



UNIVERSIDAD PERUNA CAYETANO HEREDIA

MÉTODOS NUMÉRICOS Y OPTIMIZACIÓN PARA MACHINE LEARNING

INFORME

“Predicción de deserción de estudiantes universitarios”



Fuente: de Prensa, O. (s/f). *Tasa de deserción en educación universitaria.*

INTEGRANTES:

- Aybar Escobar Edithson Ricardo
- Llanos Angeles Leily Marlith
- Luque Mamani Magno Ricardo
- Mendoza Villar Antony Iván

DOCENTES:

- Castro Zarate Nelson Enrique
- Mauricio Salazar Josue Angel

Lima-Perú

2024

1. Introducción:

La deserción estudiantil en la educación superior es un problema crítico con implicancias profundas, tanto para los estudiantes como para las instituciones educativas y la sociedad en general. El abandono de los estudios no solo afecta el futuro académico y profesional de quienes desertan, reduciendo sus oportunidades laborales y personales, sino que también implica un uso ineficiente de los recursos educativos y presenta desafíos en la planificación institucional. Diversos estudios han identificado factores que contribuyen a esta problemática, como las características sociodemográficas, el rendimiento académico, las condiciones familiares y la situación económica (Gallego, J. C. G. 2021; Santos y Oliveira, 2019).

Ante este desafío, surge la pregunta ¿qué acciones podemos tomar para mitigar la deserción estudiantil? En la actualidad, los avances tecnológicos han proporcionado diversas herramientas para analizar grandes volúmenes de datos, y el machine learning se ha destacado como una solución eficaz para identificar patrones en los datos que podrían pasar desapercibidos con métodos tradicionales (Peña y Pérez, 2019; Maluenda-Albornoz et al., 2022). Un modelo de machine learning puede ayudar a predecir qué estudiantes están en riesgo de abandonar sus estudios, permitiendo así que las instituciones tomen medidas preventivas.

No obstante, muchos de los modelos actuales se construyen utilizando frameworks como TensorFlow o Scikit-learn, que si bien simplifican el proceso, tienden a abstraer los detalles técnicos subyacentes (Rivas et al., 2019; Peixoto & Silveira, 2020). En este trabajo se propone un enfoque diferente: desarrollar un modelo de predicción de la deserción estudiantil sin el uso de frameworks preexistentes. Este enfoque ofrece una mayor flexibilidad y control sobre cada fase del proceso, permitiendo personalizar el modelo según las características específicas del problema y los datos disponibles (Fernández & Ortega, 2022). Además, construir el modelo desde cero proporciona una comprensión más profunda de las etapas críticas del proceso de predicción, tales como la selección de características, el ajuste de parámetros y la validación de resultados (Peña & Pérez, 2019).

El conjunto de datos utilizado en este estudio incluye una amplia variedad de características estudiantiles, tales como estado civil, tipo de aplicación a la universidad, número de solicitudes enviadas, curso al que se inscribe el estudiante, y calificaciones de admisión. También se consideran factores económicos, como la tasa de desempleo, la inflación y el

Producto Interno Bruto (PIB) del país, los cuales influyen significativamente en el rendimiento académico (Santos & Oliveira, 2019; Fernández & Ortega, 2022). La diversidad de estas variables permite una visión integral de los elementos que pueden influir en la decisión de un estudiante de abandonar sus estudios.

2. Estado del Arte

La deserción estudiantil universitaria es un problema creciente en Latinoamérica. Las altas tasas de abandono representan un gran desafío para las instituciones educativas, ya que afectan tanto la continuidad académica de los estudiantes como el desarrollo del capital humano en la región (Pérez A. M. et al., 2018). En respuesta a esta problemática, se han desarrollado modelos predictivos que utilizan datos históricos y características individuales para identificar factores de riesgo de deserción, optimizar recursos y apoyar la toma de decisiones sobre estrategias de retención (Pérez A. M. et al., 2018).

2.1. Revisión de Literatura y Estudios Previos

Para predecir la deserción escolar, se han utilizado técnicas de aprendizaje automático como la regresión logística, los árboles de decisión, las redes neuronales y el K-Nearest Neighbors (KNN). Universidades como la Bernardo O'Higgins y la Simón Bolívar han reportado alta precisión con la regresión logística (86.4%) (Pérez A. M. et al., 2018). Los árboles de decisión, como el algoritmo J48, han demostrado ser aún más eficientes para clasificar estudiantes (Amaya Torrado Y. K. et al., 2024).

Aunque las variables académicas y socioeconómicas, como el promedio de notas y los años cursados, son factores clave para predecir la deserción, falta discusión sobre cómo estos modelos se comportan con diferentes calidades de datos y contextos. Las redes neuronales en la Universidad de Las Américas mostraron problemas debido a la baja calidad de los datos, subrayando la importancia de bases de datos completas (Fischer Angulo E. S., 2012). Los modelos deben seleccionarse con cuidado, considerando su capacidad para manejar datos incompletos y adaptarse al contexto.

Los modelos predictivos cuantifican la deserción, pero el apoyo emocional y social es igualmente fundamental. Estudios como el de Ramírez y Grandón (2018) muestran que el apoyo social y el asesoramiento psicológico mejoran el sentido de pertenencia y reducen el estrés, favoreciendo la retención. Sin embargo, muchos modelos no incluyen estos factores,

lo que limita su eficacia para ofrecer soluciones completas. La incorporación de factores psicosociales sigue siendo una laguna en los estudios sobre deserción.

2.2. Comparación de Enfoques Metodológicos

Artículo	Desequilibrio de Clases	Preprocesamiento de Datos	Selección de Características	Métricas de Evaluación
Pérez, A. M., et al. (2018)	Sobremuestreo y submuestreo	Limpieza, normalización, imputación de valores faltantes	Análisis bivariado y multivariado	Precisión, sensibilidad, especificidad, AUC
Amaya Torrado, Y. K., et al., (2024)	Sobremuestreo de clase minoritaria	Normalización, imputación de valores faltantes	Ganancia de información	Precisión, sensibilidad, especificidad, AUC
Fischer Angulo, E. S. (2012)	Muestreo para balancear clases	Normalización, imputación de valores faltantes	Métodos estadísticos y de correlación	Precisión, sensibilidad, especificidad
Ramírez, P. E., & Grandón, E. E. (2018).	Sobremuestreo	Limpieza, eliminación de outliers, normalización	Análisis de importancia basado en árboles de decisión	Precisión, sensibilidad, especificidad, AUC
Cruz, E., González, M., & Rangel, J. C. (2022)	Sobremuestreo y submuestreo	Normalización, imputación de valores faltantes, codificación	Análisis de correlación, selección automática	Precisión, recall, F1-score, AUC
Gudivada, Apon y Ding (2017)	Sobremuestreo y submuestreo	Limpieza, eliminación de duplicados,	Métodos de filtrado y envoltura	Precisión, recall, F1-score, AUC

		normalización		
--	--	---------------	--	--

La mayoría de los estudios utilizan sobremuestreo y submuestreo para corregir el desequilibrio de clases, ya que los estudiantes en riesgo de deserción son una minoría. Aunque estas técnicas mejoran la precisión, también pueden introducir sesgos por la duplicación de datos (Pérez A. M. et al., 2018). Estudios más recientes, como los de Amaya Torrado Y. K. et al. (2024), combinan sobremuestreo con imputación de valores faltantes, lo cual mejora tanto la fiabilidad como la precisión. El preprocesamiento de datos con limpieza, normalización y manejo de valores faltantes también es clave, como se muestra en la Universidad de Las Américas y la UTN Regional La Plata (Fischer Angulo E. S., 2012). Mejorar estas técnicas reduciría la dependencia de la calidad de los datos.

La selección de características es clave. Algunos estudios usan técnicas de filtrado y envoltura, mientras que otros, como Valero Cajahuanca J. E. et al. (2022), emplean árboles de decisión para identificar las variables más influyentes. Incorporar factores no académicos, como la motivación o el entorno social, podría mejorar la precisión de los modelos predictivos. En cuanto a las métricas de evaluación, la precisión, sensibilidad y AUC son comunes, pero el F1-score también es importante, ya que combina precisión y recall, lo cual es crucial para evitar falsos negativos y mejorar la identificación de estudiantes en riesgo.

2.3. Identificación de Lagunas en el Conocimiento

Una de las principales limitaciones es la calidad de los datos. La falta de bases de datos bien estructuradas y completas afecta el rendimiento de los modelos predictivos, como lo señala Fischer Angulo (2012). Además, la mayoría de los estudios se enfocan en predecir la deserción en las primeras etapas del ciclo académico, sin considerar adecuadamente las fases avanzadas, donde la carga académica y las oportunidades laborales adquieren mayor relevancia. Muchos estudiantes abandonan en los últimos años debido a estos factores, que no son suficientemente contemplados en los modelos actuales.

El análisis de factores psicológicos y motivacionales también es clave. Aunque difíciles de medir, son esenciales para comprender el abandono. La motivación intrínseca y el apoyo institucional son factores fundamentales para la retención, pero muchos modelos no los consideran, limitando la eficacia de las intervenciones. Finalmente, los modelos se centran en estudiantes en riesgo, sin abordar intervenciones preventivas. Un enfoque integral que

considere el entorno socioeconómico y cultural podría mejorar la efectividad de los modelos predictivos.

2.4. Resumen de Teorías o Modelos Aplicables

Entre las teorías más relevantes para la predicción de la deserción estudiantil se encuentra el modelo de integración académica y social de Tinto. Este modelo sugiere que los estudiantes tienen mayor probabilidad de abandonar cuando no se integran adecuadamente en la comunidad académica, destacando la importancia de la integración social y académica para la retención. Esta teoría podría beneficiarse de los avances en machine learning para desarrollar intervenciones más personalizadas. Los modelos de machine learning, como los árboles de decisión y la regresión logística, siguen siendo útiles por su interpretabilidad y efectividad en la predicción de la deserción. Sin embargo, los modelos más complejos, como las redes neuronales, aunque más precisos, requieren bases de datos más robustas, lo cual limita su aplicación en muchas instituciones.

3. Metodología

La metodología de este trabajo se basa en la realización de varias etapas incluyendo la obtención de datos de estudiantes universitarios, el preprocesamiento de datos, y la aplicación de modelos de machine learning. Cada una de estas nos permitió, mediante estrategias que se explican mejor más adelante, llegar a los resultados esperados. La idea detrás de este enfoque es tener un proceso ordenado que nos lleve a cumplir con los objetivos de forma eficiente. A continuación, se muestra los pasos:

3.1. Elección de la metodología para la extracción de información

En el contexto de los datos en Big Data y Machine Learning, Gudivada, Apon y Ding (2017) mencionan lo importante que es tener en cuenta la calidad de los datos, ya que estos impactan directamente en el rendimiento de los modelos. Por eso, tener una buena metodología que nos permita llevar a cabo un proceso adecuado para la extracción de datos es clave para una fase inicial. De esta manera, se comparó tres de las metodologías más usadas: KDD, SEMMA y CRISP-DM, las cuales mantienen relación entre ellas, para el proceso de entendimiento y/o análisis de nuestros datos.

Conocidas las diferencias entre los procesos de análisis y preparación de datos, en este trabajo, y con algunas modificaciones, nos basamos en el modelo CRISP-DM por su

simplicidad y porque se adecúa mejor a nuestros objetivos de desarrollar un modelo de predicción.

3.2. Descripción del conjunto de datos

Basándonos en las fases de la metodología elegida, iniciamos con la comprensión del conjunto de datos utilizado. Los datos que se utilizaron fueron obtenidos del repositorio de datos de la Universidad de California en Irvine. Estos datos son de estudiantes de pregrado en carreras como agronomía, diseño, educación, enfermería, periodismo, administración, servicio social y tecnologías. Incluyen una variedad de características sociodemográficas, académicas y económicas de los estudiantes. Estas características se recopilaron de fuentes institucionales y externas, como:

- Estado civil.
- Ocupación de los padres.
- Curso al que se inscribió el estudiante.
- Factores económicos (tasa de desempleo, inflación, PIB).

3.3. Preprocesamiento de los datos

Para preparar los datos, se realizaron tareas de normalización para asegurar que todas las características estuvieran a un rango específico. Como todos los registros tenían valores completos, no fue necesario aplicar imputación de valores faltantes. También evaluamos la conversión de variables categóricas a numéricas.

Al analizar los datos, se pudo notar que varias de las características eran numéricas, pero representaban categorías, lo que las clasifica como variables categóricas nominales. Por lo tanto, fue necesario seleccionar aquellas columnas de este tipo para verificar la asociación con nuestra característica de interés, denominada 'Target'.

3.4. Selección de características

La selección de características es un paso crucial en el desarrollo de modelos de machine learning, ya que permite identificar y retener las variables más relevantes que influyen en la deserción estudiantil. En este estudio, se aplicó la prueba de Chi-cuadrado, la cual mostró resultados de alta relación entre las variables. Para medir la fuerza de asociación, se utilizó la medida estadística Cramer's V.

En consecuencia, se eligieron las características de tipo categórico nominal con las que trabajaremos. Para evaluar la asociación entre los valores de interés y las variables numéricas, se aplicó la normalización. Además, para verificar la asociación, se utilizaron la prueba de Kruskal-Wallis y la prueba ANOVA. Como resultado, se seleccionaron las columnas con las que trabajaremos en la siguiente etapa.

3.5. Definición de modelos conocidos

Los modelos de predicción que se utilizan comúnmente para el análisis de deserción estudiantil incluyen una variedad de técnicas de machine learning, cada una con sus ventajas y desventajas. En primer lugar, la Regresión Logística es ampliamente utilizada debido a su simplicidad y capacidad para interpretar las variables influyentes en la deserción (Song et al., 2023). Otro modelo popular es el de Árboles de Decisión, que permite visualizar claramente el proceso de decisión, siendo especialmente útil cuando se requieren explicaciones claras sobre por qué se produce la deserción (Segura et al., 2022).

Por otro lado, técnicas más avanzadas como Gradient Boosting y XGBoost han demostrado ser altamente eficaces, ya que corrigen errores iterativos de modelos más simples y son particularmente útiles cuando se trata de conjuntos de datos desbalanceados (Villar et al., 2024). Además, las Redes Neuronales Convolucionales (CNN) han sido aplicadas en estudios recientes, donde se utilizaron para predecir la deserción de estudiantes universitarios al analizar datos de comportamiento y características de manera más eficiente (Mbunge et al., 2022).

3.6. Evaluación de modelos conocidos

Después de entrenar estos tres modelos, evaluamos sus rendimientos utilizando métricas como accuracy, precisión, recall, y F1-score. Utilizamos validación cruzada para garantizar que los resultados sean consistentes y no presenten sobreajuste de datos de entrenamiento. Del mismo modo, comparamos el rendimiento de los modelos para identificar cuál se comporta mejor en la predicción de deserción universitaria.

3.7. Diseño de nuestro modelo

Una vez realizada la comparación, implementamos un modelo desde cero inspirado en los principios del modelo que mejores resultados se obtuvo. Esta implementación nos permitirá

entender los cálculos y optimizaciones internas de los modelos, como el uso del Gradiente Descendente en la regresión y/o capas convolucionales en las CNN.

3.8. Análisis y evaluación de nuestro modelo

La validación de los resultados de nuestro modelo es fundamental para asegurar su efectividad en la predicción de la deserción estudiantil. Utilizamos un conjunto de datos de prueba, calculando métricas como accuracy, precisión, recall y F1-score para evaluar su rendimiento, así como se hizo para los modelos ya conocidos. También realizamos un análisis de errores mediante una matriz de confusión, lo que nos permitirá identificar patrones en los casos mal clasificados y comprender mejor las características influyentes. Finalmente, discutiremos cómo estos hallazgos pueden traducirse en estrategias prácticas que ayuden a las instituciones educativas a intervenir de manera efectiva en la reducción de la deserción.

3.9. Herramientas

En este trabajo se utilizaron varios frameworks para el análisis y visualización de datos: NumPy se empleó para realizar operaciones, mientras que Pandas facilitó la manipulación y limpieza de datos mediante DataFrames. Scikit-learn permitió implementar algoritmos de aprendizaje automático, abarcando desde el preprocesamiento hasta la evaluación de modelos. Para la visualización de datos, se utilizó Seaborn para crear gráficos estadísticos atractivos y Matplotlib para realizar gráficos personalizados, lo que ayudó a ilustrar los resultados.

4. Experimentación y Resultados

- - - AVANCE - - -

Antes de adentrarnos en la comprensión de los datos, se compararon tres de las metodologías más comunes en la creación de modelos de predicción. El proceso de descubrimiento del conocimiento, señalado como Knowledge Discovery in Databases (KDD) por Fayyad, Piatetsky-Shapiro y Smyth (1996), se encarga de la preparación de los datos y la interpretación de los resultados obtenidos, los cuales dan un significado a patrones encontrados. Por otro lado, tenemos al proceso SEMMA que fue planteado por el Instituto SAS. SEMMA es el acrónimo de Sample, Explore, Modify, Model y Assess (Muestrear, Explorar, Modificar, Modelar y Evaluar) (Vanrell, Bertone, & García Martínez, 2010). Y por

último se tiene al proceso CRISP-DM, el cual se trata de una metodología utilizada para la minería de datos y aprendizaje automático (Chapman et al., 2000). A continuación, se puede ilustrar las diferencias entre estos modelos.

Figura – Cuadro comparativo.

Fases	KDD	CRISP – DM	SEMMA
<i>Análisis y comprensión del negocio</i>	Comprensión del dominio de aplicación	Comprensión del negocio	
<i>Selección y preparación de los datos</i>	Crear el conjunto de datos	Entendimiento de los datos	Muestreo
	Limpieza y pre-procesamiento de los datos	Preparación de los datos	Comprensión
	Reducción y proyección de los datos		Modificación
<i>Modelado</i>	Determinar la tarea de minería Determinar el algoritmo de minería Minería de datos	Modelado	Modelado
<i>Evaluación</i>	Interpretación	Evaluación	Valoración
<i>Implementación</i>	Utilización del nuevo conocimiento	Despliegue	

Fuente: Análisis comparativo de metodologías para la gestión de proyectos de minería de datos (Moine et al., 1996).

Tomando como referencia la metodología CRISP-DM, planteamos las siguientes etapas o fases para este trabajo:

- A. Descripción del conjunto de datos
- B. Preprocesamiento de los datos
- C. Selección de características
- D. Definición de modelos conocidos
- E. Evaluación de modelos conocidos
- F. Diseño de nuestro modelo
- G. Análisis y evaluación de nuestro modelo

A. Descripción del conjunto de datos

Los datos de los universitarios de distintas carreras profesionales fueron tomados del repositorio de aprendizaje automático de la Universidad de California, Irvine. Estos datos incluyen las siguientes variables: estado civil, modo de ingreso, orden de solicitud, curso, turno de asistencia, nivel de educación, calificación de la titulación previa, nacionalidad, nivel educativo de los padres, estudiante internacional, unidades curriculares del primer y segundo semestre, tasa de desempleo, tasa de inflación, el PBI y el objetivo. De estas, la variable más importante será el objetivo o "Target" en relación con las demás, ya que se desea crear un modelo de predicción de tres categorías (abandono, matriculado y graduado) con aquellas características que contribuyen a dicha predicción. Consideramos que también se debería tomar en cuenta el aspecto psicológico como un factor que influye en la toma de decisiones para la permanencia en la universidad.

REFERENCIAS:

1. **Amaya Torrado, Y. K., Avendaño, E. B., & Vizcaíno, D. J. H.** (2024). Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos. Universidad Francisco de Paula Santander, Ocaña, Colombia; Universidad Simón Bolívar, Barranquilla, Colombia. Recuperado de: <https://documentos.redclara.net/bitstream/10786/759/1/124-22-3-2014-Modelo%20predictivo%20de%20deserci%C3%B3n%20estudiantil%20utilizando%20t%C3%A9cnicas%20de%20miner%C3%ADa%20de%20datos.pdf>
2. **Cruz, E., González, M., & Rangel, J. C.** (2022). Técnicas de machine learning aplicadas a la evaluación del rendimiento y a la predicción de la deserción de estudiantes universitarios, una revisión. *Prisma Tecnológico*, 13(1), 77-87. <https://doi.org/10.33412/pri.v13.1.3039>
3. **Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P.** (1996). From data mining to knowledge discovery in databases. *AI Magazine*. Recuperado de: <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1230>
4. **Fernández, M. L., & Ortega, E.** (2022). Factores económicos y académicos en la deserción estudiantil universitaria: un análisis predictivo utilizando técnicas de machine learning. *Innovación Educativa*, 22(3), 65-82.
5. **Fischer Angulo, E. S.** (2012). Modelo para la automatización del proceso de determinación de riesgo de deserción en alumnos universitarios (Tesis de maestría, Universidad de Chile). Repositorio Académico Universidad de Chile. Recuperado de https://repositorio.uchile.cl/bitstream/handle/2250/111188/cf-fischer_ea.pdf?sequence=1
6. **Gallego, M. G., Perez de los Cobos, A. P., & Gallego, J. C. G.** (2021). Identifying students at risk to academic dropout in higher education. *Education Sciences*, 11(8), 427.

7. **Gudivada, V., Apon, A., & Ding, J.** (2017). Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations. *International Journal on Advances in Software*. Recuperado de: https://personales.upv.es/thinkmind/dl/journals/soft/soft_v10_n12_2017/soft_v10_n12_2017_1.pdf
8. **Kabathova, J., & Drlik, M.** (2021). Towards predicting student's dropout in university courses using different machine learning techniques. *Applied Sciences*. <https://doi.org/10.3390/app11073130>
9. **León, M. A., & Cano, S. E.** (2020). Development of dropout prediction model in higher education using machine learning techniques. *International Journal of Advanced Computer Science and Applications*, 11(8), 234-242.
10. **Maluenda-Albornoz, J., Infante-Villagrán, V., Galve-González, C., Flores-Oyarzo, G., & Berríos-Riquelme, J.** (2022). Early and dynamic socio-academic variables related to dropout intention: A predictive model made during the pandemic. *Sustainability*, 14(2), 831.
11. **Mbunge, E., Gurajena, C., Fashoto, S., Akinnuwesi, B., & Metfula, A.** (2022). Predicting student dropout in massive open online courses using deep learning models - A systematic review. In *Cybernetics Perspectives in Systems*. Recuperado de: https://link.springer.com/chapter/10.1007/978-3-031-09073-8_20
12. **Moine, J. M., Gordillo, S., & Haedo, A. S.** (2011). *Análisis comparativo de metodologías para la gestión de proyectos de minería de datos*. Recuperado de: https://sedici.unlp.edu.ar/bitstream/handle/10915/18749/Documento_completo.pdf?sequence=1&isAllowed=y
13. **Peixoto, A., & Silveira, F.** (2020). Predicting university dropout using machine learning techniques. *International Journal of Educational Technology in Higher Education*, 17(32), 1-15.
14. **Peña, L., & Pérez, C.** (2019). El análisis del abandono escolar en educación superior mediante técnicas de machine learning. *Revista Iberoamericana de Educación*, 79(2), 85-101.
15. **Pérez, A. M., Escobar, C. R., Toledo, M. R., Gutierrez, L. B., & Reyes, G. M.** (2018). Modelo de predicción de la deserción estudiantil de primer año en la Universidad Bernardo O'Higgins. *Educação E Pesquisa*, 44(0). <https://doi.org/10.1590/s1678-4634201844172094>

16. **Ramírez, P. E., & Grandón, E. E.** (2018). Predicción de la deserción académica en una universidad pública chilena a través de la clasificación basada en árboles de decisión con parámetros optimizados. *Formación Universitaria*, 11(3), 3-10.
<https://doi.org/10.4067/S0718-50062018000300003>
17. **Rivas, S., Cobo, M. J., & Herrera-Viedma, E.** (2019). Machine learning and student dropout in higher education: A meta-analysis. *IEEE Transactions on Education*, 62(3), 235-244.
18. **Santos, F., & Oliveira, P.** (2019). Socioeconomic influences on university dropout: An analysis using predictive models. *Higher Education Studies*, 9(4), 89-101.
19. **Segura, M., Mello, J., & Hernández, A.** (2022). Machine learning prediction of university student dropout: Does preference play a key role? *Mathematics*.
<https://doi.org/10.3390/math10183359>
20. **Song, Z., Sung, S-H., Park, D-M., & Park, B-K.** (2023). All-year dropout prediction modeling and analysis for university students. *MDPI Applied Sciences*. Recuperado de: <https://www.mdpi.com>
21. **Villar, A., & Robledo Velini de Andrade, C.** (2024). Supervised machine learning algorithms for predicting student dropout and academic success: A comparative study. *Discover Artificial Intelligence*. Recuperado de: <https://link.springer.com/article/10.1007/s44163-023-00079-z>
22. **Xing, W., & Du, D.** (2019). Dropout prediction in MOOCs: Using deep learning for personalized intervention. *Journal of Educational Computing Research*, 57(3), 570-590. Recuperado de: https://www.researchgate.net/publication/323784695_Dropout_Prediction_in_MOOCs_Using_Deep_Learning_for_Personalized_Intervention