



MODELO BASADO EN MACHINE LEARNING PARA PREDECIR LA APROBACIÓN DE UN CURSO DE INGENIERÍA

MODEL BASED ON MACHINE LEARNING TO PREDICT THE APPROVAL OF AN ENGINEERING COURSE

¹Fernando Luis Carrascal Porras, ²Wilson de Jesús Arrubla Hoyos,
³Luis Alfonso Flórez Prias, ⁴Daniel Francisco Chica Medrano

^{1,2,3,4}Universidad Nacional Abierta y a Distancia, Colombia

Recibido: 20/10/2023 Aprobado 20/11/2023

RESUMEN

En el contexto de la educación virtual y a distancia, el curso “Introducción a la Ingeniería de Telecomunicaciones” de la Universidad Nacional Abierta y a Distancia (UNAD) ha enfrentado el desafío de la deserción estudiantil en primera matrícula. Para abordar esta problemática, se realizó una investigación centrada en desarrollar un modelo de minería de datos basado en la metodología CRISP-DM. El estudio utilizó un dataset con 808 registros de estudiantes recopilados en distintos periodos académicos del año 2021. Se aplicaron diversas técnicas de machine learning, como Random Forest, Tree Decision, Knn, SVM y Neural Network, evaluando su desempeño mediante validación cruzada k-10. Los resultados revelaron que el modelo Random Forest obtuvo el mejor rendimiento, logrando predecir con un 80 % de precisión la aprobación del curso. Este modelo es relevante para identificar los estudiantes con riesgo de deserción temprana del curso, permitiendo la implementación de estrategias proactivas para hacer frente a la deserción y mejorar los indicadores del curso. Aunque algunos resultados fueron ligeramente inferiores a investigaciones anteriores, el modelo sigue siendo una herramienta valiosa para apoyar las estrategias de retención de los estudiantes en el curso “Introducción a la Ingeniería de Telecomunicaciones” de la UNAD.

Palabras clave: clasificación predictiva, deserción estudiantil, Random Forest, retención, permanencia.

Citación: Carrascal Porras, F. L. ., Arrubla Hoyos, W. de J. ., Florez Prias, L. A. ., & Chica Medrano, D. F. . (2023). Modelo basado en machine learning para predecir la aprobación de un curso de ingeniería. *Publicaciones E Investigación*, 17(4). <https://doi.org/10.22490/25394088.7510>

¹fernando.carrascal@unad.edu.co / <https://orcid.org/0000-0002-6674-7321>

²wilson.arrubla@unad.edu.co / <https://orcid.org/0000-0001-7119-7603>

³luis.florez@unad.edu.co / <https://orcid.org/0000-0002-5350-0119>

⁴daniel.chica@unad.edu.co / <https://orcid.org/0000-0001-7558-2461>

<https://doi.org/10.22490/25394088.7510>

ABSTRACT

In the context of virtual and distance education, the course “Introduction to Telecommunications Engineering” at the National Open and Distance University (UNAD) has faced the challenge of student attrition at first enrolment. To address this problem, a research study was carried out to develop a Data Mining model based on the CRISP-DM methodology. The study used a dataset with 808 student records collected in different academic periods of the year 2021. Several Machine Learning techniques were applied, such as Random Forest, Tree Decision, Knn, SVM and Neural Network, evaluating their performance through k-10 cross-validation. The results revealed that the Random Forest model obtained the best performance, predicting with 80 % accuracy the passing rate of the course. This model is relevant for identifying students at risk of dropping out of the course early, allowing the implementation of proactive strategies to address dropout and improve course indicators. Although some results were slightly lower than previous research, the model remains a valuable tool to support student retention strategies in the “Introduction to Telecommunications Engineering” course at UNAD.

Keywords: Predictive classification, Random Forest, retention, permanency, student dropout.



1. INTRODUCCIÓN

La deserción estudiantil es un problema que aqueja a muchas Universidades de Colombia, por ello, el Ministerio de Educación Nacional MEN estableció el sistema de prevención y análisis de la deserción en las instituciones de educación superior SPADIES desde el 2006. En particular, este sistema define el concepto “primera deserción” (*first drop out*) que hace referencia a que el estudiante no se matricule en un programa académico por dos semestres consecutivos. Asimismo, este sistema busca generar un servicio de alertas para que las instituciones de educación superior ejecuten las medidas necesarias para evitar la deserción. Por un lado, este fenómeno genera afectaciones directas en los proyectos de vida de los estudiantes y por otro desequilibra la ecuación de sostenibilidad de la autonomía universitaria dictada por la Ley 30 de 1992 en cuanto a la aplicación de recursos para el cumplimiento de su misión.

Como se indicó anteriormente, la deserción es un fenómeno de gran relevancia en las IES, y más aún, cuando la Universidad tiene más de cien mil estudiantes, el modelo pedagógico se basa en la modalidad virtual y a distancia soportado en las tecnologías de la información y comunicación TIC y tiene cobertura en todo el territorio nacional. Las características

anteriormente mencionadas son particulares de la Universidad Nacional Abierta y a Distancia –UNAD–. Frente a este panorama se requiere que se desarrollen una serie de estrategias de retención y permanencia que permita mitigar los efectos de la deserción. En consecuencia, se identifica una oportunidad de adelantar investigaciones teórico-prácticas que permitan focalizar de manera temprana los estudiantes con mayor riesgo de reprobación los cursos.

El diseño de este proyecto es una aplicación práctica que se contextualiza en el curso de Introducción a la ingeniería de telecomunicaciones de la UNAD, donde se ha identificado que históricamente la reprobación del curso está por encima del 30 %. Una de las afectaciones que genera este fenómeno en el programa de Ingeniería de Telecomunicaciones es el recorte de profesionales por el número de estudiantes que avanzan a periodos superiores, asimismo, si no se busca controlar este factor es posible que se presenten problemas para acreditarse como programa de alta calidad ante el MEN.

Dicho esto, este proyecto propone desarrollar un modelo predictivo basado en técnicas de *machine learning* que permita segmentar los estudiantes que

pueden reprobar el curso Introducción a la ingeniería de telecomunicaciones de manera temprana; permitiendo focalizar las estrategias de retención y permanencia para hacer frente al fenómeno de la deserción del programa. Para lograr este objetivo se utiliza la metodología CRISP-DM ampliamente utilizada en proyectos de minería de datos que permite de manera organizada entender el negocio, comprender los datos, preparar los datos, modelar con diferentes técnicas de ML, evaluar desde las métricas de calidad que brinda la matriz de confusión y finalmente hacer despliegue del proyecto.

Este proyecto está organizado en 6 capítulos que se describen a continuación: el capítulo 1 aborda el planteamiento del problema, descripción del problema, justificación objetivos y alcance del proyecto; el capítulo 2 presenta los fundamentos teóricos del proyecto desde los ejes conceptuales del ML y la deserción estudiantil; el capítulo 3 aborda la metodología con la que se logra cumplir los objetivos trazados; por su parte, el capítulo 4 muestra la solución propuesta para dar solución al problema; en el capítulo 5 se presentan los resultados con sus respectivos análisis y discusión; finalmente, en el capítulo 6 se presentan las conclusiones seguido de las referencias bibliográficas.

2. MATERIALES Y MÉTODOS

Existen diversas técnicas disponibles para abordar proyectos de minería de datos, entre las cuales se destacan CRISP-DM, KDD y SEMMA, reconocidas metodologías que se centran en obtener resultados concretos.

La metodología de descubrimiento de conocimientos en bases de datos (KDD) se compone de cinco fases fundamentales. Estas etapas abarcan desde la comprensión del contexto empresarial y la

interpretación de los datos, hasta la preparación de los mismos, el modelado y la evaluación de los resultados obtenidos.

Estas metodologías brindan un enfoque estructurado para dirigir el desarrollo de proyectos de minería de datos, centrándose en la consecución de los resultados deseados. La implementación del modelo se erige como un paso crucial en este proceso, asegurando la efectiva puesta en práctica del modelo desarrollado.

La metodología también implica una comprensión detallada del problema o los requisitos empresariales, así como la recopilación y exploración exhaustiva de los datos. La preparación de los datos se lleva a cabo mediante técnicas descriptivas y limpieza estadística, seguido de su división con fines de entrenamiento y pruebas.

El propósito último de los proyectos de minería de datos radica en la capacidad de prever resultados futuros basados en la información disponible. Esto puede incluir la predicción de riesgos, la activación de clientes, el análisis de series temporales o la anticipación de necesidades de inventario.

3. DESARROLLO

El presente artículo se enfocó en desarrollar un modelo utilizando técnicas de minería de datos basadas en la metodología CRISP-DM. Inicialmente, se creó un dataset que comprendía 808 registros de estudiantes del curso “Introducción a la ingeniería de telecomunicaciones” de la UNAD, recopilados durante los periodos académicos 16-01, 16-02, 8-03 y 160-4 del año 2021.

La siguiente figura representa las fases llevadas a cabo durante el desarrollo del proyecto:

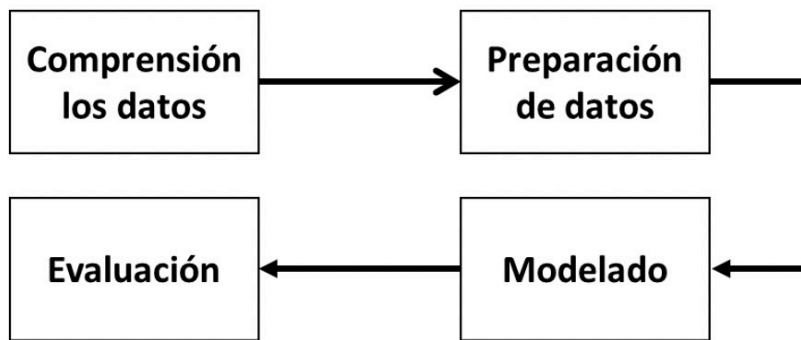


Figura 1. Fases del proyecto.

Comprensión de los datos: el dataset consta de 8 describen a continuación: variables descriptoras y una variable objetivo, que se

TABLA 1. DESCRIPCIONES VARIABLES DEL DATASET

Variable	Descripción
Repitente	Indica si el estudiante perdió el curso y lo ha matriculado nuevamente.
GE	(Generación E): refleja si el estudiante es becado por el programa del gobierno Generación E.
Zona	Indica la zona del país en la que el estudiante se encuentra inscrito, considerando las diferentes regiones de la UNAD.
Estado civil	Describe el estado civil del estudiante (soltero, casado, divorciado, unión libre, separado).
Sexo	Representa el género del estudiante (masculino o femenino).
Condición especial	Hace referencia si los estudiantes ingresan al programa por homologación (estudiante por homologación, estudiante por convenio SENA o ninguna).
Nota momento Inicial	Indica la calificación de la primera tarea del curso, correspondiente al 25 % del mismo y realizada en las dos primeras semanas del periodo académico.
Numero de matrículas	Registra el número de veces que el estudiante ha matriculado el curso.
Aprobación curso	Representa la variable objetivo del dataset e indica si el estudiante aprueba o no el curso.

Preparación de los datos: se realizaron tareas de limpieza, transformación y selección de características para garantizar la calidad de los datos y prepararlos adecuadamente para el entrenamiento del modelo.

Pandas en el entorno de Google Colab con el lenguaje de programación Python. Se procedió a entrenar varios modelos, entre ellos Random Forest (RF) y Tree Decision, aplicando validación cruzada (VC).

Modelado: se aplicaron técnicas de machine learning utilizando las librerías Scikit-learn, Matplotlib y

Evaluación: se empleó la validación cruzada k-10 en el dataset para calcular y analizar diversas métricas

de calidad, incluyendo el accuracy, precisión y recall, con el objetivo de obtener una visión integral del desempeño de los modelos predictivos. Estas métricas proporcionaron información relevante para medir la efectividad de las predicciones y determinar qué tan bien se ajustaban los modelos a los datos de prueba.

4. DISCUSIÓN

La presente investigación sugiere un enfoque en el desarrollo de un “Modelo de predicción basado en técnicas de minería de datos para determinar el nivel de riesgo en deserción estudiantil en la Universidad Nacional Abierta y a Distancia (UNAD) sede Corozal”. La discusión aborda varios aspectos, entre ellos:

Minería de datos en educación: analizar cómo las técnicas de minería de datos se aplican en el ámbito educativo, específicamente para prever el riesgo de deserción estudiantil.

Factores de riesgo en deserción estudiantil: explorar y discutir los posibles factores de riesgo que podrían llevar a la deserción estudiantil, tanto a nivel general como en el contexto específico de la UNAD sede Corozal.

Aplicación de modelos predictivos en educación: examinar la utilidad y eficacia de los modelos de predicción basados en técnicas de minería de datos para anticipar y abordar la deserción estudiantil en instituciones educativas, centrándose en la UNAD sede Corozal como caso de estudio.

Implicaciones prácticas y políticas: discutir cómo los resultados de este modelo de predicción podrían tener implicaciones prácticas en la gestión y toma de decisiones en la UNAD sede Corozal, así como su posible aplicación en otras instituciones educativas.

Ética en la utilización de datos estudiantiles: considerar y abordar las cuestiones éticas relacionadas con la recopilación y el uso de datos estudiantiles para predecir la deserción, destacando la importancia de la privacidad y la seguridad de la información.

Este tema de investigación abarcaría aspectos interdisciplinarios que involucran la educación, la minería de datos, la tecnología de la información y la toma de decisiones institucionales.

En la siguiente tabla se presentan los resultados de los distintos métodos de machine learning aplicados.

TABLA 2. RESULTADOS VALIDACIÓN CRUZADA

Técnica ML	accuracy	precision	recall	F1- Score
Tree Desicion	73 %	74 %	72 %	73 %
RF	79 %	81 %	75 %	78 %
Knn	75 %	73 %	80 %	76 %
SVM	71 %	91 %	47 %	61 %
Neural Network	60 %	65 %	62 %	57 %
Stacking	79 %	79 %	78 %	78 %

Los resultados presentados revelan que la técnica de Random Forest obtuvo un rendimiento superior. La siguiente técnica con mejor desempeño fueron los ensambles Stacking. Estos hallazgos coinciden

con la investigación realizada por Bean & Metzner (1985), donde los ensambles mostraron un rendimiento superior en comparación con otras técnicas, como las redes neuronales ANN.

La siguiente figura presenta la matriz de confusión que proporciona una visión detallada del rendimiento del modelo RF, el cual demostró un

desempeño destacado. El accuracy del modelo alcanza un 79 % de acierto, con una precisión del 81 %, y un 75 % de recall.

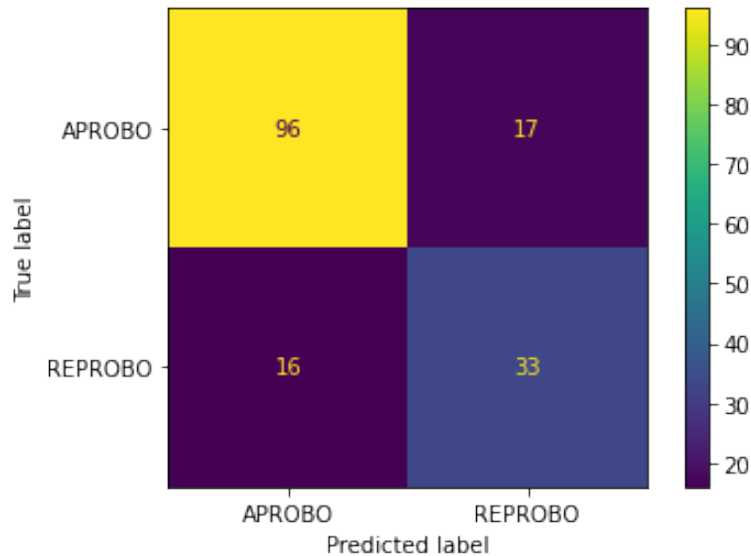


Figura 2. Matriz de confusión modelo con mayor rendimiento.

Además, la curva ROC muestra un sólido rendimiento del 82 %. Todas estas métricas se derivan

de la matriz de confusión, la cual se puede observar en la Figura 3.

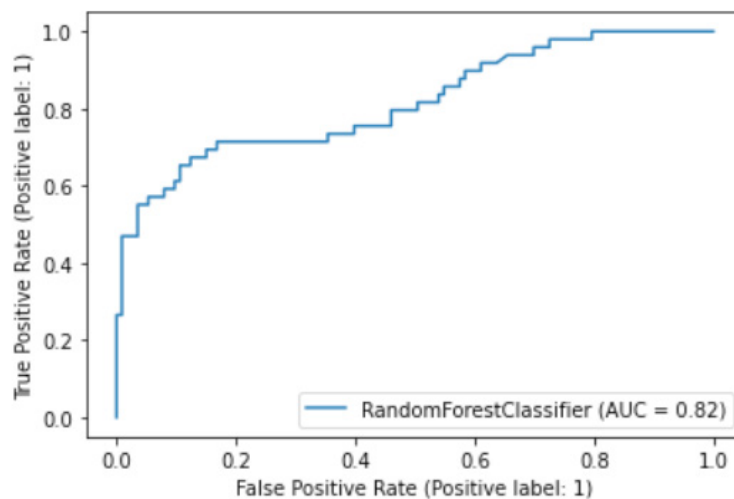


Figura 3. curva ROC modelo RF.

Por otro lado, los resultados de Acero, Achury & Morales (2019) mostraron un accuracy del 88 %, lo cual es ligeramente superior a los obtenidos en esta

investigación. Sin embargo, es importante destacar que los resultados son relevantes para la dinámica propia que tiene la UNAD. El modelo desarrollado puede servir

como un apoyo en las estrategias de retención y permanencia, ya que logra clasificar con un 80 % de exactitud a los estudiantes con riesgo de perder el curso de manera temprana. Esto permitiría a las redes de curso aplicar acciones de retención y permanencia de manera oportuna para mejorar los indicadores del curso.

Por otro lado, en la investigación de Amazon Machine Learning (2020), se empleó la técnica Knn con resultados favorables; sin embargo, en este proyecto, fue la técnica con peor rendimiento. La discrepancia en el rendimiento de Knn entre ambas investigaciones podría deberse a varias razones. En primer lugar, cada conjunto de datos es único y puede contener características particulares que favorezcan o dificulten el desempeño de ciertos algoritmos. Es posible que el conjunto de datos utilizado en este proyecto tenga una distribución diferente o una estructura más compleja, lo que podría influir en el rendimiento de Knn y explicar su bajo desempeño en esta ocasión.

5. CONCLUSIONES

Esta investigación ha culminado en la creación de un modelo predictivo con una precisión del 80 % para clasificar si los estudiantes tendrán éxito o no en el curso “Introducción a la ingeniería de telecomunicaciones”. Estos resultados no solo son destacables por su elevada exactitud, sino que también llevan consigo implicaciones significativas para las estrategias de retención estudiantil.

La capacidad demostrada para prever con precisión la aprobación del curso se erige como una herramienta estratégica para las redes de curso, permitiéndoles implementar acciones más selectivas y efectivas. Al identificar a los estudiantes con mayor riesgo de no aprobar, el modelo ofrece la oportunidad de intervenir proactivamente, brindando un apoyo personalizado que puede marcar la diferencia en la trayectoria académica de estos individuos.

La aplicabilidad práctica de este modelo se traduce en la mejora palpable de los resultados académicos y

la reducción de la deserción estudiantil en el curso de “Introducción a la ingeniería de telecomunicaciones”. Al anticipar las dificultades que podrían enfrentar los estudiantes, las estrategias de apoyo pueden ser implementadas de manera temprana y específica, contribuyendo no solo a la retención sino también al éxito académico de un número significativo de estudiantes.

En última instancia, esta investigación no solo ha perfeccionado un modelo predictivo, sino que también ha proporcionado a las instituciones educativas una herramienta práctica y valiosa. La combinación de precisión y aplicabilidad conlleva la promesa de un impacto positivo tanto en la experiencia académica individual de los estudiantes como en la salud general del curso, sentando las bases para futuras investigaciones y mejoras continuas en la gestión educativa.

REFERENCIAS

- Acero, A., Achury, J. C., & Morales, J. C. (2019). University dropout: A prediction model for an engineering program in Bogotá, Colombia. En B. Kloot (ed.). *Proceedings of the 8th Research in Engineering Education Symposium, REES 2019 - Making Connections* (pp. 483-490). <https://pure.udem.edu.mx/es/publications/university-dropout-a-prediction-model-for-an-engineering-program->
- Amaya Torrado, Y. K., Barrientos Avendaño, E., & Heredia Vizcaino, D. J. (2014). *Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos*. Universidad Simón Bolívar. <https://bonga.unisimon.edu.co/handle/20.500.12442/8280>
- Amazon Machine Learning. (2020). Validación cruzada. https://docs.aws.amazon.com/es_es/machine-learning/latest/dg/cross-validation.html
- Aquino, A. A., Molero-Castillo, G., & Rojano, R. (2018). Hacia un nuevo proceso de minería de datos centrado en el usuario. *Pistas Educativas*, 36(114). <https://pistaseducativas.celaya.tecnm.mx/index.php/pistas/article/view/303/0>
- Barrientos Martínez, R. E., Cruz Ramírez, N., Acosta Mesa, H. G., Rabatte Suárez, I., Pavón León, P. & Blázquez Morales, S. L. (2009). Árboles de decisión como herramienta en el diagnóstico médico. *Revista Médica de la Universidad Veracruzana*, 9(2). <https://www.medigraphic.com/pdfs/veracruzana/muv-2009/muv092c.pdf>
- Bean, J. P. & Metzner, B. (1985). A Conceptual Model of Non-traditional Undergraduate Student Attrition. *Review of Educational Research*, 55(4). <https://journals.sagepub.com/doi/10.3102/00346543055004485>

- Blackmist. (2022). Evaluación de los resultados de los experimentos de aprendizaje automático automatizado—Azure Machine Learning. <https://learn.microsoft.com/es-es/azure/machine-learning/how-to-understand-automated-ml>
- Boser, B. E., Guyon, I. & Vapnik, V. (1992). A Training Algorithm for Optimal Margin Classifiers. *COLT '92: Proceedings of the fifth annual workshop on Computational learning theory July 1992*. (pp. 144–152). <https://doi.org/10.1145/130385.130401>
- Chique Rodríguez, C. (2 de septiembre del 2020). Máquina de soporte vectorial (SVM). *Medium*. <https://medium.com/@csarchiquerodriguez/maquina-de-soporte-vectorial-svm-92e9f1b1b1ac>
- Cuji Chacha, B. R. & Vicente Guerrero, V. X. (2020). *Aplicación de la técnica de minería de datos para la predicción de la deserción estudiantil universitaria*. Universidad Técnica de Ambato. <https://repositorio.uta.edu.ec/8443/jspui/handle/123456789/30892>
- Donoso, S., & Schiefelbein, E. (2007). Análisis de los modelos explicativos de retención de estudiantes en la universidad: una visión desde la desigualdad social. *Estudios Pedagógicos (Valdivia)*, 33(1). <https://doi.org/10.4067/S0718-07052007000100001>
- Facundo Díaz, A. H. (2009). Análisis sobre la deserción en la educación superior a distancia y virtual: El caso de la UNAD - Colombia. *Revista de Investigaciones UNAD*, 8(2), 117. <https://doi.org/10.22490/25391887.639>
- Forero Zea, L. D., Piñeros Reina, Y. F., & Rodríguez Molano, J. I. (2019). Machine Learning for the Identification of Students at Risk of Academic Desertion. En L. Uden, D. Liberona, G. Sanchez, & S. Rodríguez-González (Eds.). *Learning Technology for Education Challenges* (pp. 462-473). Springer International Publishing. https://doi.org/10.1007/978-3-030-20798-4_40
- Gálvez Chambilla, M. B., & Flores Cornejo, K. B. (2015). *Modelo predictivo de deserción universitaria de la carrera de Ingeniería Informática en la Universidad Ricardo Palma*. Universidad Ricardo Palma. <http://repositorio.urp.edu.pe/handle/URP/2031>
- Gamboa, H. & Fred, A. (2002). Designing intelligent tutoring systems: A bayesian approach. En J. Filipe, B. Sharp & P. Miranda (eds.). *Enterprise Information Systems III* (pp. 146-152). Springer Verlag.
- Gil Vargas, L. P., Patiño Nuñez, E. J., Ruge Ruge, I. A., & Jiménez López, F. R. (2021). Influential factors in the desertion of electronic engineering students from UPTC admitted in 2015. *Proceedings - 7th International Symposium on Accreditation of Engineering and Computing Education, ICACIT 2021*. <https://doi.org/10.1109/ICACIT53544.2021.9612503>
- Giovagnoli, P. I. (2002). Determinantes de la deserción y graduación universitaria: una aplicación utilizando modelos de duración. *Documentos de Trabajo*, 37. <http://sedici.unlp.edu.ar/handle/10915/3436>
- Guzmán Ruiz, C., Duran Muriel, D., Franco Gallego, J., Castaño Vélez, E., Gallón Gómez, S., & Gómez Portilla, K. (2009). *Deserción estudiantil en la educación superior colombiana*. Ministerio de Educación Nacional.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD explorations newsletter*, 11(1), 10-18.
- Han, J., Pei, J., & Tong, H. (2022). *Data mining: Concepts and techniques*. Morgan kaufmann.
- Hernández J., R. M. J. y F. C. (2004). *Introducción a la minería de datos*. Editorial Pearson Prentice Hall.
- IBM Documentation (17 de agosto de 2021). Conceptos básicos de ayuda de CRISP-DM. <https://prod.ibmdocs-production-dal-6099123ce774e592a519d7c33db8265e-0000.us-south.containers.appdomain.cloud/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>
- Izaurieta, F., & Saavedra, C. (2000). Redes neuronales artificiales. Departamento de Física, Universidad de Concepción Chile. <https://franciscocruz.org/lectures/sistint/rna.pdf>
- Lee, S. (2014). Learning random forests for segmentation of person in self-portrait photos. *The 18th IEEE International Symposium on Consumer Electronics (ISCE 2014)*, 1-2. <https://doi.org/10.1109/ISCE.2014.6884485>
- Lévy Mangin, J. P., Flórez López, R. & Fernández, J. M. (2008). *Las redes neuronales artificiales*. Netbiblo.
- Martínez Pérez, J. A., & Pérez Martin, P. S. (2022). La curva ROC. *Medicina de Familia. SEMERGEN*, 49(1), 101821. <https://doi.org/10.1016/j.semerg.2022.101821>
- Microsoft. (2022). Optimizar los hiperparámetros del modelo—Azure Machine Learning. <https://learn.microsoft.com/es-es/azure/machine-learning/component-reference/tune-model-hyperparameters>
- Moreira da Silva, D. E., Solteiro Pires, E. J., Reis, A., de Moura Oliveira, P. B., & Barroso, J. (2022). Forecasting Students Dropout: A UTAD University Study. *Future Internet*, 14(3). <https://doi.org/10.3390/fi14030076>
- Paramo, G., & Correa Maya, C. (1999). Deserción estudiantil universitaria. Conceptualización. *Revista Universidad EAFIT*, 35(114), 65-78.
- Peralta, B., Salazar, J., Levano, M., & Nicolis, O. (2021). A causal modelling for desertion and graduation prediction using Bayesian networks: A Chilean case. *2021 IEEE International Conference on Automation/24th Congress of the Chilean Association of Automatic Control, ICA-ACCA 2021*. <https://doi.org/10.1109/ICAACCA51523.2021.9465333>
- Ríos, G., & Darío, R. (2018). *Modelo predictivo de gestión administrativa y deserción estudiantil en programa pre grado adulto trabajador de universidad privada de Lima Metropolitana, año 2017*. Universidad Privada del Norte. <https://repositorio.upn.edu.pe/handle/11537/13723>
- Shin. (20 de mayo del 2020). Comprensión de la matriz de confusión y cómo implementarla en Python. *DataSource.ai*. <https://www.datasource.ai/es/data-science-articles/view-source:https://www.datasource.ai/es/data-science-articles/comprension-de-la-matriz-de-confusion-y-como-implementarla-en-python>

- SPADIES. (2019). Estadísticas de Deserción—Sistemas información. https://www.mineducacion.gov.co/sistemasdeinformacion/1735/w3-article-357549.html?_noredirect=1
- Tinto, V. (1989). Definir la deserción: una cuestión de perspectiva. *Revista de la Educación Superior*, 71. http://publicaciones.anui.es.mx/pdfs/revista/Revista71_S1A3ES.pdf
- UNAD. (2018). Acuerdo 002 del 30 de enero de 2018. Universidad Nacional Abierta y a Distancia. https://sgeneral.unad.edu.co/images/documentos/consejoSuperior/acuerdos/2018/COSU_ACUE_002_20180130.pdf
- UNAD. (2022). Acerca de la UNAD: Misión. <https://informacion.unad.edu.co/acerca-de-la-unad>
- Valle Benavides, A. R. (2017). *Curvas ROC (Receiver-Operating-Characteristic) y sus aplicaciones*. Universidad de Sevilla.
- Vallejo Ballesteros, H. F., Guevara Iñiguez, E., & Medina Velasco, S. R. (2018). Minería de datos. *Mundo de la Investigación y el Conocimiento*, 2(1), 339-349. <https://doi.org/10.26820/recimundo/2.esp.2018.339-349>
- Ventura Apaza, L. A., Rosales Huamani, J. A., Alfaro Bernedo, J. O., & Zamudio Chauca, A. G. (2021). A proposal of Machine Learning model to improve learning process and reduce dropout rate at technical training institutes. *Iberian Conference on Information Systems and Technologies, CISTI*. <https://doi.org/10.23919/CISTI52073.2021.9476370>