



FACULTAD DE INGENIERÍA

Carrera de Ingeniería de Sistemas Computacionales

**APLICACIÓN DE ALGORITMOS DE MACHINE LEARNING PARA
LA PREDICCIÓN DEL BAJO RENDIMIENTO ACADÉMICO EN LOS
ESTUDIANTES DEL 6° GRADO DE PRIMARIA EN LOS CURSOS DE
MATEMÁTICA Y COMUNICACIÓN DE LA I.E N°16698 MANUEL
GONZALES PRADA LA BERMEJA, AÑOS 2024 - 2025**

**Tesis para optar al título profesional de:
Ingeniero de Sistemas Computacionales**

Autores:

Jean Miguel Lalangui Quinde
Omar Francisco Mori Jaramillo

Asesor:

Mg. José Antonio Ogosi Auqui
<https://orcid.org/0000-0002-4708-610X>

Lima - Perú

2025

JURADO EVALUADOR

Jurado 1 Presidente(a)	NEICER CAMPOS VASQUEZ
	Nombre y Apellidos

Jurado 2	LAURA SOFIA BAZAN DIAZ
	Nombre y Apellidos

Jurado 3	JOSE ANTONIO OGOSI AUQUI
	Nombre y Apellidos

Informe de Similitud

E-8-N00223594-DB8E2A6D49

INFORME DE ORIGINALIDAD

19% INDICE DE SIMILITUD	18% FUENTES DE INTERNET	10% PUBLICACIONES	10% TRABAJOS DEL ESTUDIANTE
-----------------------------------	-----------------------------------	-----------------------------	---------------------------------------

FUENTES PRIMARIAS

1	repositorio.unprg.edu.pe Fuente de Internet	6%
2	www.clubensayos.com Fuente de Internet	3%
3	repositorio.ucv.edu.pe Fuente de Internet	1%
4	repositorio.upn.edu.pe Fuente de Internet	<1%
5	Submitted to Universidad Privada del Norte Trabajo del estudiante	<1%
6	repositorio.upse.edu.ec Fuente de Internet	<1%
7	hdl.handle.net Fuente de Internet	<1%
8	repositorio.uchile.cl Fuente de Internet	<1%
9	repositorio.unsm.edu.pe Fuente de Internet	<1%
10	www.coursehero.com Fuente de Internet	<1%
11	repositorio.unheval.edu.pe Fuente de Internet	<1%

Tabla de contenidos

Índice de tablas	6
Índice de Figuras	7
Resumen	8
CAPÍTULO I: INTRODUCCIÓN.....	9
1.1. Realidad problemática	9
1.2. Formulación del problema.....	19
1.3. Objetivos.....	20
1.4. Hipótesis	20
CAPÍTULO II: METODOLOGÍA	25
CAPÍTULO III: RESULTADOS	50
CAPÍTULO IV: DISCUSIÓN Y CONCLUSIONES	66
REFERENCIAS	73
ANEXOS.....	79

Índice de tablas

Tabla 1. Tabla comparativa entre modelo de datos	18
Tabla 2. Tipificación de la investigación.....	27
Tabla 3. Población y muestra de la investigación	28
Tabla 4. Criterio de inclusión y exclusión para la muestra	30
Tabla 5. Escala de calificaciones	31
Tabla 6. Escala de rendimiento académico.....	32
Tabla 7. Correlación de escalas académicas.....	32
Tabla 8. Fases de metodología SEMMA.....	33
Tabla 9. Fases de metodología KDD	34
Tabla 10 Fases de metodología TDSP	35
Tabla 11 Fases de metodología CRISP-DM.....	36
Tabla 12. Distribución de clases en el conjunto de prueba.....	51
Tabla 13. Correlación de variables con el riesgo.....	52
Tabla 14. Matriz de confusión para comunicación.....	55
Tabla 15. Matriz de confusión para matemática.....	55
Tabla 16. Resultados por clases y modelo entrenado para comunicación.....	56
Tabla 17. Resultados por clases y modelo entrenado para matemática.....	57
Tabla 18. Resultados de indicadores macro para modelos de ML de matemática y comunicación.....	58
Tabla 19. Clasificación binaria para DT.....	61
Tabla 20. Vista binaria de intervención de DT	61

Índice de Figuras

Figura 1. Esquema organizacional de la investigación	25
Figura 2. Esquema de metodología de la investigación	29
Figura 3 Flujograma del proceso CRISP-DM.....	38
Figura 4. Distribución de registros académicos	50
Figura 5. Comparación correlacional	53
Figura 6. Comparativo Matemática.....	59
Figura 7. Comparativo comunicación.	60

Resumen

El estudio aborda la necesidad de predecir el bajo rendimiento en sexto grado, utilizando datos institucionales de calificaciones, conducta y asistencia, para apoyar decisiones pedagógicas oportunas. El objetivo fue comparar la efectividad de tres algoritmos de aprendizaje automático, Regresión Logística (RL), Decision Tree (DT) y Random Forest (RF) en los cursos matemática y comunicación durante 2024 y 2025. El diseño fue aplicado, no experimental, longitudinal; Los modelos fueron entrenados con datos del 2024 y se evaluó con 2025 (B1–B2). Se analizó la correlación (punto-biserial) y se evaluó desempeño mediante accuracy, precisión, recall y F1-Score. Los resultados confirmaron una asociación negativa y significativa entre variables académicas y riesgo (nota: $r = -0.714$ en matemática; -0.799 en comunicación; $p < 0.001$). En pruebas, DT obtuvo el mejor rendimiento (comunicación: accuracy 0.975; F1 0.978; matemática: 0.950; F1 0.953), superando a RF y RL. En la vista binaria de riesgo logró cero falsos positivos y detección completa de casos severos, generando listas de intervención accionables y priorizables por severidad. Se concluye que los modelos de ML son eficaces para el problema estudiado y que Decisión Tree es el modelo recomendado para su implementación escolar, con reentrenamientos periódicos, validación docente y salvaguardas éticas.

Palabras Claves

Rendimiento académico; Machine learning; decision tree; CRIPS-DM; Scikit-learn

NOTA

El contenido de la investigación no se encuentra disponible en **acceso abierto** por determinación de los propios autores, en concordancia con en el Texto Integrado del Reglamento RENATI (artículo 12), la Directiva N°048-2020-CONCYTEC-P que regula el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto (ALICIA) administrado por el pliego Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica - CONCYTEC y la Ley N° 29733, Ley de Protección de Datos Personales.

REFERENCIAS

- Aguilar, J. E., Mejia, E. F., Morocho, T. P., & Velasco, G. (2025). Estudio del rendimiento académico mediante la comparación de modelos de regresión y árboles de clasificación. *Telos: Revista de Estudios Interdisciplinarios en Ciencias Sociales*, 27(1), 94-115.
- Alfaro, A. D. J., & Ospina, J. V. D. (2021). Revisión sistemática de literatura: Técnicas de aprendizaje automático (Machine Learning). *Cuaderno activa*, 13(1), Article 1.
- Asana. (2025, febrero 21). *El árbol de decisiones: Un análisis de 5 pasos para tomar mejores decisiones [2025]* • Asana. Asana. <https://asana.com/es/resources/decision-tree-analysis>
- Barrera, D. C. R. P. (2022). La representatividad de una muestra en investigaciones educativas. *CIENCIAS PEDAGÓGICAS*, 15(1), Article 1.
- Belcic, I., & Stryker, cole. (2024, diciembre 28). *¿Qué es el aprendizaje supervisado? | IBM. ¿Qué es el aprendizaje supervisado?* <https://www.ibm.com/mx-es/think/topics/supervised-learning>
- Bustamante, G. J. (2022). *Factores que inciden en el rendimiento académico de los estudiantes de bachillerato en el cantón Sucúa-Ecuador*. <https://dspace.ucacue.edu.ec/handle/ucacue/13051>
- Candia, D. I. (2019). Predicción del rendimiento académico de los estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático. *Universidad Nacional de San Antonio Abad del Cusco*. <https://repositorio.unsaac.edu.pe/handle/20.500.12918/4120>
- Castañeda, G., Canales, S., Ovalle, C., & Rabanal, E. H. (2023). Intelligent predictive model applying Data Mining strategies for a credit evaluation of a commercial

- company. *LACCEI*, *I*(8), Article 8.
<https://doi.org/10.18687/LACCEI2023.1.1.1148>
- Cimas, G. (2022, julio 22). *Visual Studio Code: Editor de código para desarrolladores | OpenWebinars*. OpenWebinars.net. <https://openwebinars.net/blog/que-es-visual-studio-code-y-que-ventajas-ofrece/>
- Clemmensen, L. H., & Kjaersgaard, R. D. (2022, marzo 9). *Data Representativity for Machine Learning and AI Systems*. arXiv.Org.
<https://arxiv.org/abs/2203.04706v2>
- Del Carpio, R. (2024). Predicción del rendimiento académico utilizando modelos de aprendizaje automático: Una revisión sistemática de la literatura. *593 Digital Publisher CEIT*, *9*(6), 1038-1054.
- Díaz, B., Meleán, R. A., & Marín, W. J. (2021). Rendimiento académico de estudiantes en Educación Superior: Predicciones de factores influyentes a partir de árboles de decisión. *Telos: Revista de Estudios Interdisciplinarios en Ciencias Sociales*, *23*(3 (septiembre-diciembre)), 616-639.
- Díaz, M. A., Ahumada, M. de los A., & Melo, J. P. (2021). Árboles de Decisión como Metodología para Determinar el Rendimiento Académico en Educación Superior. *Revista Lasallista de Investigación*, *18*(2), 94-104.
<https://doi.org/10.22507/rli.v18n2a8>
- Espino, C. (2017). *Análisis predictivo: Técnicas y modelos utilizados y aplicaciones del mismo - herramientas Open Source que permiten su uso*.
<https://hdl.handle.net/10609/59565>
- Espino, J. Q. (2022). *Algoritmo para la extracción del conocimiento (KDD) a través del análisis de textos aplicado en la investigación social*. <https://ri-ng.uaq.mx/handle/123456789/3657>

- Espinosa, J. J. (2020). Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública. *Ingeniería, investigación y tecnología*, 21(1). <https://doi.org/10.22201/fi.25940732e.2020.21n1.008>
- Faneite, S. F. A. (2023). Los enfoques de investigación en las Ciencias Sociales. *Revista Latinoamericana Ogmios*, 3(8), Article 8. <https://doi.org/10.53595/rlo.v3.i8.084>
- Fernández, S. (2022). *Métodos de regresión y clasificación basados en árboles*. <https://uvadoc.uva.es/handle/10324/53822>
- Fois, G., Agüero Crovella, G. A., & Britos, P. V. (2020). *Evaluación comparativa de las metodologías Team Data Science Process TDSP y Analytics Solutions Unified Method for Data Mining ASUM-DM desde la perspectiva de la ciencia de datos*. Instituto Antioqueño de Investigación. Medellín, Antioquia. <https://doi.org/10.5281/zenodo.4031253>
- González, R. (2010). *Python para todos*. Self publishing. <http://archive.org/details/2010PythonParaTodos>
- Hernández, & Mendoza, C. P. (2018). *Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta*. <https://virtual.cuautitlan.unam.mx/rudics/?p=2612>
- Hernández, O. (2021). Aproximación a los distintos tipos de muestreo no probabilístico que existen. *Revista Cubana de Medicina General Integral*, 37(3). http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S0864-21252021000300002&lng=es&nrm=iso&tlng=es
- Hidalgo, A. (2024). *Comenzando con Numpy y Pandas*. <http://hdl.handle.net/10637/16011>

- Holguín, G. A., Moreno, D. A., & Holguín, M. (2024). *Introducción a la ciencia y análisis de datos con Python: Una visión empresarial*.
<https://doi.org/10.22517/9789587229714>
- IDLE — *documentación de Python*—3.8.20. (s. f.). Recuperado 17 de julio de 2025, de <https://docs.python.org/es/3.8/library/idle.html>
- INEI. (2023, octubre 31). *Perú: Indicadores de Educación según Departamentos, 2012-2022*. <https://www.gob.pe/institucion/inei/informes-publicaciones/4793457-peru-indicadores-de-educacion-segun-departamentos-2012-2022>
- Jiang, Q. (2023). *Métodos de Clasificación con Python: Aplicaciones Empresariales*.
<https://riunet.upv.es/handle/10251/195236>
- López, D. F. (2024). *Implementación de algoritmos de parallel boosting*.
<https://hdl.handle.net/1992/73842>
- Martínez, J. A., & Pérez, P. S. (2024). Regresión logística. *Medicina de Familia. SEMERGEN*, 50(1), 102086. <https://doi.org/10.1016/j.semerg.2023.102086>
- MINEDU. (2023a, abril 3). *Evaluación Muestral de Estudiantes 2022 presenta resultados más bajos que los de 2019*.
<https://www.gob.pe/institucion/minedu/noticias/737725-evaluacion-muestral-de-estudiantes-2022-presenta-resultados-mas-bajos-que-los-de-2019>
- MINEDU, O. de prensa. (2023b, diciembre 12). *Colegios utilizarán este año escala literal para calificar a estudiantes*.
<https://www.gob.pe/institucion/minedu/noticias/880490-colegios-utilizaran-este-ano-escala-literal-para-calificar-a-estudiantes>
- Ministerio para la Transformación Digital y de la Función Pública. (2021, enero 26). *¿Cómo sé si mi modelo de predicción es realmente bueno? | datos.gob.es*.

<https://datos.gob.es/es/blog/como-se-si-mi-modelo-de-prediccion-es-realmente-bueno>

MINSA. (s. f.). *Resolución Ministerial N.º 339-2023-MINSA*. Recuperado 3 de septiembre de 2025, de <https://www.gob.pe/institucion/minsa/normas-legales/4034345-339-2023-minsa>

Mucha, L. F., Chamorro, R., Oseda, M. E., & Alania, R. D. (2021). Evaluación de procedimientos empleados para determinar la población y muestra en trabajos de investigación de posgrado. *Desafíos*, 12(1), Article 1. <https://doi.org/10.37711/desafios.2021.12.1.253>

Novoa, Y. G. (2021). Nociones de la investigación y el potencial de la investigación aplicada como estrategia de innovación. *Encuentro SENNOVA del Oriente Antioqueño*, 7(1), Article 1. <https://doi.org/10.23850/26652447/7/1/3720>

Ogosi, J. A. (2025). *Implementación de una aplicación móvil con machine learning aplicando la nueva metodología CSKT para el proceso de contratación de personal en el sector privado*. <https://repositorio.unfv.edu.pe/handle/20.500.13084/10220>

Ordoñez, Á. F. (2025). Metodología de la Investigación Metodología académica con aplicación a las investigaciones sociales: Enfoques, tipos, métodos y diseños. *Sociedad & Tecnología*, 8(2), Article 2. <https://doi.org/10.51247/st.v8i2.484>

Perez, B. R. (2020). Comparación de técnicas de minería de datos para identificar indicios de deserción estudiantil, a partir del desempeño académico. *Revista UIS Ingenierías*, 19(1(2020)), 193-204. <https://doi.org/10.18273/revuin.v19n1-2020018>

Pineda, R. (2021). *Comparación en términos de precisión predictiva de tres algoritmos de aprendizaje supervisado aplicado a la detección de sitios Phishing haciendo*

uso de la librería Scikit-Learn [Other, Universidad de San Carlos de Guatemala].

<https://biblioteca.ingenieria.usac.edu.gt/>

Quiroz, A. G. del M., & Silva, A. L. (2025). *Regresión logística ordinal para identificar factores asociados al desempeño escolar en estudiantes de educación secundaria*

de una institución educativa pública. Suyo 2023.

<http://repositorio.unprg.edu.pe/handle/20.500.12893/14453>

Scikit-learn, developers. (2025). 3.4. *Metrics and scoring: Quantifying the quality of predictions*. Scikit-Learn. [https://scikit-](https://scikit-learn/stable/modules/model_evaluation.html)

[learn/stable/modules/model_evaluation.html](https://scikit-learn/stable/modules/model_evaluation.html)

Streamlit. (2021, enero 14). *Streamlit • A faster way to build and share data apps*

[Software]. Streamlit Documentation. <https://streamlit.io/>

Universidad Europea. (2025, febrero 10). *Datasets: Qué es, tipos y dónde encontrarlos* |

Blog UE [Institucional]. Datasets: qué es, tipos y dónde encontrarlos.

<https://universidadeuropea.com/blog/datasets-que-es/>