I + T + C Investigación, Tecnología y Ciencia **ISSN: 1909-5775** **e-ISSN: 2805-7201**

Artículo de revisión

Prediction and Prevention of School Dropout through A.I.: A Review to Identify Models and Relevant Factors.

Predicción y prevención de deserción escolar mediante I.A.: Una revisión a fin de identificar modelos y factores relevantes.

Juan Carreño [1*], Diego Martínez [2], Deisy Paez [3*]

[1]Universidad Santo Tomas Facultad de Ingeniería Mecatrónica; juandavid.carreno@ustabuca.edu.co
[2]Universidad Santo Tomas Facultad de Ingeniería Mecatrónica; diegoandres.martinez@ustabuca.edu.co
[3]Universidad Santo Tomas Facultad de Ingeniería Mecatrónica; deisy.paez@ustabuca.edu.co

- Corespondencia: deisy.paez@ustabuca.edu.co

Citación: Carreño, J .; Martínez, D .; Páez, D. Predicción y prevención de deserción escolar mediante I.A.: Una revisión a fin de identificar modelos y factores relevantes. I + T + C Investigación, Tecnología y Ciencia. Vol 1. Num. 17. 2023.

Nota del editor: El Sello editorial Unicomfacauca se mantiene neutral con respecto a los reclamos derivados de los resultados de este trabajo

Resumen: La deserción estudiantil representa una preocupación latente en las instituciones educativas, según estadísticas del Ministerio de Educación de Colombia donde se informa que 473.786 niños y jóvenes estudiantes han interrumpido sus estudios entre noviembre de 2022 a mayo de 2023[1]. Especialmente en programas académicos de ciencia, tecnología, ingeniería y matemáticas (Science, Technology, Engineering, and Mathematics STEM) [2]. Abordar este desafío requiere la incorporación de herramientas de Tecnologías de la Información (TI) que ofrezcan seguimiento eficaz y oportuno a las áreas encargadas del control académico. El propósito de esta revisión bibliografica es explorar las variables que tengan relación con la deserción académica y encontrar modelos predictivos apropiados para el procesamiento de datos, además de identificar variables y modelos utilizados anteriormente en el tópico. Para lograr esto se propone una investigación mediante el uso de plataformas de búsqueda de carácter académico como Lens.org y Google académico. Una vez hecha la investigación se identifican las variables relevantes en el contexto nacional como rendimiento académico, edad, genero, condición familiar, aspectos psicológicos, entre otras, ya que se consideran relevantes para llegar a una predicción correcta y se selecciona el modelo de árboles de decisión C4.5 ya que se considera el que mejores resultados obtuvo en la investigación, su amplio uso en el campo y su bajo costo computacional.

Palabras clave: Modelos predictivos; Estudiantes en riesgo; Seguimiento académico; deserción estudiantil, arboles de decisión C4,5,

Abstract: School dropout is a pressing concern in educational institutions, as per statistics from the Ministry of Education of Colombia, which report that 473,786 children and young students have discontinued their studies between November 2022 and May 2023 [1]. This issue is especially prominent in Science, Technology, Engineering, and Mathematics (STEM) academic programs [2]. Addressing this challenge requires the integration of Information Technology (IT) tools that provide effective and timely monitoring to the academic control departments. The purpose of this literature review is to explore the variables related to academic dropout and find suitable predictive models for data processing, while also identifying variables and models previously used in the field. To achieve this, research is proposed using academic search platforms such as Lens.org and Google Scholar. After conducting the research, relevant variables in the national context are identified, such as academic performance, age, gender, family status, psychological aspects, among others, as they are considered crucial for accurate prediction. The C4.5 decision tree model was chosen due to its excellent performance in research, widespread usage in the field, and low computational cost.

Keywords: Predictive models; At-risk students; Academic monitoring; Student dropout; C4.5 decision trees.

DOI: 10.57173/ritc.v1n17a2

1. Introducción

La deserción escolar constituye un desafío significativo en el ámbito educativo, con implicaciones tanto para el proyecto de vida de los niños y jóvenes como para la sociedad, en [3] se observa la tendencia que ha llevado la deserción escolar en Colombia entre 2015 y 2020 la cual, si bien se aprecia descendente, se considera significativa; además de este se tiene un reporte [1] del Ministerio de Educación de Colombia en el que se observa que 473.786 jóvenes han abandonado el sistema educativo entre noviembre de 2022 a mayo de 2023. La necesidad de tratar este problema lleva a la búsqueda de herramientas de seguimiento y prevención que permitan intervenir de manera oportuna. La aplicación de la inteligencia artificial emerge como una herramienta eficiente para abordar este problema y diseñar intervenciones preventivas, como lo demuestran en [4] con el despliegue de una herramienta de predicción para educación superior donde consiguieron un nivel de acierto de 91.7%.

Este artículo se enfoca en analizar cómo las técnicas de clasificación apoyan en la identificación temprana de factores que contribuyen a la deserción estudiantil. Diversos estudios comparan técnicas de árboles de decisión, redes neuronales, máquinas de soporte vectorial, regresión logística, kNN. En [5] los algoritmos de árboles de decisión alcanzaron la presión más alta 74,6% para un conjunto de datos basados solo en desempeño académico. Para lograr esto, se plantea un enfoque metodológico orientado a identificar variables relevantes que influyen en la decisión de un estudiante de abandonar su educación, que pueden ser: académicas, socioeconómicas, culturales o personales.

El objetivo principal de este artículo es construir sobre estas experiencias previas y explorar la adaptación de una herramienta de prevención de deserción estudiantil basada en inteligencia artificial en el contexto educativo de Latinoamérica. Al enfocarse en este contexto específico, se aspira a desarrollar una solución replicable que pueda ser implementada en otras instituciones educativas y ciudades, contribuyendo así a la creación de estrategias más efectivas y personalizadas para combatir la deserción escolar y fortalecer el vínculo entre los estudiantes y su educación. Como referencia podemos ver un caso de éxito en Chile [6] para estudiantes de primaria y bachillerato.

2. Materiales y métodos

A partir de la selección de palabras clave se aplica la ecuación de búsqueda, ver Tabla 1, en las bases de datos de acceso libre lens.org donde se aplica el filtro de búsqueda para los últimos cinco años, en total arroja 16446 artículos de revista (3890 corresponden a año inmediatamente anterior), 828 conferencias y 511 capítulos de libro. Además, usando el software libre VOSViewer se representa de manera gráfica la red bibliográfica de los resultados de lens ver Figura 1.

Tabla 1. Estructura de la búsqueda

Plataforma	**https://www.lens.org/**
Tópicos de interés	**Tópico 1. Identificación de variables** Q1.1. ¿Cuáles son las variables que han sido usadas para entrenar los modelos de predicción de deserción académica? **Tópico 2. Identificación de modelos** Q2.1.¿Cuáles son los modelos de inteligencia artificial que han sido aplicados para predecir deserción académica?
Ecuación de búsqueda	(("machine learning" OR "artificial intelligence") AND ("student" OR "education") AND ("dropout" OR "attrition" OR "risk" OR "early warning" OR "academic performance"))
Filtro de campo	(Education, General Engineering, Artificial Intelligence, Software, Information Systems)

Filtro de palabras clave	(Education, Assessment, Curriculum, Machine Learning, Learning Outcomes, Evaluation, Higher Education, Students)

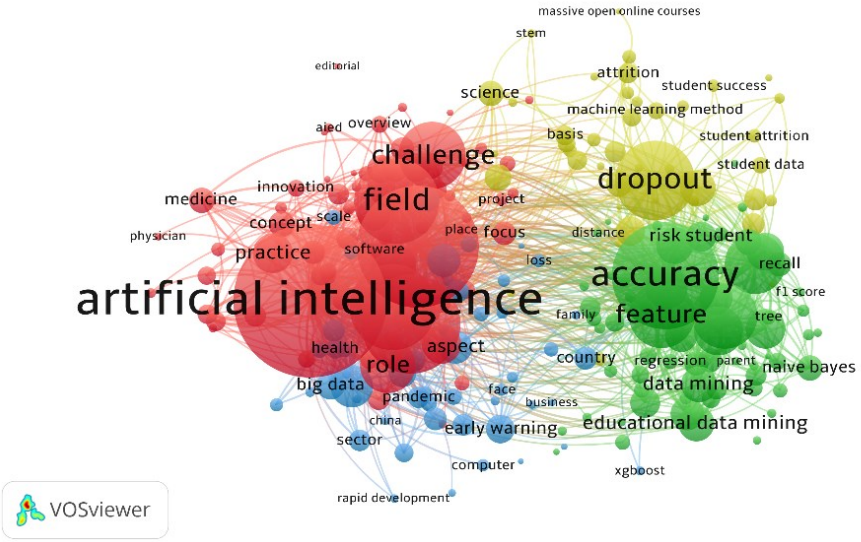

Figura 1. Red bibliográfica de la ecuación de búsqueda en VOSViewer

3. Resultados

Para esta revisión se realizaron dos investigaciones con el propósito de determinar el camino a seguir: una investigación referente al tópico de la deserción escolar para identificar los factores, ya sean sociales o académicos, que pueden afectar el rendimiento de un estudiante y llevarlo a abandonar sus estudios; y una investigación sobre modelos de inteligencia artificial predictivos para poder determinar qué modelo se ajusta a las necesidades del proyecto.

3.1. Determinación de Variables

3.1.1 Variables Académicas

La deserción escolar es un fenómeno multifacético que se ve influido por diversos factores académicos. Uno de los aspectos clave es el rendimiento académico del estudiante, como se plantea en [4] donde se encuentra evidencia de la relación negativa entre el rendimiento académico del estudiante como cantidad de materias aprobadas con el riesgo de abandono. Las dificultades persistentes en el logro de resultados satisfactorios pueden llevar a la pérdida de motivación y a una disminución en la autoconfianza del estudiante, lo que a su vez puede conducir al abandono de sus estudios. La falta de apoyo educativo adecuado, como la ausencia de programas de tutoría o recursos de aprendizaje adaptados a las necesidades individuales, también puede contribuir a este problema [7]. Además, la elección de cursos y programas inapropiados para las habilidades e intereses del estudiante podría generar desinterés y desconexión, incrementando la posibilidad de abandono. En este contexto, es esencial examinar cómo las variables académicas, como el rendimiento, la asistencia y el ajuste curricular, interactúan para determinar el riesgo de deserción escolar.

3.1.2 Variables Socioeconómicas

Los factores sociales desempeñan un papel crucial en la determinación del riesgo de deserción escolar. El entorno socioeconómico del estudiante puede tener un impacto significativo, como se encuentra en trabajos sobre deserción estudiantil un riesgo de abandono menor para estudiantes que provienen de hogares con un mayor estrato socioeconómico [7,8]. Además, aquellos provenientes de familias de bajos ingresos pueden enfrentar desafíos adicionales, como la necesidad de contribuir al sustento familiar o la falta de acceso a recursos educativos [9]. La falta de apoyo familiar y la ausencia de modelos a seguir en el ámbito educativo también pueden influir en la decisión de abandonar la educación. Además, la pertenencia a grupos minoritarios étnicos o culturales puede exponer a los estudiantes a experiencias de discriminación y marginación, lo que a su vez puede afectar negativamente su sentido de pertenencia a la comunidad educativa. La presión de pares y las dinámicas sociales en la escuela también juegan un papel importante, ya que la falta de relaciones positivas y el acoso pueden contribuir a la sensación de aislamiento y alienación, aumentando la probabilidad de deserción. Es imperativo considerar cómo estos factores sociales interaccionan y se combinan para influir en la persistencia educativa de los estudiantes.

3.1.3 Comparativa uso de variables

En el 100% de los estudios consultados las características de desempeño académico se incluyen en los modelos, además se observa que el 60% de los estudios tienen un enfoque por género y la edad. Se incluyen otras variables como lugar de procedencia [10][11][17], composición familiar [10-14][44], nivel de educación de los padres [10][11][44], como se presenta en la tabla 2.

Diversas variables se incluyen en diferentes estudios, como: estilos de aprendizaje, Consumo de sustancias sicoactivas, niveles de depresión se incluyen en [15], ausencia a clase [16][42], actividades extraescolares, categoría del colegio [6][43], costos y distancia de la escuela [44]

Tabla 2. Q1.1. ¿Cuáles son las variables que han sido usadas para entrenar los modelos de predicción de deserción académica?

Ref.	Rendimiento académico	No. Materias	Edad	Genero	Lugar de procedencia	Composición familiar	Nivel de educación padres	Necesidad de trabajo
[10]	x	x	x	x	x	x	x	
[11]	x			x		x	x	
[12]	x					x		x
[13]	x				x	x		
[14]	x		x	x		x		x
[5]	x							
[15]	x			x				
[16]	x							
[17]	x			x	x			
[6]	x							
[42]	x		x	x				
[43]	x		x	x				
[44]	x					x	x	
[46]	x		x	x				

3.2. Modelos de aprendizaje automático

3.2.1. K-Nearest Neighbor

K nearest neighbor (KNN) es un algoritmo de aprendizaje automático supervisado fundamental para aplicaciones en las que no se conoce claramente la distribución de los datos a utilizar, este es utilizado para tareas de clasificación y regresión de datos [18]. Este algoritmo se considera un método de lazy learning debido a que no implica un proceso explícito de entrenamiento, en cambio, al recibir un nuevo ejemplo de entrada busca en el conjunto de datos de entrenamiento los K ejemplos más cercanos a éste utilizando una medida de distancia para calcular la proximidad entre las muestras ya sea distancia euclidiana, manhattan o minkoswki dependiendo del problema y naturaleza de los datos [19,20]. El algoritmo KNN es fácil de entender y su implementación es relativamente simple. Sin embargo, su rendimiento puede ser afectado por la elección de la métrica de distancia, la normalización de las características y la selección adecuada de "K" [21]. Además, puede volverse computacionalmente costoso cuando se trabaja con conjuntos de datos muy grandes, ya que debe calcular las distancias entre el nuevo ejemplo y todos los ejemplos de entrenamiento.

3.2.2. Logistic Regression

La regresión logística es otro algoritmo de aprendizaje supervisado utilizado en problemas de clasificación binaria donde el objetivo es predecir la pertenencia a una de dos posibles clases, generalmente etiquetadas como 0 y 1. Este modelo utiliza una función logística o sigmoide para mapear las predicciones a una probabilidad entre 0 y 1. La hipótesis del modelo estima la probabilidad de pertenencia a la clase positiva basándose en coeficientes y características [22]. La función de costo evalúa la diferencia entre las predicciones y las etiquetas reales del conjunto de entrenamiento. El objetivo del entrenamiento en este modelo es ajustar los coeficientes para minimizar la función de costo [23,24]. Luego, se aplica un umbral de decisión para clasificar las muestras como positivas o negativas. La regresión logística también puede ser propensa al overfitting, especialmente cuando hay un alto número de variables predictoras en el modelo. La regularización se utiliza típicamente para penalizar los coeficientes grandes de los parámetros cuando el modelo sufre de alta dimensionalidad [25].

3.2.3. Support Vector Machine Clasiffier

Support Vector Machine es un algoritmo de aprendizaje supervisado utilizado principalmente en tareas de clasificación, el objetivo de un modelo SVC es encontrar un hiperplano, el cual es un plano que divide el espacio en dos regiones, en un espacio multidimensional que mejor separe las diferentes clases de datos de entrada [26]. Su característica distintiva es que el modelo busca hallar un margen máximo, siendo este la distancia entre el hiperplano y las muestras más cercanas a cada clase, maximizando el este margen para mejorar la capacidad de generalización del modelo y reducir el riesgo de overfitting [27]. En caso de que los datos no puedan ser separados perfectamente mediante un solo hiperplano, el algoritmo utiliza un “Kernel Trick” para mapear los datos a un espacio mayor de dimensionalidad donde puedan ser separados lo que permite realizar transformaciones no lineales en los datos, también es robusto frente a datos ambiguos [28]. Las principales debilidades de SVM son su rendimiento depende de una buena función de kernel y el parámetro de regularización [29], además su rendimiento puede ser menos eficiente que otros clasificadores específicamente diseñados para problemas de múltiples clases, como Random Forest o métodos de Redes Neuronales.

3.2.4. Gaussian Naive Bayes

El modelo GaussianNB es un algoritmo de aprendizaje supervisado utilizado para tareas de clasificación, hace referencia a un enfoque de clasificación basado en el teorema de Bayes, que se utiliza para calcular la probabilidad condicional de que un elemento pertenezca a una determinada clase, dada la observación de sus características y ciertas suposiciones simplificadas sobre la independencia de las características [30,31]. El término "Ingenuo" en Naive Bayes se refiere a la suposición de independencia condicional, lo que significa que asumimos que las características son independientes entre sí, dado que conocemos la clase [32]. Aunque esta suposición puede no ser cierta en la realidad, en la práctica, el algoritmo Naive Bayes funciona sorprendentemente bien en muchos casos y es ampliamente utilizado debido a su eficiencia y simplicidad. Este se utiliza principalmente para problemas donde las características de los datos siguen una distribución normal [33]. Por lo tanto, es adecuado para datos numéricos o continuos.

3.2.5. Decision trees

Los árboles de decisión son un método de clasificación supervisada usada en la inteligencia artificial. La idea nace de la forma de un árbol común, parte de un nodo inicial (raíz), cada nodo representa una característica o variable de los datos y sus ramificaciones serían el rango de valores que dicha variable puede adquirir [34]. Los árboles de decisión tienen distintas variaciones a partir de cuál algoritmo es usado para su construcción, teniendo varias opciones con distintas ventajas y desventajas. Una de estas opciones es el algoritmo ID3, o su variación el C4.5 que resulta en un árbol con menos nodos de decisión y por ende un costo computacional menor [35,36]. En cuanto a lo negativo se puede hablar de lo inexactos que son estos para cálculos complejos o toma de decisiones que dependan de muchas variables, por lo que es mejor trabajar con una cantidad de variables pequeña si se quiere usar este método.

3.2.6. Neural networks

Una red neuronal es un método de aprendizaje profundo (deep learning) con una intervención humana bastante limitada con respecto a métodos de machine learning [37]. El objetivo de esta es simular el funcionamiento de una red neuronal de un cerebro humano, donde capas de múltiples neuronas se conectan entre sí asignando distintos valores que se ajustan de acuerdo al conjunto de datos de entrenamiento [38]. Lo llamativo de estas redes puede ser la capacidad de que una computadora aprenda de sus errores, haciendo el modelo mucho más flexible que otros, también es usado para tareas bastante complejas como lo puede ser la visión artificial, pero esta es también una de sus desventajas, ya que, al usarse para tareas complejas, su costo computacional tiende a ser elevado, haciendo que no valga la pena ser usado para tareas más simples pero que involucren grandes volúmenes de datos. Esto se puede solucionar con redes neuronales convolucionales, pero tienen a ser más complejas en su realización y difíciles de entender [39].

3.2.7. Modelos utilizados en artículos relevantes

Este es un problema de clasificación que se puede solucionar por diversos métodos, en la Tabla 3 se presenta la tendencia de los algoritmos más utilizados

Las métricas más usadas para la medición de la eficiencia son: Accuracy, Precision, Recall, F1 [6][17][42-47] y otras técnicas como: matriz de confusión, área bajo la curva (ROC-AUC) [5], media geométrica GM [6][15].

Además se observa el usa de diversas técnicas: SMOTE, SHAP, CATBoost, Head map XGBoost-, lightGBM, RF, J48, REPTree [15-17][42-47]

Tabla 3. Q2.1.¿Cuáles son los modelos de inteligencia artificial que han sido aplicados para predecir deserción académica?

Ref.	Árboles de decisión	Redes neuronales	Regresión Logística	Máquina soporte vectorial	k-means	kNN	Random Forest (RF)
[5]	x	x	x	x	x		
[15]			x				
[16]	x						x
[17]					x		
[6]	x						
[42]	x		x	x			
[43]	x						x
[44]			x			x	x
[45]				x		x	x
[46]	x					x	
[47]	x					x	x

4. Discusión

Una vez terminada la investigación se pueden analizar los resultados de esta, donde se empieza mencionando la importancia de usar no solamente variables netamente académicas como lo podrían ser notas que sigan el proceso educativo o un registro de asistencia de cada espacio académico, también se deben tomar en cuenta variables de tipo social, económico o cultural que puedan interferir en la relación entre el estudiante y su educación. Algo importante a destacar referente a las variables, es la diversidad que se puede encontrar en los diferentes estudios, como podemos observar en la tabla 2, donde los autores usan variables comunes, pero cada uno tiene un enfoque distinto llevando sus estudios más al lado social o al lado psicológico.

Por otro lado, se tiene el estudio realizado a los posibles métodos que se pueden usar para realizar un modelo que permita hacer la predicción a tiempo de un riesgo de deserción. Para ello se toma en cuenta que sea un modelo de clasificación que agrupe a los estudiantes en dos grupos según las variables, los que pasan y los que no. Además, centrando la herramienta a algo general que pueda ser usado en cualquier institución, se toma en cuenta también el costo computacional de cada uno de los métodos.

Una vez mencionados los criterios de selección, se deja de lado el método KNN y redes neuronales por su alto costo computacional al trabajar con grandes volúmenes de datos, la cual sería la condición de trabajo ideal en una institución que tenga un gran historial de estudiantes en su base de datos. También descartamos *Suport Vector Machine Clasiffier* por lo mencionado anteriormente, existen clasificadores con un mejor rendimiento como los *Random forest* (basados en los árboles de decisión). Esto se refuerza analizando los resultados obtenidos en la tabla 2, donde podemos analizar la tendencia de uso de árboles de decisión para estos casos.

Finalmente quedan tres métodos que cumplen con los requisitos esperados, pero que siguen sin ser perfectos. Se empieza analizando la regresión lineal, la cual con grandes volúmenes de datos puede sufrir de *overfitting* (sobre entrenamiento), mientras que de *GaussianNB* y los árboles de decisión hay poco que resaltar, siendo los que más se ajustan a lo requerido, con la única aclaración de encontrar el mejor algoritmo para los árboles de decisión, teniendo como opción principal por bajo costo computacional, el C4.5.

5. Conclusiones

La investigación revela que la predicción del riesgo de deserción escolar no puede limitarse únicamente a variables académicas, como calificaciones y asistencia. Es importante reconocer la influencia de factores sociales, económicos y culturales en la relación entre el estudiante y su educación. Bajo esto se destaca la necesidad de un enfoque completo al abordar el problema de la deserción escolar, considerando tanto los factores académicos como sociales para lograr una evaluación más precisa y efectiva. Integrar estas variables en modelos de predicción puede proporcionar una visión más completa de la situación y permitir intervenciones personalizadas que aborden las necesidades únicas de cada estudiante.

En la investigación de posibles métodos de predicción de riesgo de deserción se destaca la importancia de elegir herramientas adecuadas para abordar este desafío. Se ha propuesto un enfoque de clasificación que divide a los estudiantes en dos grupos: los que pasan y los que no. Si bien se han identificado múltiples métodos prometedores, la consideración del costo computacional es crucial, especialmente en instituciones con grandes bases de datos. La selección final se inclina hacia el uso de algoritmos como Random Forest y árboles de decisión, que han demostrado un rendimiento sólido. La elección del algoritmo C4.5 se perfila como una opción principal debido a su eficiencia computacional, aunque se reconoce la necesidad de seguir explorando y optimizando los algoritmos de árboles de decisión para alcanzar el máximo potencial en la predicción del riesgo de deserción escolar.

Fondos: Universidad Santo Tomás de Aquino seccional Bucaramanga. XI convocatoria interna de semilleros de investigación 2023. Desarrollado por el semillero TURING con el proyecto "Alertas tempranas de desempeño académico usando inteligencia artificial. Piloto en secundaria y educación superior."

Conflictos de interés: Los autores declaran no tener ningún conflicto de intereses.

Referencias

1. Radio Nacional de Colombia. Disponible en línea: https://www.radionacional.co/actualidad/educacion/la-desercion-escolar-en-colombia-aumento-en-el-2023-panorama-preocupante#:~:text=2023%20-%2016%3A30-,Según%20el%20Ministerio%20de%20Educación%2C%20la%20deserción%20escolar%20aumentó%20en,comparación%20con%20los%20años%20anteriores (consultado el 28, 07, 2023).
2. Nagy, M., Molontay, R., 2023. Interpretable Dropout Prediction: Towards XAI-Based Personalized Intervention. International Journal of Artificial Intelligence in Education.. https://doi.org/10.1007/s40593-023-00331-8
3. Ministerio de Educación Nacional. (2022, Ago. 5) DESERCIÓN ESCOLAR EN COLOMBIA: ANÁLISIS, DETERMINANTES Y POLÍTICA DE ACOGIDA, BIENESTAR Y PERMANENCIA. [Online]. Disponible: https://www.mineducacion.gov.co/1780/articles-363488_recurso_34.pdf
4. O. Castrillón, W. Sarache y S. Ruiz. "Predicción del rendimiento académico por medio de técnicas de inteligencia artificial" *Form. Univ.* vol.13, no.1, pp.93-102, Febrero 2020
5. Yağcı, M., 2022. Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learning Environments 9. https://doi.org/10.1186/s40561-022-00192-z
6. Rodríguez, P., Villanueva, A., Dombrovskaia, L., Valenzuela, J.P., 2023. A methodology to design, develop, and evaluate machine learning models for predicting dropout in school systems: the case of Chile. Education and Information Technologies 28, 10103–10149.. https://doi.org/10.1007/s10639-022-11515-5
7. Castaño, E., Gallón, S., Gómez, K. y Vásquez, J. Deserción estudiantil universitaria: una aplicación de modelos de duración. Lecturas de economía, 2004. 60, 39-65.
8. Benites, R. M. El papel de la tutoría académica para elevar el rendimiento académico de los estudiantes universitarios. Revista Conrado. 2020. 16(77), 315-321.
9. Ishitani, T. Studying attrition and degree completion behavior among first generation college students in the United States. The Journal of Higher Education, 2006. 77(5), 861-885.
10. Castillo Caicedoa, M., Osorio Mejíab, A. M. y Montero Cuartasc, S. Deserción y retención, en la carrera de Economía de la Pontificia Universidad Javeriana Cali: un análisis de supervivencia, 2000-2008. Economía, Gestión y Desarrollo, 2010 9, 11- 33.
11. Giovanoli, P. Determinantes de la deserción y graduación universitaria: una aplicación utilizando modelos de duración. Documento de trabajo, 37. Argentina: Universidad Nacional de La Plata. 2002.

12. García Ramírez, R. G. García Montejo, J. S. ANÁLISIS CARACTERÍSTICO DE LOS FACTORES DE LA DESERCIÓN EN EDUCACIÓN SUPERIOR. Revista de divulgación científica y tecnológica. 2022. Vol 7, No. 3. 21-31
13. Jim←nes Garc←s, C. Vieyra Reyes, P. Trujillo Condes, V. E. Hernandez Gonzales, M. M. Factores asociados al rendimiento acad ←mico y deserci◉n escolar en educaci◉n media superior: Reflexiones. AMeditores. 2022
14. Guayacán, J. Estado de la deserción escolar en los establecimientos oficiales de Colombia. 2015 Recuperado de: http://hdl.handle.net/20.500.12209/779.
15. Hoyos Osorio, J.K., Daza Santacoloma, G., 2023. Predictive Model to Identify College Students with High Dropout Rates. Revista Electrónica de Investigación Educativa 25, 1–10.. https://doi.org/10.24320/redie.2023.25.e13.5398
16. Lee, S., Chung, J.Y., 2019. The Machine Learning-Based Dropout Early Warning System for Improving the Performance of Dropout Prediction. Applied Sciences 9, 3093.. https://doi.org/10.3390/app9153093
17. Kim, S., Choi, E., Jun, Y.-K., Lee, S., 2023. Student Dropout Prediction for University with High Precision and Recall. Applied Sciences 13, 6275.. https://doi.org/10.3390/app13106275
18. F. Pacho and D. Chiqui. "Estudio de las causas de la deserción escolar," B.S. Thesis. Cuenca, 2011. [Online]. Available: http://dspace.ucuenca.edu.ec/handle/123456789/1868
19. E. Ortega de Ávila, B. V. Alvarado de la Torre, M. G. Balderrábano Saucedo, C. A. Martínez Cardona, & J. O. Bautista Acosta. Implicaciones de la deserción escolar a nivel superior en Ingeniería en Sistemas e Informática. Coloquio de investigación multidisciplinaria, 2019. 7(1), 2383–2390.
20. Leif E. Peterson, K-nearest neighbor. Scholarpedia. 2009. Disponible en línea: http://scholarpedia.org/article/K-nearest_neighbor
21. Kramer, O. K-Nearest Neighbors. Intelligent Systems Reference Library, 2013. 13–23.
22. Daniel T. Larose; Chantal D. Larose. k-Nearest Neighbor Algorithm. Discovering Knowledge in Data: An Introduction to Data Mining. 2014. pp.149-164,
23. Dudani, S.A. The distance-weighted k-nearest-neighbor rule. IEEE Trans. Syst. Man Cybern., SMC-6:325–327, 1976.
24. Moore, A. W., & Komarek, P. Logistic regression for data mining and high-dimensional classification. Carnegie Mellon University ProQuest Dissertations Publishing, 2004. 18–20
25. Sperandei, S. Understanding logistic regression analysis. Biochemia Medica, 2014. 12–18.
26. Bisong, E. Logistic Regression. In: Building Machine Learning and Deep Learning Models on Google Cloud Platform. Apress, Berkeley, CA. 2019.
27. Zou, X., Hu, Y., Tian, Z., & Shen, K. Logistic Regression Model Optimization and Case Analysis. 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT). 2019
28. Noble, W. S. What is a support vector machine? Nature Biotechnology, 2006. 24(12), 1565–1567.
29. Mammone, A., Turchi, M., & Cristianini, N. Support vector machines. Wiley Interdisciplinary Reviews: Computational Statistics, 2009. 1(3), 283–289.
30. Otchere, D. A., Ganat, T. a. O., Gholami, R., & Ridha, S. Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models. Journal of Petroleum Science and Engineering, 2021. 200,
31. ZHAO, C., ZHANG, H., ZHANG, X., LIU, M., HU, Z., & FAN, B. Application of support vector machine (SVM) for prediction toxic activity of different data sets. Toxicology, 2006. 217(2-3), 105–119.
32. Kamel, H.; Abdulah, D.; Al-Tuwaijari, J. M. Cancer Classification Using Gaussian Naive Bayes Algorithm. 2019 International Engineering Conference (IEC). 2019
33. Gayathri, B., & Sumathi, C. P. An Automated Technique using Gaussian Naïve Bayes Classifier to Classify Breast Cancer. International Journal of Computer Applications, 2016. 148(6), 16–21.
34. Hemachandran, K., Tayal, S., George, P. M., Singla, P., & Kose, U. Bayesian reasoning and Gaussian processes for machine learning applications. In Chapman and Hall/CRC eBooks. 2022 3-5
35. Ontivero-Ortega, M., Lage-Castellanos, A., Valente, G., Goebel, R., & Valdes-Sosa, M. Fast Gaussian Naïve Bayes for searchlight classification analysis. NeuroImage, 2017. 163, 471–479.
36. Kingsford, C., & Salzberg, S. L. What are decision trees? Nature Biotechnology, 2008. 26(9), 1011–1013.
37. Adhatrao, K., Gaykar, A., Dhawan, A., Jha, R., & Honrao, V. Predicting students' performance using ID3 and C4.5 classification algorithms. International Journal of Data Mining & Knowledge Management Process, 2013. 3(5), 39–52.
38. Ozsoy, S., Gümüş, G., & Khalilov, S. C4.5 versus other decision trees: A review. Computer Engineering and Applications, 2015. 4(3), 173–182.
39. Lawrence, J. Introduction to neural networks. California Scientific Software, USA. 1993.
40. Naim, A. E-Learning Engagement through Convolution Neural Networks in Business Education. European Journal of Innovationin Nonformal Education. 2022. Volumen 2 497-501
41. Aggarwal, C. C. Neural networks and deep learning: A Textbook. Springer. 2018.
42. Song, Z., Sung, S.-H., Park, D.-M., Park, B.-K., 2023. All-Year Dropout Prediction Modeling and Analysis for University Students. Applied Sciences 13, 1143.. https://doi.org/10.3390/app13021143
43. Flores V, Heras S, Julian V. Comparison of Predictive Models with Balanced Classes Using the SMOTE Method for the Forecast of Student Dropout in Higher Education. Electronics. 2022; 11(3):457. https://doi.org/10.3390/electronics11030457

44. Mnyawami, Y.N., Maziku, H.H., Mushi, J.C., 2022. Enhanced Model for Predicting Student Dropouts in Developing Countries Using Automated Machine Learning Approach: A Case of Tanzanian's Secondary Schools. Applied Artificial Intelligence 36.. https://doi.org/10.1080/08839514.2022.2071406
45. Adnan, M., Habib, A., Ashraf, J., Mussadiq, S., Raza, A.A., Abid, M., Bashir, M., & Khan, S.U. (2021). Predicting at-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models. IEEE Access, 9, 7519-7539.
46. Iam-On, N., & Boongoen, T. (2015). Improved student dropout prediction in Thai University using ensemble of mixed-type data clusterings. International Journal of Machine Learning and Cybernetics, 8(2), 497–510. doi:10.1007/s13042-015-0341-x
47. Livieris, I. E., Kotsilieris, T., Tampakas, V., & Pintelas, P. (2018). Improving the evaluation process of students' performance utilizing a decision support software. Neural Computing and Applications. doi:10.1007/s00521-018-3756-y