

# CAPÍTULO 1

## Planteamiento del problema



Esta obra está bajo una licencia  
internacional Creative Commons  
Atribución 4.0.



# CAPÍTULO 1

## Planteamiento del problema

### Problem statement

**DOI:** <https://doi.org/10.71112/kv7zdx52>

#### Resumen

El capítulo plantea el problema de la baja integración del aprendizaje automático en la gestión académica de la EPISI de la Universidad Nacional de Moquegua, pese a la digitalización de sus procesos. Se identifica que no se aplican adecuadamente algoritmos como árboles de decisión, Naive Bayes y Random Forest para analizar la productividad educativa basada en el rendimiento académico. La investigación busca determinar la relación entre productividad educativa y aprendizaje automático, justificando su relevancia teórica, práctica, metodológica y social para mejorar la calidad educativa universitaria

**Palabras clave:** Productividad educativa, aprendizaje automático, rendimiento académico, algoritmos, educación superior.

#### Abstract

This chapter addresses the problem of the low integration of machine learning into the academic management of the EPISI at the National University of Moquegua, despite the digitalization of its processes. It identifies the inadequate use of algorithms such as decision trees, Naive Bayes, and Random Forest to analyze educational productivity based on academic performance. The research aims to determine the relationship between educational productivity and machine learning, justifying its theoretical, practical, methodological, and social relevance to improve the quality of university education.

**Keywords:** Educational productivity, machine learning, academic performance, algorithms, higher education.

#### Descripción de la realidad problemática

El nuevo escenario digital denominado cuarta revolución industrial ha transformado las sociedades, no cabe duda que ha alterado la forma de vivir y laboralmente trabajamos de una manera distinta, las nuevas tecnologías, como la computación en la nube, internet de las cosas, inteligencia artificial, etc., se ha integrado a nuestras actividades de manera directa o indirecta,

esto significa que estamos digitalizados (McGinnis, 2020); por ello la digitalización académica no es ajena en UNAM, los procesos están automatizados, esto significa que se encuentran estructurados según el modelo de cada escuela profesional, para nuestro caso *EPISI*, *en donde los procesos automatizados generan información distribuida sin tomar decisiones significativas concernientes a la productividad educativa basados en el rendimiento académico*, es decir a partir de los datos tenemos conocimiento de la realidad académica de los actores involucrados pero no existe una herramienta de aprendizaje automático que analice de manera inteligente el comportamiento de los datos a fin de modelar y evaluar la productividad educativa de los estudiantes mientras van reconociendo debilidades y se van mejorando los resultados educativos desde diferentes enfoques como los algoritmos de aprendizaje automático Arboles de decisión, Naive Bayes y Random Forest.

Por ello, la revisión realizada *destaca* que la generación masiva de datos ha estimulado a la inteligencia artificial a través de los diferentes algoritmos funcionales a aprender de nuestras actividades, asimismo se ha adaptado a nuestras actividades cognitivas; considerando a Sarker (2021) en donde menciona que para analizar inteligentemente los datos y desarrollar aplicaciones inteligentes, es importante aplicar el conocimiento de la inteligencia artificial (IA) particularmente el aprendizaje automático (ML), asimismo según Edgard y Manz (2017), menciona que el aprendizaje automático es el campo de la informática que se encarga de enseñar a las computadoras cómo aprender de los datos, dicho aprendizaje depende del algoritmo que es un conjunto de reglas, en donde indica lo que debe realizar una computadora; Ray (2019) menciona algunos algoritmos de aprendizaje automático más utilizados con mayor frecuencia y populares: Algoritmo de descenso de gradientes, Algoritmo de regresión lineal, Análisis de regresión multivariado, Regresión logística, Árbol de decisión, Máquinas de vectores soporte, Aprendizaje bayesiano, Naïve bayes,c, K means algoritmo de clustering, Algoritmo de propagación de espalda, etc., *dichos algoritmos de aprendizaje automático que en los diferentes ámbitos empresariales han sido muy importantes*, asimismo dichas aplicaciones han generado aportes significativos en la toma de decisiones considerando el nivel de confianza que brindan las diferentes técnicas.

Mediante Ley 30220, SUNEDU en su art. 1, destaca la “promoción del mejoramiento continuo de la calidad educativa de las instituciones universitarias como entes fundamentales del desarrollo nacional, de la investigación y de la cultura”, el tema de calidad educativa universitaria es un punto crucial que ha enfocado SUNEDU en las universidades del país, por tanto, la productividad educativa a través del rendimiento académico es una variable importante de eficiencia y calidad en la educación superior; es aquí en donde los datos juegan

un rol importante a fin de generar información relevante respecto a lo mencionado, según (Hawkins y Bailey, 2020) destaca la importancia del uso y el valor de los datos de los entes que componen la universidad a fin de generar información destacable para diferentes propósitos y niveles de la administración académica; asimismo, a fin de medir la productividad educativa, se debe definir los indicadores necesarios con la finalidad de obtener el producto esperado (Hanushek y Ettema, 2017, como se citó en Triwiyanto T. et al., 2020); por tanto la calidad educativa basado en los datos y cómo van generando valor desde los diferentes niveles establecerán información crucial e importante para el conocimiento de la productividad educativa, cabe recordar que la productividad educativa de los estudiantes es importante en todo sistema educativo, los indicadores influyentes deben cumplir aspectos influyentes a fin de obtener información relevante (Brew et al., 2021); según, Medina N. et al. (2018), recalcan que la medición del rendimiento académico y su predicción considerando las variables que lo determinan representan un tema central en las investigaciones.

Por dichas razones mencionadas hoy en día las universidades cuentan con procesos académicos digitalizados y a nivel de aprendizaje automático se han desarrollado diferentes trabajos de investigación que son mencionados en los antecedentes nacionales e internacionales, utilizando diferentes algoritmos inteligentes, adicionalmente podemos mencionar que según, Castrillón et al. (2020) utilizó el Algoritmo J4; Fernandes et al. (2019) utilizó Minería de datos; Contreras et al. (2020) utilizó Algoritmos de Máquina de Vectores de Soporte y Perceptrón (red neuronal), acordes a los escenarios del ámbito de estudio.

Hasta ahora las diversas investigaciones predictivas basadas en el rendimiento académico han aplicado algoritmos de aprendizaje automático de manera independiente considerando el ámbito de la predicción, tal como se citan en los **antecedentes** del presente trabajo, por tanto, el problema destaca si el algoritmo de aprendizaje automático se relaciona mejor para todos los problemas y si es relevante en la predicción; *por ello el presente trabajo de investigación está enfocado en aplicar diferentes algoritmos de aprendizaje en paralelo a fin de conocer el grado de entrenamiento, aprendizaje y efectividad en los resultados predictivos del ámbito de estudio de la presente investigación.* Asimismo, acorde a la revisión bibliográfica el nuevo conocimiento busca ampliar procedimientos novedosos a fin de magnificar el prestigio y mejora de la calidad educativa en EPISI de UNAM.

En EPISI UNAM aún no ha logrado integrar la productividad educativa a nivel del aprendizaje automático en la gestión académica.

El problema identificado será planteado mediante la aplicación del método analítico que consiste en presentar los síntomas, luego identificar las causas de dichos síntomas,

posteriormente realizar un pronóstico de lo que pasaría si es que no se hace nada respecto a la problemática y finalmente se presenta al control del pronóstico mediante la alternativa de solución de la presente investigación.

El problema de la productividad en EPISI de UNAM se describe de la siguiente manera:

Síntomas de la problemática investigada:

Al respecto se ha determinado lo siguiente:

1. La clasificación de datos y las predicciones de los algoritmos de árboles de decisión son importantes y aplicables al caso, pero no se viene dando en la escuela profesional de ingeniería de sistemas e informática de la UNAM, debido a que no cuenta con una adecuada información debido a que *no se realiza el proceso de inspección, limpia, transformación y modelación del conjunto de datos* con el objetivo de encontrar información útil que sugiera conclusiones que apoyen a toma de decisiones en las actividades académicas.
2. En la escuela profesional de ingeniería de sistemas e informática de la UNAM en sus procesos *no realizan exploración*, clasificación de datos *para el* modelamiento de probabilidades predictivas del algoritmo naive bayes en las actividades académicas.
3. Asimismo, en la escuela profesional de ingeniería de sistemas e informática de la UNAM en sus procesos académicos no se extraen perspectivas valiosas de los datos, *no realizan sobreajustes en el análisis de datos* del algoritmo random forest que apoyen en la toma de decisiones en las actividades académicas.
4. Se ha determinado que en la escuela profesional de ingeniería de sistemas e informática de la UNAM en sus procesos automatizados no aplican el aprendizaje automático, limitando la ubicación de patrones de comportamiento que permitan desarrollar conocimiento para la toma de decisiones.

Causas de la problemática identificada:

Al respecto se ha determinado lo siguiente:

- a) La falta de ponderación en los algoritmos automatizados de la escuela profesional de ingeniería de sistemas e informática de la UNAM, no ha permitido disponer de políticas ni estrategias algorítmicas para la aplicación del algoritmo de aprendizaje automático de árboles de decisión.
- b) Tampoco se dispone de un marco tecnológico para promover la innovación en la gestión de datos que apoyen a los entrenamientos y modelamientos de los datos en base al algoritmo de aprendizaje naive bayes.

- c) Además, la escuela profesional de ingeniería de sistemas e informática de la UNAM no se cuenta con una estrategia basados en algoritmos agiles, paralelos, multifuncional tipo random forest que apoyen en la reducción de tiempo y costos.
- d) Asimismo, la escuela profesional de ingeniería de sistemas e informática de la UNAM no se dispone de políticas de innovación tecnológica para la implementación de sistemas de aprendizaje automático.

#### Pronostico

1. Si la escuela profesional de ingeniería de sistemas e informática de la UNAM no aplicara medidas correctivas respecto a los síntomas continuarán las causas y los efectos antes indicados.
2. *Si no se realiza el proceso de inspección, limpia, transformación y modelación del conjunto de datos* no se permitirá encontrar información útil que sugiera conclusiones que apoyen a toma de decisiones en las actividades académicas.
3. No se logrará organizar, concretar con éxito la exploración de datos y sobreajustes para una adecuada clasificación de datos y predicción de los procesos académicos.
4. No se logrará implementar el sistema de aprendizaje automático que ayude a ubicar patrones de comportamiento en el proceso académico.

#### Control del pronóstico de la situación problemática

Para solucionar los síntomas como las causas antes mencionadas en la presente investigación, se propone como control del pronóstico en la productividad educativa mediante los siguientes elementos:

1. *En la aptitud* académica permitirá obtener información referente a la capacidad de los conocimientos adquiridos, motivaciones adquiridas y el desarrollo planificado de sus actividades productivas en la EPISI de UNAM.
2. *En el Medio ambiente* académico, permitirá conocer apoyo del entorno familiar, el desenvolvimiento en el entorno comunitario social, asimismo la influencia de los medios de comunicación en la productividad educativa en EPISI de UNAM.
3. *En referencia a la Instrucción* académica permitirá planificar los estudios, uso de medios tecnológicos y la calidad de instrucción recibida en el proceso de productividad educativa en EPISI de UNAM.
4. Asimismo, el aprendizaje académico permitirá conocer el logro cognitivo obtenido en el proceso de la productividad educativa.

## **Delimitación de la investigación**

### **a) Delimitación espacial**

El presente trabajo de investigación se realizó en la Universidad Nacional de Moquegua en el año académico 2023, ubicado en Ciudad Jardín s/n del distrito de Pacocha, provincia de Ilo, departamento de Moquegua.

### **b) Delimitación social**

La unidad de análisis correspondió a los estudiantes de la Escuela Profesional de Ingeniería de Sistemas e Informática de la Universidad Nacional de Moquegua, periodo 2023.

### **c) Delimitación temporal**

Esta investigación se inició el 01 de marzo y se culminó en diciembre del año 2023.

### **d) Delimitación conceptual**

Se estudiaron las variables:

**a) Productividad educativa:** “Se centra en identificar y comprender el valor atribuido en los conocimientos de aprendizaje adquiridos que contribuyen al éxito de la productividad educativa mejorando la efectividad de los sistemas educativos; propone los componentes de Aptitud, Instrucción, Medio ambiente y Aprendizaje (Figura 4) que se combinan e influyen entre sí, a fin de determinar el éxito del rendimiento final del estudiante” (Mazana et al., 2019), (Peconcillo Jr et al., 2020).

**b) Algoritmos de aprendizaje automático:** “Es parte de la ciencia de datos, que mediante el uso de métodos estadísticos, los algoritmos son programados, entrenados a fin de realizar clasificaciones o predicciones a fin de conocer información importante, en donde dichos conocimientos permitirán la toma de decisiones dentro de las aplicaciones y las organizaciones empresariales, a fin de generar impacto en su cadena de valores (IBM Cloud Education, 2020), esto significa que los algoritmos se vuelven inteligentes y más en el tiempo, permitiendo establecer efectos positivos de fortalecimiento institucional” (Sharma et al., 2021).

## **Problemas de investigación**

### **Problema principal**

- ¿Cuál es el nivel de relación existe entre la productividad educativa y los algoritmos de aprendizaje en estudiantes EPISI de UNAM, 2023?

## **Problemas específicos**

- ¿Cuál es el nivel de relación que existe entre la dimensión aptitud y los algoritmos de aprendizaje automático en estudiantes EPISI de UNAM, 2023?
- ¿Cuál es el nivel de relación que existe entre la dimensión instrucción y los algoritmos de aprendizaje automático en estudiantes EPISI de UNAM, 2023?
- ¿Cuál es el nivel de relación que existe entre la dimensión medio ambiente y los algoritmos de aprendizaje automático en estudiantes EPISI de UNAM, 2023?
- ¿Cuál es el nivel de relación que existe entre la dimensión aprendizaje y los algoritmos de aprendizaje automático en estudiantes EPISI de UNAM, 2023?

## **Objetivos de la investigación**

### **Objetivo general**

- Determinar el nivel de relación que existe entre la productividad educativa y los algoritmos de aprendizaje automático en estudiantes EPISI de UNAM, 2023

### **Objetivos específicos**

- Determinar el nivel de relación que existe entre la dimensión aptitud y los algoritmos de aprendizaje automático en estudiantes EPISI de UNAM, 2023
- Establecer el nivel de relación que existe entre la dimensión instrucción y los algoritmos de aprendizaje automático en estudiantes EPISI de UNAM, 2023
- Especificar el nivel de relación que existe entre la dimensión medio ambiente y los algoritmos de aprendizaje automático en estudiantes EPISI de UNAM, 2023
- Analizar el nivel de relación que existe entre la dimensión aprendizaje y los algoritmos de aprendizaje automático en estudiantes EPISI de UNAM, 2023

## **Justificación e importancia de la investigación.**

Según Hernández-Sampieri R. (2018) en la investigación, la justificación de una investigación implica detallar las razones que respaldan su realización. Esto significa que justificar un estudio consiste en presentar los motivos que demuestran la relevancia de llevar a cabo dicho análisis, fundamentándose en los objetivos y las preguntas planteadas en la investigación actual.

### **a) Justificación teórica**

Considerando la existencia del enfoque de modelado de datos y considerando la existencia de algoritmos iniciales en donde son interpretados; machine learning automatiza el proceso de

datos mediante la producción de modelos tanto en los datos actuales, como en los futuros, proporcionando una aceleración sin precedentes concernientes a la productividad educativa basado en el rendimiento académico de EPISI de la Universidad Nacional de Moquegua. El marco teórico concerniente a la productividad educativa esta referenciado por el modelo de productividad educativa propuesto por Herbert J. Walberg, basado en la teoría del aprendizaje cognitivo; en referencia a los algoritmos de aprendizaje automático se consideran Arboles de decisión, Naive Bayes y Random Forest, a fin de realizar un análisis predictivo mediante el aprendizaje de los datos de reciente recogida sin necesidad de intervención humana y sin hacer falta de una programación explícita.

**b) Justificación práctica**

Conscientes de la importancia que va adquiriendo el uso correcto de los sistemas inteligentes basados en modelos de aprendizaje predictivo en tecnologías de la información enfocados al rendimiento académico en la Universidad Nacional de Moquegua y los indicadores edad, genero, lugar de procedencia, centro laboral, tipo de preparación, promedio final, modalidad de ingreso, cursos matriculados, escuela profesional, tipo de matrícula, porcentaje de predicción acertada y el uso de la eficiencia de los algoritmos de aprendizaje automática y que esta justificación metodológica ayudara en la mejora de la productividad académica en la Universidad Nacional de Moquegua.

El presente trabajo de investigación tuvo como objetivo observar la relevancia del uso del uso de la eficiencia de los algoritmos de aprendizaje automático para el sistema académico basado en la productividad educativa de EPISI de la UNAM.

**c) Justificación metodológica**

Acorde a los objetivos de la investigación, se llevó a cabo un proceso metodológico estructurado y sistemático, empleando técnicas de investigación cuantitativa enfocadas en el análisis y la síntesis respecto al uso de algoritmos de aprendizaje automático para la predicción de la productividad educativa en base al rendimiento académico en la escuela profesional de ingeniería de sistemas e informática de la Universidad Nacional de Moquegua., determinando en ambas variables los procedimientos para la jerarquización de los factores descriptivos y explicativos.

**d) Justificación social**

Las universidades hoy en día se involucran en actividades sociales mediante la divulgación de los conocimientos, para ello es importante la productividad académica a fin de elevar los conocimientos junto a la innovación y transferencia tecnológica, basado en la sociedad del conocimiento. El protagonismo de las universidades peruanas en la creación, transmisión y

aplicación de conocimiento hace prescindible el uso de algoritmos de aprendizaje para predecir los rendimientos académicos y fortalecer los conocimientos adquiridos en la UNAM.

### **Importancia**

La importancia de aplicar nuevas técnicas inteligentes en el presente trabajo de investigación establecerá nuevos horizontes para realizar estudios sustantivos que presenten situaciones similares planteadas en la presente investigación, asimismo el marco referencial contribuirá a la implementación de políticas educativas relacionadas al mejoramiento de la productividad educativa basado en logros de aprendizaje en un marco de equidad y justicia. También considerar que las nuevas técnicas de aprendizaje automático brindarán información valiosa para mejorar las prácticas pedagógicas y establecer las estrategias de aprendizaje que afectan el rendimiento académico.

### **Factibilidad de la investigación**

Se dispuso de recursos como tiempo, asesoría continua y acceso a información, aunque en algunos casos esta fue limitada. También se contaron con recursos económicos restringidos y una actitud personal comprometida para llevar a cabo la investigación de manera responsable y completa, con el fin de graduarme. Los recursos mencionados se utilizaron para evaluar la viabilidad de la investigación, mediante el cumplimiento del objetivo específico.

### **Limitaciones del estudio**

El desarrollo tuvo algunas limitaciones referentes a la obtención de los datos para el presente estudio, para ello se consideró estudiantes regulares e irregulares matriculados en la escuela profesional de ingeniería de sistemas de UNAM, con un promedio de edad 20 hasta los 26 años de edad aproximadamente concernientes a EPISI.

Es importante mencionar que el presente estudio estuvo orientado a obtener información relevante conforme al modelo de productividad educativa propuesto por Herbert J. Walberg en las dimensiones de aptitud, instrucción, medio ambiente y aprendizaje, ampliado con indicadores correspondiente a fin de fortalecer las variables de estudio, asimismo los algoritmos en estudio desarrollaran el análisis, clasificación y aprendizaje respectivo afianzando el propósito del presente trabajo de investigación.

El acceso a los datos a la población objetivo puede resultar difícil para la investigación, para ello los cuestionarios se aplicaron en forma directa en las inmediaciones de la Escuela Profesional de Ingeniería de Sistema e Informática de la Universidad Nacional de Moquegua.

Las limitaciones temporales no pudieron afectar la cantidad de datos recopilados, para ello se tuvo que planificar cuidadosamente el cronograma de investigación, establecer metas realistas y utilizar métodos eficientes de recolección de datos, junto a la flexibilización y la adaptación ante imprevistos.

## Referencias

Brew, E. A. y Nketiah, B. y Koranteng, R. (2021). A Literature Review of Academic Performance, an Insight into Factors and their Influences on Academic Outcomes of Students at Senior High Schools. *OALib*, 08(06), 1–14. <https://doi.org/10.4236/oalib.1107423>

Castrillón, O. D. y Sarache, W. y Ruiz-Herrera, S. (2020). Predicción del rendimiento académico por medio de técnicas de inteligencia artificial. *Formación Universitaria*, 13(1), 93–102. <https://doi.org/10.4067/S0718-50062020000100093>

Contreras, L. E. y Fuentes, H. J. y Rodríguez, J. I. (2020). Predicción del rendimiento académico como indicador de éxito/fracaso de los estudiantes de ingeniería, mediante aprendizaje automático. *Formación Universitaria*, 13(5), 233–246. <https://doi.org/10.4067/S0718-50062020000500233>

Edgard, T. W. y Manz, D. O. (2017). *Research Methods for Cyber Security* (1ra ed.). Elsevier Inc.

Fernandes, E. y Holanda, M. y Victorino, M. y Borges, V. y Carvalho, R. y Erven, G. van. (2019). Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research*, 94, 335–343. <https://doi.org/10.1016/j.jbusres.2018.02.012>

Hawkins, C. y Bailey, L. E. (2020). A New Data Landscape: IR's Role in Academic Analytics. *New Directions for Institutional Research*, 2020(185–186), 87–103. <https://doi.org/10.1002/ir.20331>

Hernández-Sampieri R. (2018). *Metodología de la investigación* (1ra Ed.).

IBM Cloud Education. (2020). *Machine Learning This introduction to machine learning provides an overview of its history, important definitions, applications, and concerns within businesses today*. <https://www.ibm.com/cloud/learn/machine-learning>

Mazana, M. Y. y Montero, C. S. y Casmir, R. O. (2019). Investigating Students' Attitude towards Learning Mathematics. *International Electronic Journal of Mathematics Education*, 14(1). <https://doi.org/10.29333/iejme/3997>

McGinnis, D. (2020). *¿Qué es la Cuarta Revolución Industrial? | Fuerza de ventas.* <https://www.salesforce.com/blog/what-is-the-fourth-industrial-revolution-4ir/>

Medina N. y Fereira J. y Marzol R. (2018). Factores personales que inciden en el bajo rendimiento académico de los estudiantes de geometría. *Revista de Estudios Interdisciplinarios En Ciencias Sociales*.

Peconcillo Jr, L. B. y D. Peteros, E. y O. Mamites, I. y T. Sanchez, D. y L. Tenerife, J. J. y L. Suson, R. (2020). Structuring Determinants to Level Up Students Performance. *International Journal of Education and Practice*, 8(4), 638–651. <https://doi.org/10.18488/journal.61.2020.84.638.651>

Ray, S. (2019). A Quick Review of Machine Learning Algorithms. *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 35–39. <https://doi.org/10.1109/COMITCon.2019.8862451>

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2(3), 160. <https://doi.org/10.1007/s42979-021-00592-x>

Sharma, N. y Sharma, R. y Jindal, N. (2021). Machine Learning and Deep Learning Applications-A Vision. *Global Transitions Proceedings*, 2(1), 24–28. <https://doi.org/10.1016/j.gltip.2021.01.004>

Triwiyanto T. y Suyanto y Prasojo L.D. y Wardana Y. (2020). Factors Affecting Educational Productivity at Private Elementary Schools in Indonesia. *Advances in Social Science, Education and Humanities Research*, 487, 318–323.