

PREDICCIÓN DE LA DESERCIÓN ESTUDIANTIL EN CARRERAS DE INGENIERÍA INFORMÁTICA

Línea Temática: Teorías y factores asociados a la permanencia y el abandono. Tipos y perfiles de abandono

Niurys Lázaro-Alvarez, Universidad de las Ciencias Informáticas, nlazaro@uci.cu

Zoraida Callejas Carrión, Universidad de Granada, zoraida@ugr.es

David Griol Barres, Universidad de Granada, dgriol@ugr.es

Resumen. La demanda de graduados en las ramas de Ciencia, Tecnología, Ingeniería y Matemáticas (STEM, por sus siglas en inglés) ha experimentado un incremento debido al creciente desarrollo científico-tecnológico y sus aplicaciones. Sin embargo, las estadísticas recientes muestran que la deserción en estas carreras es alta y existe gran preocupación por elevar la permanencia de los estudiantes en la educación superior. El trabajo que se presenta muestra los resultados de la primera fase de una investigación doctoral realizada por los autores que tienen como objetivo: identificar los factores que determinan la deserción estudiantil en carreras de Ingeniería Informática. Como resultado de una sistematización rigurosa de la bibliografía se obtiene un modelo teórico de deserción en carreras de Ingeniería Informática. Posteriormente se determinan los factores predictivos de deserción estudiantil mediante un estudio cuantitativo correlacional, multivariado y predictivo, empleando una muestra representativa de estudiantes de todas las provincias cubanas matriculados en una carrera del perfil de la Informática. Las variables utilizadas se agrupan en: previas al ingreso y posteriores al ingreso a la Educación Superior. Se obtuvieron como factores predictivos de la deserción estudiantil: el rendimiento académico en Matemática y Programación, la nota en el examen de ingreso en Matemática, la opción en que solicitó la carrera y provincia de procedencia. La metodología de estudio puede ser replicada en otros contextos e incluir nuevas variables.

Descriptorios o Palabras Clave: deserción estudiantil, ingeniería informática, metodología cuantitativa, predicción.

1. Introducción

El tema desarrollado es de actualidad y se relaciona directamente con el Objetivo 4 de los Objetivos de Desarrollo Sostenible (ODS) “Garantizar una educación inclusiva y equitativa de calidad y promover oportunidades de aprendizaje permanente para todos” (Naciones Unidas, 2018, p.27).

La deserción de los estudiantes en la Universidad es una problemática importante, multifactorial y de alcance internacional. A propósito, en la III Conferencia Regional de Educación Superior para América Latina y el Caribe desarrollada en 2018, se analizó la contradicción entre el alto interés en el ingreso y la limitada atención a la deserción (Gacel-Ávila *et al.*, 2018). En consecuencia, se derivó un Plan de Acción para el periodo del 2018 al 2028.

Una de las estrategias indicativas definidas en dicho plan forma parte de la justificación de la presente investigación. Se trata de la estrategia indicativa 5.1.2 que plantea: “Diseñar e implementar sistemas de seguimiento de estudiantes para identificar las dificultades (intra-institucionales y extra-

institucionales) que llevan a la deserción temporaria o permanente” (UNESCO-IESALC, 2018, p.36). La investigación que se somete a su consideración tributa al cumplimiento de esta estrategia.

La deserción estudiantil es un desafío muy relevante para las instituciones de Educación Superior. Son estudios necesarios en áreas de STEM y escasos en carreras de Ingeniería Informática.

En esta investigación se ha propuesto un modelo teórico de deserción estudiantil en carreras de Ingeniería Informática basado en una revisión profunda y sistemática sobre la deserción estudiantil en la Educación Superior, con mayor énfasis en carreras de áreas STEM. Los autores publicaron un resumen de dicho estudio teórico en la VII Conferencia Latinoamericana sobre el Abandono en la Educación Superior (CLABES) (Lázaro, *et al.*, 2017).

Los resultados científicos en la investigación del doctorado para el desarrollo del área del conocimiento impactan en la gestión universitaria pues aporta diversas técnicas estadísticas y de clasificación automática para detección de factores predictivos de deserción estudiantil y estudiantes en riesgo de abandono.

Se ha realizado un estudio de campo riguroso y abarcador combinando múltiples técnicas cuantitativas, que ha permitido identificar (a partir de datos reales a los que se tuvo acceso) factores predictivos de la deserción estudiantil, utilizando una muestra de estudiantes de todas las provincias cubanas matriculados en una carrera del perfil de la Informática.

2. Modelo teórico de deserción en carreras de Ingeniería Informática

Se realizó una búsqueda en la Web of Science, donde se encontraron solo 46 estudios enfocados en la deserción en carreras STEM hasta el 31/05/2019. Luego de eliminar los que están relacionados con cursos en la modalidad a distancia y que no analizan factores de deserción, quedaron finalmente 25 artículos a analizar relacionados con el estudio de causas de deserción en Informática. Dichos artículos se clasificaron según las variables estudiadas en los modelos iniciales en sociológicos, psicológicos, interaccionista, organizacional o una combinación de ellos.

La sistematización de los estudios analizados en los capítulos 2 y 3 de la investigación doctoral (Lázaro, 2020) mediante los métodos histórico lógico y análisis síntesis, sirvió de base para la propuesta de un Modelo Teórico de Deserción Estudiantil en Carreras de Perfil Ingeniería Informática, que se muestra en la Figura 1. Dicho modelo se utilizó para realizar el planteamiento de las hipótesis de investigación, a partir de los datos a los que se tuvo acceso en el estudio de campo.

Los factores previos al ingreso en los estudios del grado en Informática (con borde naranja en la Figura 1) se consideran, además de los tradicionales, la formación académica previa en Matemática e Informática, la nota de acceso a la Universidad y la opción en que se solicita la carrera para aquellos sistemas educativos que la realizan, la orientación vocacional previa, el lugar de procedencia, el nivel educativo de los padres, estado civil y las condiciones socioeconómicas de la familia. (Amaya *et al.*, 2015; Lacave *et al.*, 2018; Kori *et al.*, 2016; Vila *et al.*, 2019).

Como factores que ocurren dentro de la institución de Educación Superior, después del ingreso a los estudios de Informática se consideran entre otros: el contenido curricular de las asignaturas de Matemática e Informática; la dedicación y el desarrollo de métodos y habilidades de estudio en los estudiantes; el apoyo docente para facilitar el aprendizaje en Matemática y Programación; la orientación profesional docente por la Informática; las competencias en la utilización de las TIC de profesores y estudiantes; el apoyo docente educativo para facilitar la adaptación; el hecho de vivir alejado de la familia en una residencia estudiantil y la ayuda económica institucional (Alzen *et al.*, 2018; Canedo *et al.*, 2018; Ferrero, y Oloriz, 2016; Kori *et al.*, 2018).

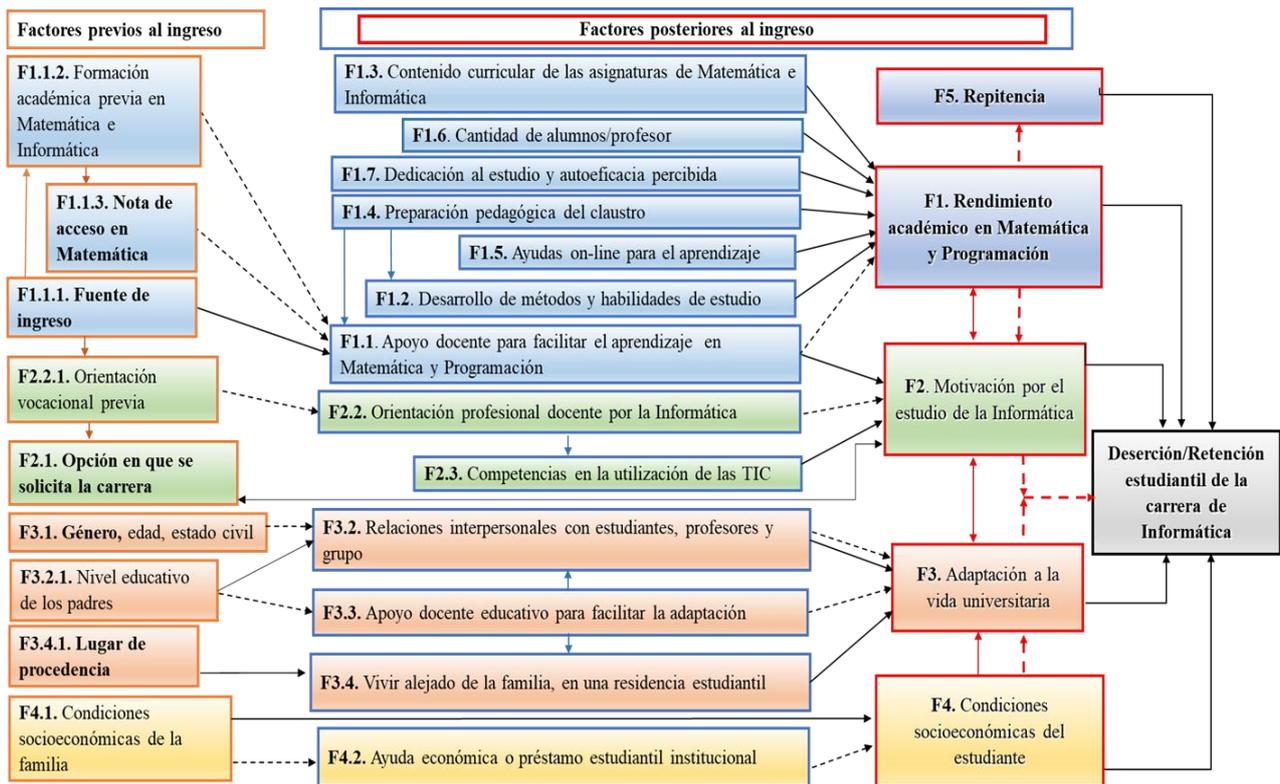


Fig. 1. Modelo de deserción estudiantil en carreras de Ingeniería Informática. Nota: Vertical desde la izquierda, en la primera columna aparecen los factores previos al ingreso (borde naranja) en la segunda y tercera columna (borde azul y rojo respectivamente) los factores posteriores al ingreso y en la tercera columna los factores directa y definitivamente inciden en la deserción. Horizontal los rectángulos de fondo azul son factores académicos, los de fondo verde factores motivacionales, de fondo rojo personales y de fondo naranja socioeconómicos. Fuente: (Lázaro, 2020).

2.1 Hipótesis de investigación

De entre las variables identificadas en el Modelo Teórico propuesto, se tuvo acceso en el Sistema de Gestión Universitaria a las que aparecen en las hipótesis planteadas.

Hipótesis 1 (H1). El *género*, el *lugar de procedencia*, la *opción de carrera*, la *fuentes de ingreso* y la *nota en el examen de ingreso en Matemática (NEIM)*, inciden en la *deserción estudiantil* en carreras de Ingeniería Informática.

Hipótesis 2 (H2). La NEIM incide directamente en el *rendimiento académico en Matemática y Programación (RAMP)*.

Hipótesis 3 (H3). El RAMP incide inversamente en la deserción estudiantil en carreras de Ingeniería Informática.

Hipótesis 4 (H4). La *repitencia de un año académico* incide directamente en la deserción estudiantil en carreras de Ingeniería Informática.

Para el análisis del cumplimiento de éstas hipótesis se realizó un estudio cuantitativo.

3. Descripción de la metodología

El estudio cuantitativo de la investigación es correlacional, predictivo y multivariado. Se utiliza análisis descriptivo, de regresión logística, análisis discriminante y de clasificación automática.

En primer lugar, se realiza un análisis descriptivo de las variables en estudio para inferir características de las mismas en la población, así como un análisis preliminar de la correlación existente entre las variables en estudio.

En segundo lugar, se utiliza la regresión logística con el objetivo de obtener el modelo matemático que sirva para calcular la probabilidad de que un estudiante cause baja de la institución, a partir de los valores en las diferentes variables independientes en estudio, que sean incluidas en el modelo finalmente obtenido, con ello, predecir si un estudiante está en riesgo de deserción.

Los modelos de regresión logística son modelos estadísticos en los que se desea conocer la relación entre una variable dependiente cualitativa y una o más variables explicativas independientes, ya sean cualitativas o cuantitativas. Esta técnica permite modelar cómo influye en la probabilidad de aparición de un suceso, habitualmente dicotómico (Hipótesis 1, 3 y 4) la presencia o no de diversos factores y su valor o nivel.

En el caso de la Hipótesis 2 se utilizó la regresión logística multivariada, pues la variable dependiente RAMP no es dicotómica, sino que toma tres valores: alto, medio o bajo. El programa estadístico SPSS permite obtener los coeficientes que se utilizan en dicho modelo. Las variables se introdujeron siguiendo el procedimiento paso a paso, en un orden lógico, según la secuencia temporal de ocurrencia; es decir, desde los atributos de preinscripción, hasta los resultados de aprendizaje en el primer año, en relación con la permanencia o deserción en el segundo curso.

Posteriormente, se realizó un análisis discriminante para identificar las variables cuantitativas que discriminan o diferencian a los estudiantes en los que desertan de la carrera, repiten el primer año académico, o promueven; así como, para crear una función discriminante capaz de clasificar de forma más precisa a los estudiantes a partir de conocer los datos de las variables que discriminan. En este método la variable dependiente debe tener dos o más grupos de clasificación y las variables independientes (discriminantes) son variables cuantitativas.

Para proveer mayor rigor en el análisis cuantitativo, se realizó un segundo estudio que consistió en la utilización de técnicas de aprendizaje automática proporcionados por el software Weka. Se aplicaron Árbol de Decisión J48 y Red Neuronal Multicapa con el objetivo de detectar los estudiantes en riesgo de deserción o repetir el año académico, lo más temprano posible. Se analizan los valores de exactitud, exhaustividad, precisión y el valor F.

3.1 Población y muestra

La población seleccionada para el estudio está compuesta por los 1.022 estudiantes de todas las provincias cubanas, que matricularon el primer año en carreras de perfil Ingeniería Informática en el curso académico 2013-2014. La muestra fue seleccionada de forma intencional, eliminando aquellos estudiantes que carecían de al menos una de las variables en estudio. Formada por $N=485$ estudiantes de nuevo ingreso de todas las provincias matriculados en una de éstas carreras, lo que representa el 48% de los estudiantes matriculados en primer año ese curso en todo el país en carreras de dicho perfil. De ellos, 178 (37%) son mujeres y 307 (63%) hombres.

4. Principales resultados

De las cuatro hipótesis de investigación planteadas, se ha aceptado parcialmente la primera, se han aceptado totalmente la segunda y tercera, y se ha rechazado la cuarta.

En concreto, como muestra la Tabla 1, para la hipótesis H1 se obtuvo como resultado de la correlación y la regresión logística que las variables previas al ingreso: provincia de procedencia, opción en que solicitó la carrera y NEIM son predictivas de la deserción estudiantil, no siendo así, el género y la fuente de ingreso.

Tabla 1. Resultados de correlación de Spearman entre las variables género, provincia de procedencia, opción en que solicitó la carrera, fuente de ingreso y NEIM con la deserción estudiantil en carreras de Ingeniería Informática. Fuente: Elaboración propia.

Rho de Spearman
con Deserción (r_s)

Género	Coefficiente de correlación	-,049
	Sig. (bilateral)	,280
Provincia de procedencia	Coefficiente de correlación	,191**
	Sig. (bilateral)	,000
Fuente de ingreso	Coefficiente de correlación	,041
	Sig. (bilateral)	,368
Opción en que solicitó la carrera	Coefficiente de correlación	-,094*
	Sig. (bilateral)	,039
NEIM	Coefficiente de correlación	,228**
	Sig. (bilateral)	,000

** La correlación es significativa en el nivel de ,01 (dos colas)

* La correlación es significativa en el nivel de ,05 (dos colas)

En la hipótesis H2 se obtuvo que la NEIM se relaciona directamente con el RAMP y, que el RAMP bajo y medio se puede predecir a partir de la NEIM.

En la hipótesis H3 existe una fuerte relación inversa entre el RAMP y la deserción en carreras de perfil Ingeniería Informática. En los casos donde más alto es el rendimiento, más baja la deserción y, donde más bajo es el rendimiento académico, más alta es la deserción.

La hipótesis H4 se rechaza por no ser significativo en esta cohorte, no existe correlación entre la repitencia y la deserción; lo que coincide en el análisis predictivo.

La regresión logística predefinida con un 95% de confianza permitió identificar como factores predictivos de deserción estudiantil en el primer año de carreras de perfil Ingeniería Informática, en orden descendente: el RAMP, la NEIM y la provincia de procedencia con una significación por debajo de $p < ,05$ y se obtuvieron funciones logísticas para predecir la deserción a partir de dichos factores. Asimismo, en la Tabla 2 se identificó que la variable opción de carrera con un valor de significación $p = ,089$ debe tenerse en cuenta.

Tabla 2. Resultado de regresión logística de las variables género, opción de carrera, fuente de ingreso, provincia de procedencia con la variable dependiente deserción estudiantil con el SPSS. Nota: Se eliminaron las provincias y fuentes de ingreso no significativas. Fuente: Elaboración propia.

	Variables en la ecuación			
	B	Error estándar	gl	Sig.
NEIM	,067	,014	1	,000
Opción de carrera	,101	,059	1	,089
Provincia			14	,019
Provincia(1)	-1,633	,691	1	,018
Provincia(3)	-1,860	,644	1	,004
Provincia(4)	-1,792	,642	1	,005
Provincia(8)	-2,136	,545	1	,000
Provincia(11)	-1,332	,550	1	,015
Provincia(13)	-1,203	,525	1	,022
Género(1)	,357	,268	1	,183
Fuente de ingreso			3	,335
Constante	-3,097	1,305	1	,018

La NEIM puede predecir el RAMP bajo o medio con una significación $p < ,01$. Las funciones de regresión obtenidas permiten inferir que la mayor probabilidad para que un estudiante tenga bajo rendimiento es que tenga la NEIM en el rango de 60 a 69, y es de 83%.

Mediante el análisis discriminante se comprobó que el RAMP, la NEIM y la opción en que solicitó la carrera pueden predecir si un estudiante abandona, repita el primer año académico, o promueve a segundo año como muestra la Figura 2. Se obtuvo una función discriminante que permite clasificar

los estudiantes en cualquiera de las tres categorías (abandona o causa baja, repite y promueve), con un 91,5% de clasificados correctamente.

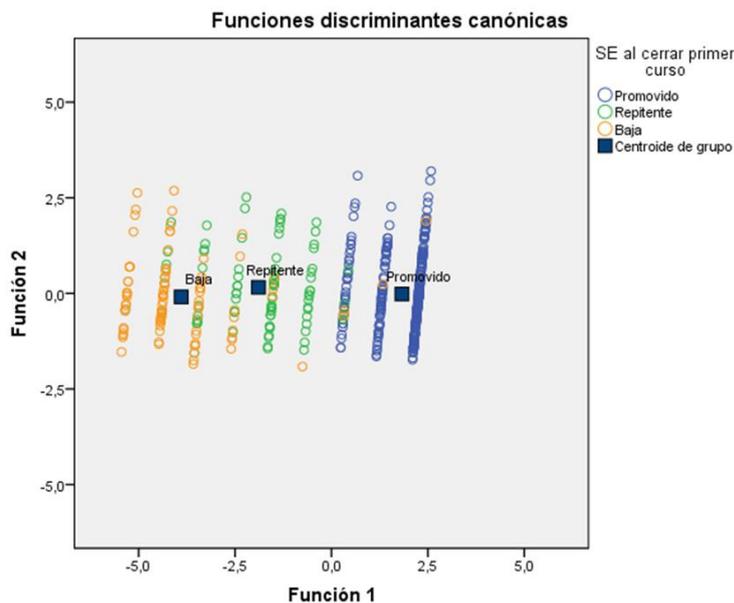


Fig. 2. Representación del diagrama de dispersión de las Funciones Discriminantes 1 y 2 por grupos de clasificación

Este estudio en el contexto cubano corrobora algunos resultados de estudios previos y también arroja resultados novedosos. Con respecto a la variable NEIM el resultado indica que esta es una variable muy importante tanto para el RAMP en primer año como para la deserción. Coincidiendo con los resultados obtenidos por diversos autores en otros contextos (Kori et al., 2018; Salazar-Fernández *et al.*, 2019).

La variable género no resultó predictiva de la deserción en el contexto cubano, lo que no coincide con el resultado de (Miliszewska et al., 2006) en el contexto australiano. Igual ocurre en el caso de la provincia de residencia como factor predictivo; esta variable fue estudiada por (Lacave *et al.*, 2018) y no resultó predictiva dentro del perfil de deserción obtenido.

Los resultados experimentales de clasificación automática descritos en la sección 6.4 de la tesis (Lázaro, 2020) muestran un rendimiento considerable en los árboles de decisión obtenidos con dos clases que utilizan todas las características analizadas con un 97,7% de precisión, donde las variables relacionadas con los resultados académicos son los más relevantes.

Además, al usar solo las características disponibles al momento de la inscripción y después del primer semestre, también se obtuvieron resultados muy positivos: 65% y 95% de exactitud (*accuracy*) respectivamente, con una alta exhaustividad (*recall*) de los estudiantes que no promueven. Coincidiendo con trabajos similares (Nagy y Molontay, 2018; Amaya *et al.*, 2015).

5. Conclusiones

La deserción estudiantil es un desafío muy relevante para las instituciones de Educación Superior. Son estudios necesarios en áreas de STEM y escasos en carreras de Ingeniería Informática. El análisis de correlación, regresión logística y discriminante ha permitido identificar los siguientes factores predictivos de la deserción estudiantil: *RAMP*, *NEIM*, *opción en que solicitó la carrera* y *provincia de procedencia*. Mediante el aprendizaje automático, se identificaron las características que permiten la detección de estudiantes en riesgo lo más temprano posible, luego de su matrícula inicial, con alto rendimiento en los árboles de decisión J48. Ello demuestra que es posible obtener

una evaluación temprana de los estudiantes en riesgo de deserción estudiantil, lo que puede ayudar a definir mecanismos de prevención más exitosos en el contexto cubano. La metodología de estudio puede ser replicado en otros contextos e incluir nuevas variables.

Agradecimientos

Se agradece a la Asociación Universitaria Iberoamericana de Posgrado (AUIP) por el Programa de Formación de Doctores en Educación, con énfasis en Tecnologías educativas. Convenio específico de colaboración entre AUIP, la UCI de Cuba, UGR y US de España.

Referencias

- Alzen, J. L., Langdon, L. S., y Otero, V. K. (2018). A logistic regression investigation of the relationship between the Learning Assistant model and failure rates in introductory STEM courses. *International Journal of STEM Education*, 5(1), 1–12. <https://doi.org/10.1186/s40594-018-0152-1>
- Amaya, Y., Barrientos, E., y Heredia, D. (2015). Student Dropout Predictive Model Using Data Mining Techniques. *IEEE Latin America Transactions*, 13(9), 3127–3134. <https://doi.org/10.1109/TLA.2015.7350068>
- Canedo, E. D., Santos, G. A., y Leite, L. L. (2018). An Assessment of the Teaching-Learning Methodologies Used in the Introductory Programming Courses at a Brazilian University. *Informatics in Education*, 17(1), 45–59. <https://doi.org/10.15388/infedu.2018.03>
- Ferrero, E. L., y Oloriz, M. G. (2016). Aplicación de estrategias motivacionales para mejorar la enseñanza de matemática introductoria en la educación superior. *Congresos CLABES*. Recuperado el 25 de marzo de 2020 de: <https://revistas.utp.ac.pa/index.php/clabes/article/view/1166>
- Gacel-Ávila, J. (coord.). (2018). *La educación superior, internacionalización e integración regional de América Latina y el Caribe* (P. H. Guajardo y H. Juri (eds.)). UNESCO-IESALC. Recuperado el 25 de marzo de 2020 en: http://erasmusplusriesal.org/sites/default/files/libro_la_es_inter_e_integracion_cres.pdf
- Kori, K., Pedaste, M., Altin, H., Tõnisson, E., y Palts, T. (2016). Factors That Influence Students' Motivation to Start and to Continue Studying Information Technology in Estonia. *IEEE Transactions on Education*, 59(4), 255–262. <https://doi.org/10.1109/TE.2016.2528889>
- Kori, K., Pedaste, M., y Must, O. (2018). The Academic, Social, and Professional Integration Profiles of Information Technology Students. *ACM Trans. Comput. Educ.*, 18(4), 20:1-20:19. <https://doi.org/10.1145/3183343>
- Lacave, C., Molina, A. I., y Cruz-Lemus, J. A. (2018). Learning Analytics to identify dropout factors of Computer Science studies through Bayesian networks. *Behaviour & Information Technology*, 1–15. <https://doi.org/10.1080/0144929X.2018.1485053>
- Lázaro, N., Callejas, Z., Griol, D., y Durán, M. (2017). La deserción estudiantil en educación superior: S.O.S. en carreras de Ingeniería Informática. *Congresos CLABES*. Recuperado el 25 de marzo de 2020 en: <http://revistas.utp.ac.pa/index.php/clabes/article/view/1674>
- Lázaro, N. (2020). *Acciones tutoriales con TIC atendiendo a factores predictivos de la deserción estudiantil en carreras de Ingeniería Informática* (Tesis doctoral). [Universidad de Granada]. <https://digibug.ugr.es/handle/10481/64571>
- Miliszewska, I., Barker, G., Henderson, F., y Sztendur, E. (2006). The Issue of Gender Equity in Computer Science – What Students Say. *Journal of Information Technology Education*, 5. Recuperado el 25 de marzo de 2020 en: <https://cutt.ly/Kr8Fusz>
- Naciones Unidas (2018). *La Agenda 2030 y los Objetivos de Desarrollo Sostenible: una oportunidad para América Latina y el Caribe* (LC/G.2681-P/Rev.3), Santiago. Recuperado el 25 de marzo de 2020 en: <https://www.cepal.org/es/publicaciones/40155-la-agenda-2030-objetivos-desarrollo-sostenible-oportunidad-america-latina-caribe>
- Nagy, M., y Molontay, R. (2018). Predicting Dropout in Higher Education based on Secondary School Performance. *2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES)*, 389–394. <https://doi.org/10.1109/INES.2018.8523888>
- Salazar-Fernández, J. P., Sepúlveda, M., y Muñoz-Gama, J. (2019). Influence of Student Diversity on Educational Trajectories in Engineering High-Failure Rate Courses that Lead to Late Dropout. *IEEE Global Engineering Education Conference (EDUCON)*, 607–616. Recuperado el 25 de marzo de 2020 en: <https://ieeexplore.ieee.org/document/8725143>
- UNESCO-IESALC. (2018). *Plan de Acción 2018 - 2028. III Conferencia Regional de Educación Superior para América Latina y el Caribe*. Recuperado el 25 de marzo de 2020 en: <https://www.iesalc.unesco.org/2019/02/22/plan-de-accion-cres-2018-2028/>
- Vila, D., Cisneros, S., Granda, P., y Ortega, C. (2019). Detection of Desertion Patterns in University Students Using Data Mining Techniques: A Case Study. En B.T.M., P.G., Z.P.M., D.M., & Z.S.M. (eds.), *Technology Trends. CITT 2018. Communications in Computer and Information Science*. 895, 420–429. Springer, Cham. https://doi.org/10.1007/978-3-030-05532-5_31