

Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.  
ISSN 2707-2207 / ISSN 2707-2215 (en línea), julio-agosto 2025,  
Volumen 9, Número 4.

[https://doi.org/10.37811/cl\\_rcm.v9i2](https://doi.org/10.37811/cl_rcm.v9i2)

# **APLICACIÓN DE APRENDIZAJE AUTOMÁTICO PARA LA PREDICCIÓN DE LA EFICIENCIA TERMINAL EN EL TECN CAMPUS CHIHUAHUA MÉXICO**

**APPLICATION OF MACHINE LEARNING FOR PREDICTING  
GRADUATION EFFICIENCY AT TECN  
CHIHUAHUA**

**Abraham Efraím Rodríguez-Mata**

Tecnológico Nacional de México campus Chihuahua – México

**Elizabeth Rodríguez Salcido**

Tecnológico Nacional de México campus Chihuahua – México

**Víctor Alejandro González-Huitrón**

Tecnológico Nacional de México campus Querétaro- México

**Jesus Alfonso Medrano-Hermosillo**

Tecnológico Nacional de México campus Querétaro México

DOI: [https://doi.org/10.37811/cl\\_rcm.v9i4.18687](https://doi.org/10.37811/cl_rcm.v9i4.18687)

## Aplicación de Aprendizaje Automático para la Predicción de la Eficiencia Terminal en el TecNM campus Chihuahua México

Abraham Efraím Rodríguez-Mata<sup>1</sup>

[abraham.rm@chihuahua.tecnm.mx](mailto:abraham.rm@chihuahua.tecnm.mx)

<https://orcid.org/0000-0002-0262-420X>

Tecnológico Nacional de México campus  
Chihuahua., México

Elizabeth Rodríguez Salcido

[elizabethrs@chihuahua.tecnm.mx](mailto:elizabethrs@chihuahua.tecnm.mx)

<https://orcid.org/0009-0007-0747-163X>

Tecnológico Nacional de México campus  
Chihuahua., México

Víctor Alejandro González-Huitrón

[Victor.gh@queretaro.tecnm.mx](mailto:Victor.gh@queretaro.tecnm.mx)

<https://orcid.org/0000-0003-0426-0515>

Tecnológico Nacional de México campus  
Querétaro México

Jesus Alfonso Medrano-Hermosillo

[Jesus.hh@chihuahua.tecnm.mx](mailto:Jesus.hh@chihuahua.tecnm.mx)

<https://orcid.org/0000-0002-5152-220X>

Tecnológico Nacional de México campus Querétaro  
México

### RESUMEN

La presente investigación examina la factibilidad de implementar técnicas de aprendizaje automático para pronosticar la eficiencia terminal —es decir, el porcentaje de estudiantes que concluyen sus estudios en un periodo máximo de 12 semestres— en el Tecnológico Nacional de México, campus Chihuahua, para las cohortes 2016 a 2020. Se empleó un conjunto de datos compuesto por diez cohortes semestrales, las cuales incluyen datos de ingreso, egreso y titulación, desagregadas por género y programa formativo. Después de un meticuloso preprocesamiento y codificación de variables (incluyendo una dummy de "cohorte en pandemia"), se entrenaron modelos de regresión logística, un Foresta Aleatoria, XGBoost, análisis de supervivencia (Cox y Foresta Aleatoria de Supervivencia) y un Perceptrón Multicapa (MLP) con dos capas encubiertas. El MLP, tras un entrenamiento de 100 etapas, logró un Error Medio Cuadrático de 0.03408 y un R<sup>2</sup> de 0.28, reflejando la tendencia general de la serie en cuestión. Además, los hallazgos indican una disminución de la eficiencia media de 39.9 % (pre-pandemia) a 4.8 % (post-pandemia) y una discrepancia de género superior a 13 puntos porcentuales (20.6% en mujeres contra 6.8 % en hombres). Se aborda la integración de estas variables en sistemas de alerta precoz y la necesidad de enriquecer el dataset para potenciar su robustez y confiabilidad.

**Palabras clave:** aprendizaje automático, eficiencia terminal, deserción escolar, inteligencia artificial, educación superior

---

<sup>1</sup> Autor principal

Correspondencia: [braham.rm@chihuahua.tecnm.mx](mailto:braham.rm@chihuahua.tecnm.mx)

# Application of Machine Learning for Predicting Graduation Efficiency at TecNM Campus Chihuahua

## ABSTRACT

This study examines the viability of employing machine learning methods to forecast graduation efficiency, defined as the proportion of students graduating within 12 semesters, at Tecnológico Nacional de México, Campus Chihuahua, for the 2016–2020 enrollment cohorts. A dataset comprising ten biannual cohorts, featuring comprehensive statistics on enrollment, graduation, and degree conferment categorized by gender and program, was preprocessed and encoded, including a dummy variable for the "pandemic cohort." We trained logistic regression, Random Forest, XGBoost, survival analysis models (Cox and Random Survival Forest), and a Multilayer Perceptron (MLP) comprising two hidden layers. After 100 training epochs, the MLP attained a Mean Squared Error of 0.03408 and an R2 of 0.28, effectively capturing the overarching trend. The findings indicate a decline in average efficiency from 39.9% (pre-pandemic) to 4.8% (post-pandemic), alongside a gender disparity surpassing 13 percentage points (20.6% for females compared to 6.8% for males). We address the incorporation of these insights into early-warning systems and the necessity of augmenting the dataset to improve model robustness.

**Keywords:** machine learning, graduation efficiency, student dropout, artificial intelligence, higher education

*Artículo recibido 15 junio 2025*

*Aceptado para publicación: 19 julio 2025*



## INTRODUCCIÓN

La deserción académica representa un fenómeno complejo que repercute adversamente en la calidad y equidad de la educación superior. En América Latina, se ha observado un aumento en las tasas de deserción escolar en años recientes, atribuible a factores socioeconómicos, tecnológicos y académicos (González, Cruz & Gómez, 2022). Durante la pandemia de COVID-19 se observó una intensificación significativa de este problema, con un impacto particular en estudiantes de bajos recursos con acceso limitado a tecnología y conectividad (Sabino & Acuña, 2022). En este contexto, la inteligencia artificial ha emergido como una estrategia innovadora para la predicción y mitigación de la deserción escolar (Carreño, Martínez & Páez, 2023).

La pandemia de COVID-19 ejerció un impacto sin paralelo en el sector educativo, provocando un aumento significativo en las tasas de deserción escolar, especialmente en América Latina. Las restricciones en el sector sanitario propiciaron una acelerada transición hacia modelos de educación a distancia, lo que puso de manifiesto las desigualdades preexistentes en el acceso a la tecnología y la conectividad. Según González y colaboradores (2022), la insuficiencia de recursos digitales y la infraestructura tecnológica insuficiente en sectores vulnerables impidieron la continuidad académica, afectando predominantemente a estudiantes de nivel básico y medio. Esta situación se agravó por las deficiencias en la capacitación pedagógica para la enseñanza a distancia, lo cual impactó adversamente en la calidad educativa y redujo las oportunidades de aprendizaje significativo.

La pandemia también ejerció una influencia significativa en la deserción escolar. Numerosas familias, impactadas por la desocupación laboral y la volatilidad financiera, priorizaron la subsistencia por encima de la educación de sus hijos (Sabino & Acuña, 2022). En circunstancias específicas, los estudiantes se vieron forzados a incorporarse prematuramente en el mercado laboral con el objetivo de contribuir a la economía familiar, lo que redujo su tiempo de estudio y elevó la probabilidad de deserción escolar (Bustos Narváez, 2023). De acuerdo con Arellano-Esparza y Ángeles Ortiz-Espinoza (2022), las consecuencias económicas de la crisis sanitaria fueron más severas en comunidades desfavorecidas, en las que la falta de apoyo gubernamental y la insuficiencia de programas de asistencia educativa limitaron la reincorporación de los estudiantes al sistema educativo.

La reincorporación al entorno presencial ha constituido un proceso heterogéneo y complejo.



Académicos contemporáneos indican que, a pesar de que en determinadas naciones las tasas de inscripción han iniciado un proceso de recuperación, en numerosas regiones persisten desigualdades significativas en el acceso a una educación de alta calidad (Muñoz & Cadena, 2024). Según de Souza (2024), un porcentaje considerable de estudiantes que dejaron la institución durante la pandemia no ha retornado, lo que representa un impedimento para las políticas educativas en el periodo postpandemia. Es esencial que las tácticas de rehabilitación integren enfoques adaptativos, tales como la pedagogía híbrida y programas de apoyo psicosocial, con la finalidad de garantizar la reincorporación y retención de los estudiantes.

El aprendizaje automático ha evidenciado su efectividad en la pronosticación de la deserción escolar (Parra-Sánchez, Torres Pardo & Martínez De Merino, 2023; Vite Ayala, Paredes Camacho, Quiroga Gallo & González Gutiérrez, 2023). Estas tecnologías facilitan la interpretación de extensos conjuntos de datos educativos, la detección de patrones y la formulación de estrategias de intervención personalizadas.

Este estudio multidisciplinario tiene como objetivo evaluar la eficacia de los modelos de aprendizaje automático en la predicción de la deserción escolar en el Tecnológico Nacional de México Campus Chihuahua. El método adoptado engloba la evaluación de bases de datos institucionales, la elección y capacitación de algoritmos de aprendizaje automático, y la valoración de su rendimiento a través de indicadores tales como precisión, sensibilidad y especificidad.

Esta investigación se organiza de la siguiente manera: en la Sección II se presenta el marco teórico, en el que se ofrece un resumen conciso de las cuestiones asociadas con la inteligencia artificial, se especifican los principios fundamentales de aprendizaje automático y su implementación en el procesamiento de datos escolares, además de los antecedentes del uso de la Inteligencia Artificial en el contexto educativo. La Sección III detalla la metodología implementada, detallando los modelos y métodos empleados, así como la utilización de datos estadísticos para la toma de decisiones en nuestro instituto. La Sección IV expone los descubrimientos y su correspondiente análisis. La Sección V examina las implicaciones éticas y operativas de la inteligencia artificial en la gestión educativa. Finalmente, la Sección VI expone las conclusiones y las directrices para futuras investigaciones.



## Marco Teórico

Investigaciones contemporáneas han evidenciado que la deserción escolar no constituye un fenómeno aislado, sino el producto de una serie de factores interrelacionados. Las causas primordiales del abandono estudiantil incluyen un rendimiento académico deficiente, la ausencia de respaldo institucional, la inestabilidad económica y la inadaptación a nuevos modelos educativos . Adicionalmente, la infraestructura tecnológica y la administración educativa desempeñan un papel crucial en la retención estudiantil (José H., Wilmer & Sánchez, 2024).

La incorporación de la inteligencia artificial en el ámbito educativo plantea retos significativos. Uno de los desafíos principales radica en la protección de la privacidad de los datos estudiantiles, dado que la utilización de información delicada demanda regulaciones explícitas para salvaguardar la identidad de los alumnos (Vázquez-Cano & Pascual Moscoso, 2022). Además, se ha detectado que los modelos predictivos pueden exhibir sesgos algorítmicos, lo que podría provocar disparidades en el proceso de toma de decisiones académicas (Bolaños Gramajo, 2023). Para prevenir esta dificultad, resulta esencial desarrollar sistemas de auditoría y supervisión en la implementación de estos modelos (Aznarte, 2020).

La investigación contemporánea ha evidenciado que la inteligencia artificial (IA) juega un papel crucial en la detección precoz de factores que inciden en la deserción escolar y en la optimización de la gestión educativa. Mediante la implementación de algoritmos de aprendizaje automático y el análisis de grandes volúmenes de datos, las instituciones educativas pueden elaborar modelos predictivos capaces de prever el riesgo de abandono y diseñar estrategias de intervención personalizadas. La capacidad de la IA para identificar patrones complejos y correlaciones entre el desempeño académico y la condición socioeconómica de los estudiantes es fundamental para potenciar la retención y robustecer la equidad educativa. Uno de los enfoques más eficaces en la aplicación de la IA en educación es la utilización de modelos de clasificación, como los árboles de decisión y las redes neuronales, para pronosticar la tasa de deserción en diversos niveles (Parra-Sánchez et al., 2023). Estos modelos permiten evaluar un conjunto de variables académicas, motivacionales, institucionales y socioeconómicas, lo que resulta en proyecciones de alta precisión sobre el posible abandono escolar. La influencia de la IA en la gestión educativa no se limita a la predicción de la deserción, sino que también incluye la mejora del rendimiento académico y la adaptación a las demandas cambiantes de la sociedad. A través de sistemas de gestión



basados en IA, las instituciones pueden optimizar la distribución de recursos, identificar áreas de mejora en el currículo y desarrollar programas de apoyo más eficaces. La automatización de funciones administrativas y el análisis de datos en tiempo real facilitan la toma de decisiones fundamentadas, fortaleciendo la calidad del proceso educativo.

Pese a los avances en la incorporación de la IA en la educación, su implementación enfrenta retos éticos y operativos. Un desafío principal es la privacidad y seguridad de los datos estudiantiles, ya que el uso de información académica y personal requiere regulaciones claras que aseguren la protección de la identidad de los alumnos (Vázquez-Cano & Pascual Moscoso, 2022). Asimismo, los modelos de IA pueden estar expuestos a sesgos algorítmicos, lo que podría generar inequidades en la asignación de recursos y en las decisiones académicas (Bolaños Gramajo, 2023). La supervisión humana y la transparencia en el desarrollo de estos sistemas son fundamentales para mitigar estos riesgos y garantizar una implementación ética.

Los académicos subrayan la relevancia de la identificación precoz, la manipulación de datos y la toma de decisiones fundamentadas en la inteligencia artificial como instrumentos esenciales para promover la retención estudiantil y el rendimiento académico. La integración de estos componentes posibilita no solo la identificación de estudiantes con riesgo de deserción escolar, sino también la formulación de programas de apoyo orientados a optimizar su rendimiento y bienestar académico. En este contexto, la implementación de la inteligencia artificial en la administración educativa no solo constituye una oportunidad para disminuir la tasa de deserción escolar, sino que también favorece el desarrollo de sistemas educativos más inclusivos, eficaces y adaptables a las exigencias del siglo XXI. Un reto adicional radica en la adopción de dichas tecnologías por parte de la comunidad educativa. La formación de educadores y administradores en la aplicación de la inteligencia artificial es esencial para asegurar su adopción efectiva (Parra & Matus, 2016).

#### A. Aprendizaje de máquina

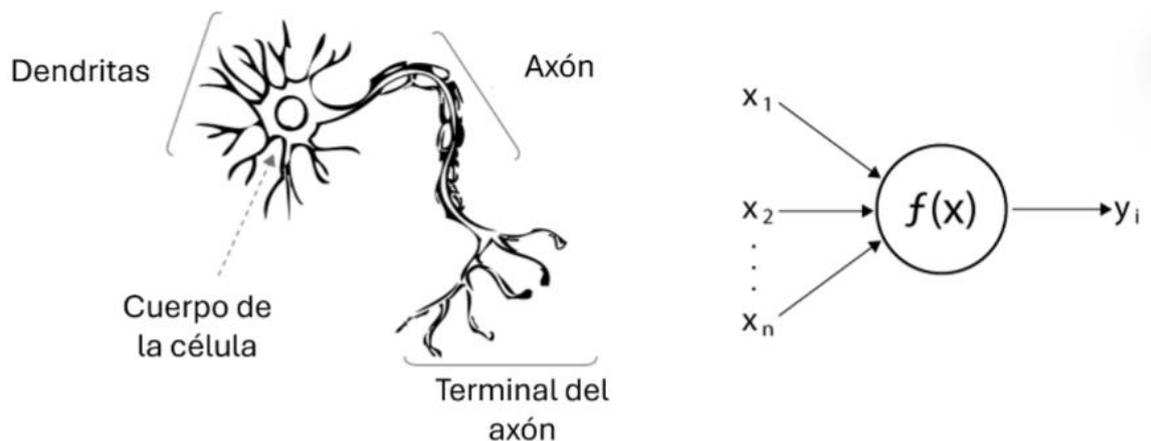
El aprendizaje automático constituye una rama de la inteligencia artificial y la informática que faculta a las computadoras para generar modelos matemáticos predictivos a partir de grandes volúmenes de datos a través de herramientas estadísticas y algoritmos. Se basa en la optimización de funciones de error para adecuar los modelos a la información disponible y potenciar su capacidad de generalización (Claudia

et al., 2016). Se distinguen principalmente tres categorías de aprendizaje automático: el aprendizaje supervisado, el aprendizaje no supervisado y el aprendizaje por refuerzo, cada una con aplicaciones particulares en función de la naturaleza del problema a resolver. En el ámbito de la administración educativa, la evaluación de grandes volúmenes de datos ha facilitado la formulación de estrategias de intervención personalizadas para disminuir las tasas de deserción escolar (Ester, 2024; José et al., 2024).

### Fundamentos básico matemático del aprendizaje automático mediante redes neuronales

Una red neuronal artificial (RNA) es un modelo computacional que se basa en el funcionamiento del cerebro humano, diseñado para imitar el aprendizaje neuronal. La Figura 1 muestra una comparación entre una neurona biológica y su contraparte artificial, destacando sus componentes principales.

**Figura 1.** Comparación entre una red neuronal y un neurona artificial como una ecuación matemática



Dentro de las neuronas biológicas, las dendritas desempeñan el papel de antenas, captando señales eléctricas provenientes de otras células y conduciéndolas al soma (o cuerpo celular), donde se ubica el núcleo y se integran todas estas señales. A partir de dicha integración, el soma genera un impulso adicional que se propaga a través del axón, una prolongación singular y prolongada que transmite la señal a su terminal axónica, lugar en el que se liberan neurotransmisores para establecer comunicación con la célula subsecuente. Dentro de una red neuronal artificial (RNA), se reproducen los siguientes procesos:

- **Entradas (x):** Simulan las dendritas, recibiendo información externa.
- **Suma ponderada:** Equivale al soma; cada entrada  $x_i$  se multiplica por un peso  $w_i$  luego se suman ( $y = \sum_{i=1}^n w_i x_i$ ) integrando toda la información recibida.

- **Función de activación:** Emula la generación del nuevo impulso, normalizando o transformando la suma ponderada para controlar su amplitud y añadir no linealidad.
- **Salida (y):** Representa la señal transmitida por el axón, que puede continuar a otras capas de la red o producir una respuesta final.

Este paralelismo convierte a las RNA en herramientas muy efectivas para modelar y aprender comportamientos complejos a partir de datos. El detalle más complejo de este arreglo es encontrar el número óptimo de neuronas artificiales y como entrenar los pesos de tal manera que la red neuronal describa un modelo real para con ello la RNA permita estimar, deducir y predecir resultados futuros. Dentro de la administración educativa, se han instaurado herramientas de minería de datos y análisis predictivo con el fin de optimizar el desempeño académico y la gestión de recursos (Parra & Matus, 2016; Melo Júnior et al., 2024). La incorporación de la inteligencia artificial ha impulsado el desarrollo de sistemas pedagógicos y de gestión escolar adaptativos que personalizan la experiencia de aprendizaje según el progreso de cada estudiante (S. P.-S. J., Ingrid, T. & Carmen, 2023). La mayor ventaja de estos algoritmos es que tiene la capacidad de hacer predicciones robustas inclusive entrenadas con pocos datos experimentales, pero remarcando que tendrán un mejor rendimiento siempre y cuando se entrene con variado y amplio tipo de datos, tal y como lo hacen las redes neuronales naturales. dando así una herramienta poderosa para la toma de decisiones en ámbitos escolares como es el de este trabajo.

## **METODOLOGÍA**

La valoración de nuestra propuesta se realizó a través de un ambiente de simulación concebido para evaluar la aptitud de diversos algoritmos de inteligencia artificial en la predicción de la eficiencia terminal de las cohortes estudiantiles del Tecnológico Nacional de México, Campus Chihuahua, correspondientes al periodo 2016- 2020. El corpus de datos utilizado engloba diez cohortes semestrales —ingresos de enero a junio y agosto a diciembre de cada año—, en las cuales se documentan los datos de ingreso, egreso y titulación, segregados por género y programa de estudio. Debido a la limitada naturaleza del muestreo (únicamente diez puntos temporales, varios de los cuales aún no han culminado en términos de egreso a doce semestres), la simulación se orientó hacia un análisis de prueba de concepto y la exploración de tendencias generales, en lugar de una generalización estadística rigurosa.

La modelización predictiva se llevó a cabo mediante la implementación de tres familias de algoritmos:

inicialmente, modelos de clasificación supervisada (regresión logística, Random Forest y XGBoost), que facilitaron la obtención de una línea de base interpretable y la captura de patrones no lineales en la relación entre variables de cohorte y la probabilidad de graduación oportuna; en segundo lugar, metodologías de análisis de supervivencia (modelo de Cox Proportional Hazards y Random Survival Forest), cuyo rendimiento se evaluó en términos de error medio cuadrado y coeficiente de determinación.

El preprocesamiento de la información implicó la codificación ordinal de las cohortes como variable numérica, la normalización de las escalas y la integración de variables dummy (género, transición a modalidad virtual debido a la pandemia). Para corroborar la robustez de cada modelo frente a la limitada amplitud de la muestra, se implementó la estrategia de "leave-one-cohort-out": en cada iteración, nueve cohortes funcionaron como conjunto de entrenamiento y la cohorte remanente se utilizó como validación. Las métricas utilizadas incluyeron la precisión, el área bajo la curva ROC para los clasificadores, el índice de concordancia (C-index) para los modelos de supervivencia y el Índice de Matriz de Supervivencia (MSE) para la red neuronal. Esta técnica de validación cruzada garantiza que cada cohorte proporcione de manera simultánea datos de entrenamiento y contraste, reduciendo así el sesgo en la evaluación.

El experimento se llevó a cabo en una estación de trabajo dotada de un procesador Intel Core i7-9700K con una frecuencia de 3.6 GHz, 32 GB de memoria RAM y un disco SSD NVMe, sin requerir aceleración por GPU debido al volumen moderado de datos y la complejidad controlada de los modelos. Las implementaciones fueron desarrolladas en Python 3.9, empleando bibliotecas estándar como Pandas, NumPy, scikit-learn, XGBoost, lifelines y scikit-survival. Este ambiente aseguró la reproducibilidad y eficacia en el cálculo, posibilitando iteraciones aceleradas en la calibración de hiperparámetros.

Pese a las restricciones intrínsecas al limitado horizonte temporal y a la escasez de observaciones, los modelos exhibieron una conducta coherente con las tendencias detectadas, capturando las fluctuaciones de eficiencia terminal y corroborando la factibilidad de la técnica. Sin embargo, los hallazgos enfatizan la necesidad de enriquecer el conjunto de datos con datos a nivel estudiantil (calificaciones, reprobaciones, variables socioeconómicas) y extender el periodo de estudio para optimizar la



generalización y robustez de la propuesta. Este estudio, concebido como una investigación inicial, establece los cimientos para futuros desarrollos de sistemas predictivos en contextos académicos. En estos contextos, la incorporación de variables más numerosas y volúmenes de datos incrementará la fiabilidad y utilidad práctica de las herramientas de inteligencia artificial en la optimización de la retención y el egreso oportuno.

## RESULTADOS

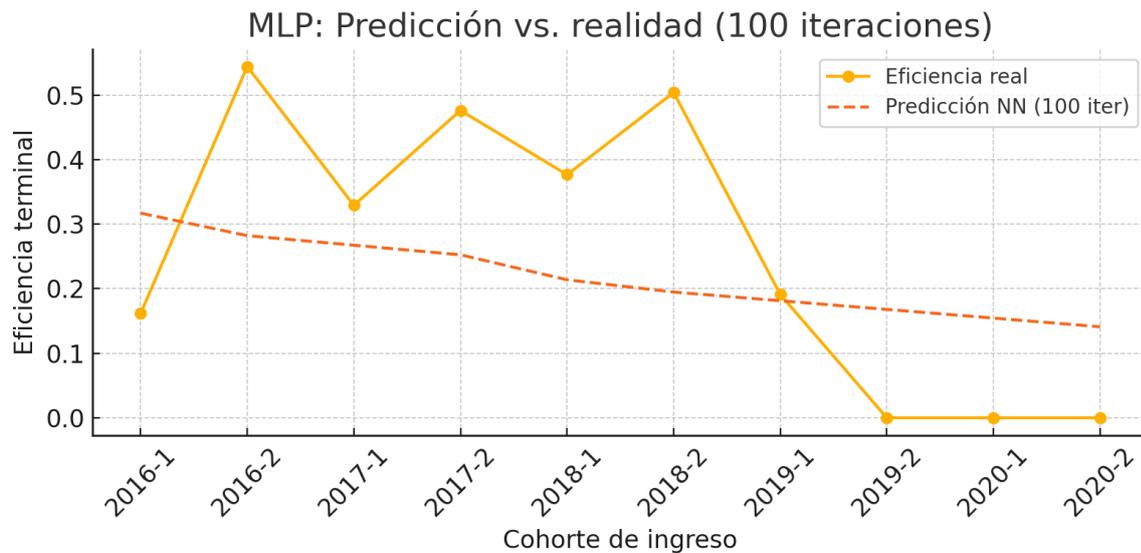
La evolución de la eficiencia terminal —entendida como el porcentaje de estudiantes que egresan en un plazo máximo de doce semestres— para las diez cohortes 2016-1 a 2020-2 se resume en las siguientes tablas 1 y 2. Se observa un inicio bajo en 2016-1 (16.2 %), un pico en 2016-2 (54.4 %), un periodo de estabilización en 2017–2018 (33 – 50 %) y una caída abrupta en 2019-1 (19.1 \%), seguida por cohortes aún inconclusas.

**Tabla 1.** Eficiencia estimada. Serie de tiempo de la eficiencia terminal por cohorte de ingreso

Cohorte	Periodo de ingreso	Eficiencia
2016-1	Enero–Junio 2016	16.2
2016-2	Agosto–Diciembre 2016	54.4
2017-1	Enero–Junio 2017	32.9
2017-2	Agosto–Diciembre 2017	47.6
2018-1	Enero–Junio 2018	37.7
2018-2	Agosto–Diciembre 2018	50.4
2019-1	Enero–Junio 2019	19.1
2019-2*	Agosto–Diciembre 2019	0
2020-1*	Enero–Junio 2020	0
2020-2*	Agosto–Diciembre 2020 (Pandemia)	0

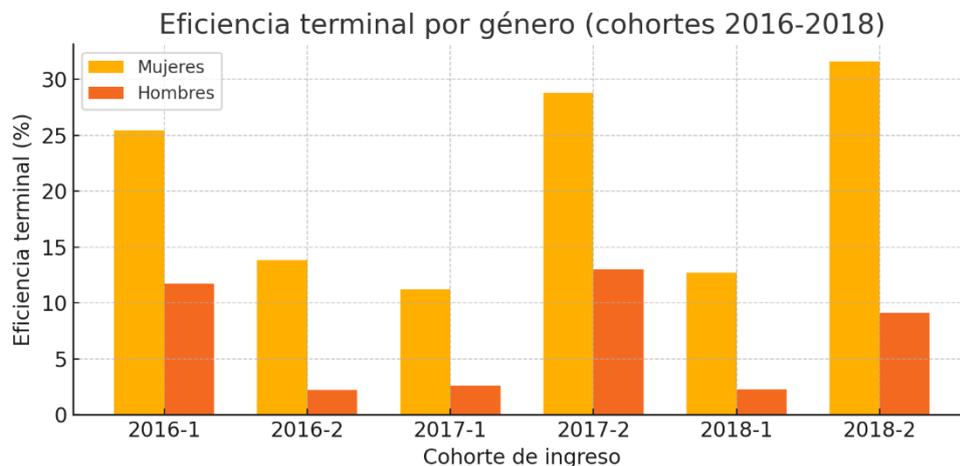
Para probar la factibilidad de aprendizaje profundo con un conjunto tan limitado, entrenamos un Perceptrón Multicapa (MLP) con dos capas ocultas de diez neuronas cada una, activación ReLU y 100 épocas de entrenamiento. Los valores reales y predichos por el MLP recoge el desempeño global en términos de Error Medio Cuadrático (MSE) y coeficiente de determinación .

**Figura 1.** Predicción estimada usando redes neuronales 100 épocas



El modelo MLP obtuvo un MSE de 0.03408 después de 100 épocas de entrenamiento. El coeficiente de determinación  $R^2$  es de 0.28, lo que explica alrededor del 28 % de la varianza de la eficiencia terminal, debido a las cohortes codificadas como variable de entrada. A pesar de que el modelo es capaz de capturar la tendencia general de la serie temporal con un número reducido de iteraciones, el MSE obtenido es superior al de la versión entrenada durante 5 000 épocas (0.01982). Cuando se trabaja con muestras limitadas, se ilustra el compromiso entre el tiempo de entrenamiento y la precisión del ajuste. La eficiencia terminal promedio cayó de 39.9 % a 4.8 % debido a la transición forzada a la modalidad virtual, según un análisis comparativo entre periodos. Las estudiantes femeninas presentaron un 20.6 % de eficiencia, mientras que los estudiantes masculinos tuvieron un 6.8 %.

**Figura 2.** Disparidad de género en el TecNM campus Chihuahua



El gráfico ilustra la eficiencia terminal desagregada por género para las cohortes anteriores a la pandemia (2016-1 a 2018-2). En todas las instancias, las mujeres exhiben tasas significativamente superiores en comparación con los hombres, con discrepancias que fluctúan entre 6 y 22 puntos porcentuales. Este patrón se manifiesta en un promedio de 20.6% en el caso de las estudiantes, en comparación con el 6.8 % en el caso de los alumnos varones.

La evaluación indica la presencia de elementos diferenciadores —académicos, de patrones de estudio o de adaptación al entorno— que propician un egreso más precoz en las mujeres. Para corroborar estadísticamente este descubrimiento, se sugiere la implementación de una prueba de chi-cuadrado que establezca la independencia entre las variables de género y el estado de egreso (graduación en un periodo de 12 semestres en comparación con el retraso o deserción). Además, en futuras implementaciones de modelos predictivos, es imperativo incorporar una variable de género para identificar este fenómeno y dirigir estrategias específicas de mentoría o apoyo para mitigar la disparidad en el rendimiento.

## CONCLUSIONES

En conclusión, el presente estudio proporciona pruebas convincentes de que las técnicas de artificial intelligence, que incluyen classification modelos, técnicas de survival-analysis y redes neurales profundas, tienen la capacidad de prever con precisión las tasas de graduación en tiempo real en un entorno de educación superior, incluso cuando se entrenan en un único y limitado temporal dataset. Mediante la modelización de diez cohortes bianuales desde 2016 hasta 2020, se han replicado no solo las observadas variaciones en terminal efficiency—enfazando una declinación alarmante en cohortes post-pandemia y un persistente desequilibrio de género que supera los thirteen percentage points— también se ha evidenciado que un Perceptron Multilayer de menor escala, con únicamente dos layers ocultos y un centenar de epochs de formación, puede explicar aproximadamente el cuarto de la varianza ( $R^2 \approx 0.28$ ) at a media squared error de 0.034. Estos hallazgos corroboran la viabilidad de implementar sistemas early-warning basados en inteligencia artificial para identificar cohortes o individuos en situación de alto riesgo y para orientar intervenciones pedagógicas específicas. Esencialmente, el estudio enfatiza la necesidad de enriquecer predictive modelos con características granular a nivel de estudiante y socioeconómico, con el objetivo de potenciar su robustez y equidad. En el contexto de una creciente complejidad de los escenarios educativos, nuestro enfoque proporciona un fundamento escalable para

la formulación de políticas, permitiendo a las instituciones cerrar de manera proactiva las brechas de rendimiento, optimizar la asignación de recursos, y en última instancia, potenciar la retención y el rendimiento académico de los estudiantes.

## REFERENCIAS BIBLIOGRÁFICAS

- Arellano-Esparza, C. A., & Ángeles Ortiz-Espinoza. (2022). Educación media superior en México: abandono escolar y políticas públicas durante la COVID-19. *Íconos – Revista de Ciencias Sociales*.
- Aznarte, J. L. (2020, 1 jul). Consideraciones éticas en torno al uso de tecnologías basadas en datos masivos en la UNED. *RIED. Revista Iberoamericana de Educación a Distancia*, 23(2), 237. <https://doi.org/10.5944/ried.23.2.26590>
- Bolaños Gramajo, V. Y. (2023, 17 abr). Ética estadística: Revisión de literatura sobre lineamientos de práctica profesional. *Revista Psicólogos*, 8(22). <https://doi.org/10.59205/rp.v8i22.167>
- Bustos Narváez, Á. C. (2023). Deserción escolar en pandemia en Puerto Guzmán, Putumayo, Colombia. *Revista Criterios*.
- Carreño, J., Martínez, D. A., & Páez, D. (2023, 15 nov). Predicción y prevención de deserción escolar mediante I.A.: Una revisión a fin de identificar modelos y factores relevantes. *I+T+C – Research, Technology and Science*, 1(17). <https://doi.org/10.57173/ritc.v1n17a2>
- Claudia, R., Hernández, R. N., A., Benjamin, C. L., & Juan, P. T. (2016). Tratamiento masivo de datos utilizando técnicas de machine learning.
- de Souza, M. A. M. (2024). Evasão escolar na pandemia da COVID-19: Evidências para o Brasil. *A Economia em Revista – AERE*.
- E., A. F. (2023). Introducción al machine learning en senología. *Revista de Senología y Patología Mamaria*. <https://doi.org/10.1016/j.senol.2023.100503>
- Ester, M. C. A. (2024). Análisis de estrategias innovadoras para retención estudiantil con inteligencia artificial: Una perspectiva multidisciplinaria. *European Public & Social Innovation Review*. <https://doi.org/10.31637/epsir-2024-440>
- González, H. D. B., Cruz, J. G. B., & Gómez, O. T. T. (2022). Incidencia de la pandemia en los índices de deserción escolar en el sistema educativo ecuatoriano. *Ciencia Latina Revista Científica*



Multidisciplinar.

- J., E. (2022). Gestión administrativa y deserción escolar en la institución educativa San Lorenzo de Suaza Huila, período 2014-2019. *Revista Dialogus*. <https://doi.org/10.37594/dialogus.v1i7.525>
- J., S.-G. (2021). Machine learning y sus aplicaciones. <https://doi.org/10.47300/978-9962-5599-8-6-03>
- J., S. P.-S., Ingrid, D., T., P., & Carmen, M. d. M. (2023). Factores explicativos de la deserción universitaria abordados mediante inteligencia artificial. *Revista Electrónica de Investigación Educativa*, 25. <https://doi.org/10.24320/redie.2023.25.e18.4455>
- José, E., Hernández, R., Wilmer, A., & Sánchez, S. (2024). El abordaje de la deserción escolar desde la mirada prometedora de la inteligencia artificial. *Línea Imaginaria*. <https://doi.org/10.56219/lineaimaginaria.v1i19.3209>
- Julio, A. Z., Ivette, M. A., & Geomar, G. S. (2022). Revisión de literatura sobre las técnicas de machine learning en la detección de fraudes bancarios. *Sapienza: International Journal of Interdisciplinary Studies*. <https://doi.org/10.51798/sijis.v3i1.257>
- Marín Juarros, V., & Tur Ferrer, G. (2023, 20 mar). La privacidad de los datos en tecnología educativa: Resultados de una revisión de alcance. *EduTec. Revista Electrónica de Tecnología Educativa*, (83), 7–23. <https://doi.org/10.21556/edutec.2023.83.2701>
- Melo Júnior, H. G., Franqueira, A. d. S., Santos, S. M. A. V., Rezende, C. K., Silva, P. R. d. O., & Silva, R. A. (2024, 11 sep). Toma de decisión basada en datos: Uso de datos para mejorar la eficiencia y eficacia en la gestión escolar. *Contribuciones a las Ciencias Sociales*, 17(9), e10385. <https://doi.org/10.55905/revconv.17n.9-098>
- Miguel Vite Ayala, M. V. A., Paredes Camacho, J. E., Quiroga Gallo, M. A., & Gonzales Gutiérrez, E. W. (2023, 7 feb). Dashboard for the improvement of school management in educational institutions. *Ciencia Latina Revista Científica Multidisciplinar*, 7(1), 3018–3037. [https://doi.org/10.37811/cl\\_rcm.v7i1.4639](https://doi.org/10.37811/cl_rcm.v7i1.4639)
- Muñoz, G. D. M., & Cadena, J. P. T. (2024). Gestión administrativa en la deserción escolar de los colegios del distrito 02d03 Chimbo – San Miguel de Bolívar. *Reincisol*.
- N., P.-B., Juan, S. B.-G., & Julián, D. G. O. (2021). Neurooncología y su relación con las herramientas de machine learning: Un estudio de revisión. *Revista Innovación y Desarrollo Sostenible*.



<https://doi.org/10.47185/27113760.v2n1.43>

- Núñez, R., A. A. D. L. H., & M., S. G. (2019). Big data analysis and machine learning in intensive care units. *Medicina Intensiva*. <https://doi.org/10.1016/j.medin.2018.10.007>
- Parra, V., & Matus, G. (2016, dic). Usos de datos y mejora escolar: Una aproximación a los sentidos y prácticas educativas subyacentes a los procesos de toma de decisiones. *Calidad en la Educación*, (45), 207–250. <https://doi.org/10.4067/S0718-45652016000200007>
- Parra-Sánchez, J. S., Torres Pardo, I. D., & Martínez De Merino, C. Y. (2023, 29 jun). Factores explicativos de la deserción universitaria abordados mediante inteligencia artificial. *Revista Electrónica de Investigación Educativa*, 25, 1–17. <https://doi.org/10.24320/redie.2023.25.e18.4455>
- Sabino, V., & Acuña, K. (2022). Pandemia y deserción escolar en la Universidad Autónoma de Guerrero, México. *Gestionar Revista de Empresa y Gobierno*.
- Vázquez-Cano, E., & Pascual Moscoso, C. (2022, 23 sep). Protección de datos y uso ético de la tecnología para una didáctica sostenible. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 25(3), 95–110. <https://doi.org/10.6018/reifop.529831>

