**Metodologías utilizadas:**

* **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 se puede predecir la satisfacción laboral de los docentes de educación básica en Perú mediante técnicas de aprendizaje automático. Mediante 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. Al igual que en el estudio, el proyecto puede utilizar técnicas de selección de variables para identificar factores que más influyen en el interés, motivación y rendimiento de los estudiantes en Química, como su nivel de satisfacción, motivación, dificultad percibida, o interés por los temas.

**3.1.4. Application of Artificial Neural Networks to Predict the Use of Mobile Learning by University Student** (Valencia-Arias et al., 2024)

**Aplicación de redes neuronales artificiales para predecir el uso del aprendizaje móvil por parte de estudiantes universitarios** (Valencia-Arias et al., 2024)

6 citaciones, fuente: Web Scopus

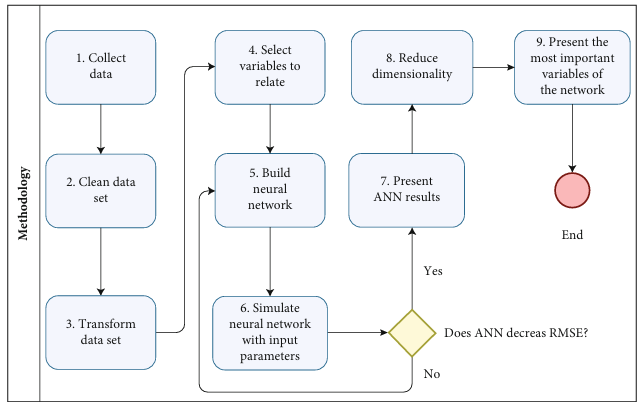
**DOI:** [**https://doi.org/10.1155/hbe2/1518987**](https://doi.org/10.1155/hbe2/1518987)

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.

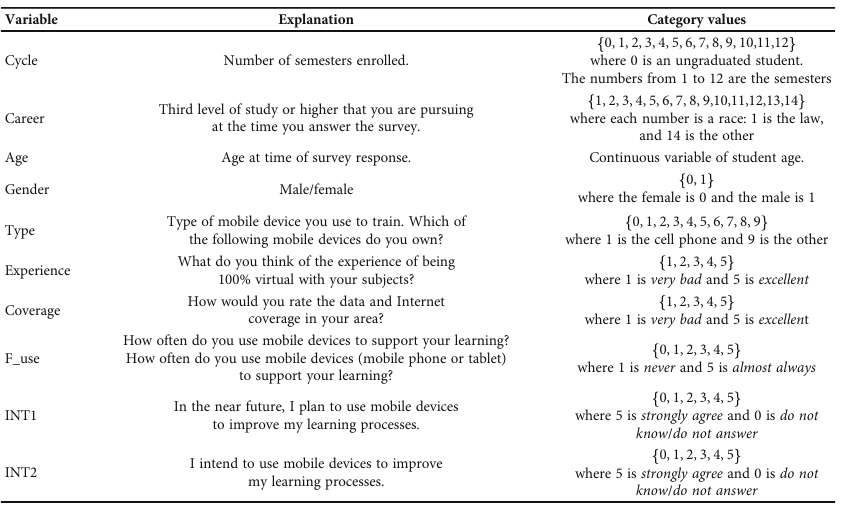
**Metodología**

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 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. Asimismo, la investigación actual resalta la importancia de la experiencia del usuario y la frecuencia de uso en la intención de emplear tecnologías educativas. Esto proporciona una base sólida para considerar estas variables en el diseño y la evaluación que permitirá desarrollar un modelo robusto para evaluar la efectividad del asistente inteligente en la mejora del aprendizaje en Química Básica y, potencialmente, predecir el rendimiento de los estudiantes basándose en sus interacciones con el sistema.

**3.1.5. Density-Based Unsupervised Learning Algorithm to Categorize College Students into Dropout Risk Levels** (Valles-Coral et al., 2022)

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

(Valles-Coral et al., 2022)

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.

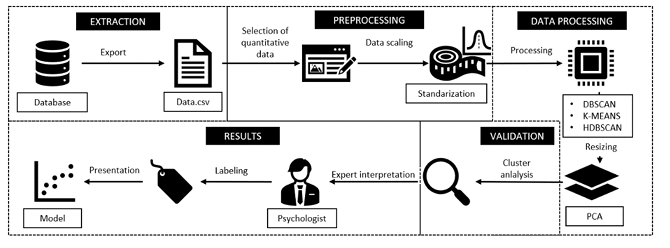
**Preprocesamiento de Datos Riguroso:** Para asegurar la calidad y la efectividad del modelo, se llevaron a cabo pasos de preprocesamiento esenciales:

**Importancia y Resultados Resaltantes:**

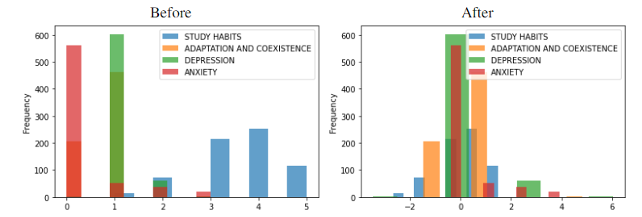
La principal **importancia** de este artículo radica en su contribución a la **predicción y prevención de la deserción universitaria** mediante un enfoque innovador de **aprendizaje no supervisado**. Esto es fundamental porque no requiere etiquetas de "abandono" previas (es decir, no necesita saber de 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 no demuestra como 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. 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.De igual forma, valida el uso de Machine Learning en un contexto educativo peruano, proporcionando un modelo de cómo recopilar y estructurar datos de estudiantes relevantes, sugiriendo la utilidad de algoritmos no supervisados para identificar patrones de aprendizaje (o dificultades) sin etiquetas previas, y fortaleciendo la justificación de la relevancia social de tu proyecto al enmarcarlo dentro de esfuerzos similares para mejorar la educación en el país.

**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.** (Koti Mani Kumar Tirumanadham et al., 2024)

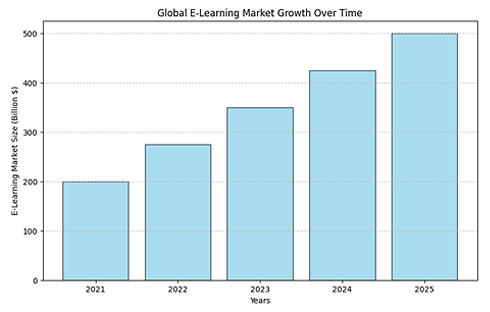
5 citaciones, fuente: Web Scopus

**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**

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, el estudio enfatiza que estas técnicas, especialmente la selección de características híbridas y el manejo de datos. En síntesis, este artículo funciona como una guía práctica, respaldada por la experimentación, que detalla la implementación de elementos esenciales de Machine Learning. Te equipa con las metodologías necesarias y te brinda la seguridad de que sus hallazgos son aplicables para optimizar el aprendizaje de la Química Básica dentro de tu propio entorno educativo.

**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**(Silva Payró et al., 2025)

**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)

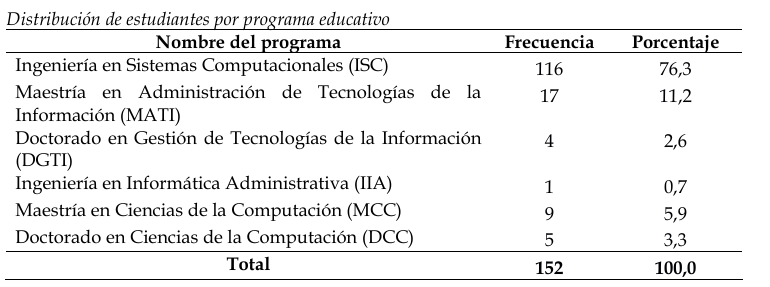
5 citaciones, fuente: Web Scopus

**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 estudio se llevó a cabo con una muestra de 152 estudiantes de Tecnologías. A pesar de que los estudiantes no están muy familiarizados con chatbots o bocinas inteligentes para el aprendizaje autodirigido, el estudio concluye que, aunque la adopción actual es limitada, existe un potencial. El articulo proporciona un valioso entendimiento del "lado humano" de la implementación de la IA en la educación. justifica la viabilidad de un asistente inteligente desde la perspectiva de la aceptación estudiantil y la relevancia educativa, y brinda consideraciones importantes para el diseño e implementación en el contexto de escuelas públicas peruanas.

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

**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**

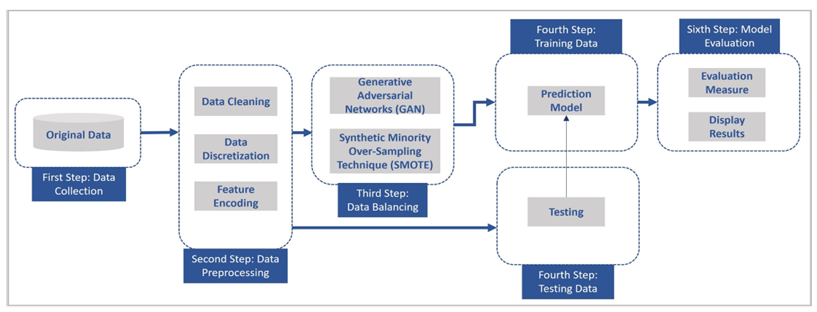
16 citaciones, fuente: Web Scopus

**DOI:**10.1109/ACCESS.2024.3350169

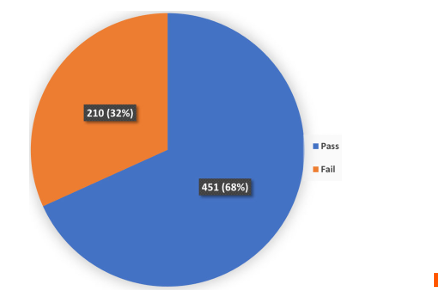
El artículo "Prediction of Students' Academic Performance in the Programming Fundamentals Course Using Long Short-Term Memory Neural Networks" se enfoca en la predicción del rendimiento académico de estudiantes universitarios peruanos en el curso de "Fundamentos de Programación" para identificar a aquellos 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**

La investigación se basa en la Minería de Datos Educativos, técnicas de Machine Learning, Recolección de Datos, Balanceo de Datos, Entrenamiento de Datos, Pruebas de Datos y Evaluación del Modelo. Este artículo constituye un pilar metodológico y empírico para la investigación, ya que proporciona una guía clara sobre cómo abordar la predicción del rendimiento académico mediante técnicas de Machine Learning en un contexto educativo similar. Ofrece herramientas avanzadas para el preprocesamiento de datos y modelos predictivos de alto rendimiento, elementos esenciales para el éxito del asistente inteligente en el curso de Química Básica.

**3.1.9.**

**BIBLIOGRAFÍA**

Chionas, G., & Emvalotis, A. (2021). How peruvian secondary students view scientists and their works: Ready, set, and draw! *International Journal of Education in Mathematics, Science and Technology*, *9*(1), 116–137. https://doi.org/10.46328/ijemst.1099

Holgado-Apaza, L. A., Carpio-Vargas, E. E., Calderon-Vilca, H. D., Maquera-Ramirez, J., Ulloa-Gallardo, N. J., Acosta-Navarrete, M. S., Barrón-Adame, J. M., Quispe-Layme, M., Hidalgo-Pozzi, R., & Valles-Coral, M. (2023). Modeling Job Satisfaction of Peruvian Basic Education Teachers Using Machine Learning Techniques. *Applied Sciences (Switzerland)*, *13*(6). https://doi.org/10.3390/app13063945

Holgado-Apaza, L. A., Ulloa-Gallardo, N. J., Aragon-Navarrete, R. N., Riva-Ruiz, R., Odagawa-Aragon, N. K., Castellon-Apaza, D. D., Carpio-Vargas, E. E., Villasante-Saravia, F. H., Alvarez-Rozas, T. P., & Quispe-Layme, M. (2024). The Exploration of Predictors for Peruvian Teachers’ Life Satisfaction through an Ensemble of Feature Selection Methods and Machine Learning. *Sustainability*, *16*(17), 7532. https://doi.org/10.3390/su16177532

Koti Mani Kumar Tirumanadham, N. S., Sekhar, T., & Muthal, S. (2024). An analysis of diverse computational models for predicting student achievement on e-learning platforms using machine learning. *International Journal of Electrical and Computer Engineering*, *14*(6), 7013–7021. https://doi.org/10.11591/ijece.v14i6.pp7013-7021

Silva Payró, M. P., Mena de la Rosa, R., & Cruz Romero, R. (2025). 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*, *10*. https://doi.org/10.31637/epsir-2025-1378

Valencia-Arias, A., Uribe-Gómez, J. A., Flores-Siapo, E., Palacios-Moya, L., Gallegos, A., & Rojas, E. M. (2024). Application of Artificial Neural Networks to Predict the Use of Mobile Learning by University Students. *Human Behavior and Emerging Technologies*, *2024*. https://doi.org/10.1155/hbe2/1518987

Valles-Coral, M. A., Salazar-Ramírez, L., Injante, R., Hernandez-Torres, E. A., Juárez-Díaz, J., Navarro-Cabrera, J. R., Pinedo, L., & Vidaurre-Rojas, P. (2022). Density-Based Unsupervised Learning Algorithm to Categorize College Students into Dropout Risk Levels. *Data*, *7*(11). https://doi.org/10.3390/data7110165