

El aprendizaje automático en la educación superior como herramienta para optimizar los índices de deserción y detectar los factores que la ocasionan

Machine learning in higher education as a tool to optimize dropout rates and detect the factors that cause them

William Ruiz Martínez*

Recibido: 13 - 03 - 2020 / Aceptado: 21 - 04 - 2020 / Publicado: 26 - 06 - 2020

Resumen:

La deserción escolar se ha convertido en una de las grandes preocupaciones de las instituciones de educación superior ya que de la permanencia de sus estudiantes depende en cierto grado su propia sustentabilidad económica. Para nadie es un secreto que la actual situación económica del país, y otros factores como la reticencia de los estudiantes a cursar programas educativos extensos, cada vez llaman menos la atención de los estudiantes. Atenuantes como el difícil y exigente panorama laboral y los elevados costos de matrículas de ciertas instituciones educativas contribuyen a desincentivar el ya minado y reducido mercado estudiantil. Es por ello que las universidades deben reinventarse y plantear nuevas alternativas y métodos de educación donde la realidad laboral coincida y sea congruente con los conocimientos impartidos en las aulas de clase. Para poder estar seguro de cuáles son las causas de la deserción escolar no basta con presumir, sino que se deben comprobar los verdaderos factores de la deserción. Consecuentemente el objetivo del presente artículo es presentar una revisión de diferentes estudios relacionados con la deserción escolar en ámbitos de educación superior y establecer cómo se puede predecir dicha deserción basándose en la aplicación de diferentes modelos y algoritmos de aprendizaje.

Abstract:

School drop-out has become one of the main concerns of higher education institutions since the permanence of their students depends to a certain extent on their own economic sustainability. It is not a secret that the current economic situation in the country, and other factors such as the reluctance of students to pursue extensive educational programs, are less and less striking. Mitigating factors such as the difficult and demanding job landscape and the high tuition costs of certain educational institutions contribute to discourage the already undermined and reduced student market. Therefore it is necessary to reinvent universities and plan new alternatives and methods of education where the market reality coincides and is consistent with the knowledge taught in classrooms. In order to be sure of the causes of dropping out it is not enough to assume the relevant factors, but the true factors must be verified, which is why the objective of this article is to present a review of different studies related to dropping out of schools of higher education and establish how such dropout can be predicted based on the application of different learning models and algorithms.

Palabras clave:

datos, inteligencia, negocios, modelos predictivos, transformación.

Cómo citar: Ruiz Martínez, W. (2020). El aprendizaje automático en la educación superior como herramienta para optimizar los índices de deserción y detectar los factores que la ocasionan. *Hashtag*, (16), 8-16

Keywords: data, intelligence, business, predictive models, transformation

Declaración de conflictos

de interés: el autor declara no tener ningún conflicto de interés

* Magister en dirección estratégica de Ingeniería de software, UNINI, Especialista en Gerencia de proyectos, Ingeniero de Sistemas, Docente asociado del programa de Ingeniería de Sistemas, miembro del grupo de investigación Axón, william_ruizmar@cun.edu.co - Bogotá/Colombia

Introducción

Una de las grandes problemáticas que se vienen presentando en los distintos programas de educación superior en todo el territorio es la deserción escolar; Dicho factor se ha convertido en una de las principales causas de que muchos de nuestros educandos no continúen sus estudios, claro que es pertinente aclarar que este no es el único factor asociado a las altas tasas de deserción escolar. Por otra parte tenemos otros factores: bajos ingresos, altos costos en las matrículas, imposibilidad de acceso a créditos educativos equitativos y justos, poco o nulo acceso al exigente y discriminativo mundo laboral, falta de claridad al seleccionar el programa de estudio, desinterés y poca motivación en el desarrollo de actividades escolares, etc. Es importante aclarar que cuando un estudiante deserta de un programa académico no solo es un impacto económico y motivacional para él no poder optar por un título tecnológico o profesional, sino que además representa un impacto que causa desequilibrio y desmotivación para el alumno. Por otra parte, también se considera un fracaso para la institución de educación superior, el perder un alumno que más tarde hubiera podido ser un gran profesional aportando conocimientos y experiencias no solo al país, sino a la institución donde se formó. Por razones como las anteriormente presentadas se hace necesario adoptar prácticas y metodologías de apoyo en base al contexto del alumno teniendo en cuenta aspectos como su ambiente familiar y su entorno económico y social. Para expertos como el PhD Dursun Delen, de la Oklahoma State University, "Es evidente que se debe apuntar hacia la retención estudiantil como un mecanismo que haga parte de muchos modelos de gestión de matrículas, debido a que este se convierte en un factor de influencia en las clasificaciones de las universidades a nivel local, regional, nacional e internacional, además de afectar la reputación y sus estados financieros". (Delen, 2010)

El profesor Delen se basó en una muestra de cinco años de datos institucionales y desarrolló una se-

rie de modelos que pretendían analizar y explicar las posibles causas que incidían en la deserción de los estudiantes de primer año. Entre sus principales conclusiones se pudieron evidenciar variables como las educativas, en términos de no asimilar contenidos y temáticas por parte de los estudiantes, y como las financieras, que no permitían dar continuidad a los estudios o que no permitían que el dinero para financiar los estudios llegara o se consiguiera en los tiempos estipulados por las instituciones. Estas variables anteriormente descritas se consideraban los principales predictores en las deserciones.

Este tipo de estudios cuenta con varias aplicaciones de tipo práctico. Con un algoritmo apropiado, aunados a la entrega adecuada de datos, el Machine Learning permite:

- Detección anticipada de estudiantes que presentan alto riesgo de deserción.
- Determinar y clasificar los factores de riesgo que ocurren con más frecuencia.
- Diseñar y presentar informes sobre las principales causas y factores de riesgo que inciden en la deserción estudiantil.

Lastimosamente la deserción estudiantil es una situación que no se puede evitar del todo, con la que toda institución de educación superior tiene que aprender a convivir y que se puede ver representada a través de indicadores de deserción. En unas instituciones estos indicadores pueden ser más altos que en otras, pero al fin y al cabo muestran que los alumnos han dejado de asistir a clase; este tipo de situaciones causan de un modo u otro un impacto en los resultados que obtienen estas instituciones de educación superior en cada periodo académico y terminan por convertirse en una avalancha de efectos colaterales para futuras adecuaciones, proyección de nuevos espacios de clase, ampliaciones de infraestructura y otros atenuantes de gran significado para la institución

y los estudiantes. Sin embargo, se trata de un reto que presenta diversas variantes y como tal debe de ser abordado. En este caso, los enfoques basados en el análisis de datos y en los modelamientos de tipo predictivo han confirmado ser efectivos para identificar posibles causas o factores de índole temprana que pueden causar afectaciones o incidir en la deserción escolar de modo tal que se puedan diseñar y elaborar planes de acción que permitan identificar a los alumnos que muestran una mayor tendencia hacia la deserción en sus respectivos programas de estudio.

Este proceso toma un primer acercamiento cuando se puede identificar en forma temprana a los alum-

nos que más posibilidades presentan de ser candidatos a desertores. Mediante el entrenamiento de modelos predictivos a través de técnicas de predicción se pueden llegar a identificar patrones de deserción en cada uno de los alumnos. A continuación, se sigue con la definición de acciones de retención que ya se encuentran estandarizadas según las causas, motivos o circunstancias que pueden dar origen a la deserción y se realiza un monitoreo y control a través de tableros de control que permiten resaltar o dar una mayor visibilidad a los estudiantes o grupos que planean desertar y resaltan la efectividad de las acciones aplicadas para que esta situación no ocurra.

Materiales y métodos

Para la obtención de las referencias documentales se utilizaron varios tipos de fuentes. Entre las principales búsquedas bibliográficas se utilizaron bases de datos como: Microsoft Academic, Google escolar, Redalyc y Base. Una vez obtenido el material se procedió a una evaluación y clasificación del

material obtenido y se seleccionaron solo aquellos documentos que tuvieran una relación directa con la aplicación de algoritmos y modelos de aprendizaje para la predicción de posibles factores y causas en la deserción escolar.

La deserción estudiantil y sus factores más relevantes

la deserción estudiantil se puede interpretar como el cese o receso de los estudiantes para con la institución educativa que originalmente los acogió y que surge debido a una combinación de factores que se presentan en diferentes contextos y situaciones, entre los más representativos: La familia, la sociedad, el ambiente escolar y la percepción individual de esta situación. (Guadalupe, Dorantes, y Can Valle, 2019).

De acuerdo con la cita anterior este tipo de contextos en la mayoría de los casos implican un abandono parcial y muchas veces definitivo

de una institución educativa o una universidad después de haberse matriculado y, por lo tanto, impiden la finalización del nivel o grado de estudio para el cual el estudiante había realizado su inscripción. Según lo expuesto por Felizzola (2018) "el abandono o deserción escolares de forma prematura es un fenómeno de gran complejidad que involucra varios factores en el ámbito social, económico, familiar, psicológico y académico del estudiante". Teniendo como base estudios previos del Ministerio de Educación Nacional (MEN), estos indican que entre los factores más concluyentes sobre la deserción se han encontrado los siguientes: género, estrato socioeconómico, nivel educativo

de los padres, ingresos económicos de la familia, clasificación según el SISBÉN, resultados de las Pruebas de Estado Saber 11°, ocupación del joven y número de personas que componen el núcleo familiar. En instituciones de educación superior donde la mayoría de los estudiantes provienen de estratos 1 y 2 esta situación se trata de manejar en el factor económico mediante el ofrecimiento de descuentos bastante representativos en las matrículas hasta del 35%. (Ministerio de Educación Nacional de Colombia, 2009, p. 129)

El MEN define la deserción de la siguiente manera:

“Un tipo de escenario al que se ve abocado un estudiante al realizar un proyecto educativo y no poderlo culminar en forma exitosa, en este tipo de situaciones se considera como desertor a aquella persona que siendo estudiante de una institución de educación superior no ha presentado ningún tipo de actividad de índole académico por el espacio de dos semestres en forma consecutiva, lo cual representaría un año de inactividad académica”.(Ministerio de Educación Nacional de Colombia, 2009, p. 79)

Pueden por lo tanto presentarse dos tipos de deserción: una que se puede encauzar en el factor tiempo y otra que puede ser encausada en el espacio. Con relación al factor tiempo la deserción o abandono escolar se clasifican como: abandono prematuro, deserción temprana y tardía. Con relación a la deserción espacial, esta se puede clasificar en deserción institucional o en deserción del programa interno que se encuentra cursando el estudiante. Si de deserción prematura hablamos, esta puede ser definida como aquella donde el alumno ha sido admitido en un programa de formación en una institución de educación superior, pero aún no se encuentra matriculado. En cuanto a la deserción temprana, esta puede ser considerada como el abandono de los estudios por parte del estudiante en los primeros semestres de la carrera. En cuanto a este último tipo de deserción

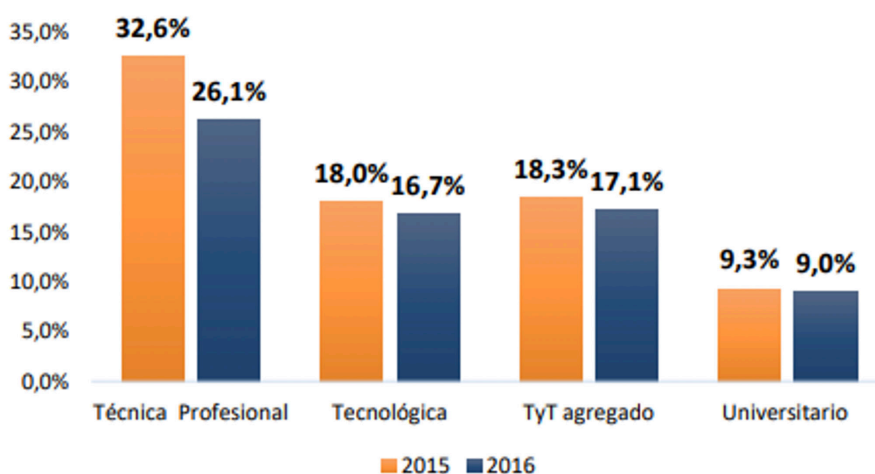
se ha encontrado que muchos de los estudiantes tenían una percepción muy diferente del programa de formación en relación a los conocimientos y temáticas vistas en el aula de clase, o simplemente se dieron cuenta de que el programa de formación no cumplía las expectativas que se habían trazado inicialmente sobre él. La deserción tardía puede ser considerada como el abandono de los estudios en los últimos semestres de la carrera (Tinto, 1982), este tipo de deserción es la menos común y se encuentra relacionada con problemas muy específicos para los cuales es bastante complejo encontrar una solución a corto plazo por parte del estudiante. En vista de la complejidad del problema y las repercusiones a gran nivel que genera en el entorno social, académico y laboral las instituciones de educación superior se encuentran muy interesadas en planear y diseñar un conjunto de estrategias para intervenir en estos casos específicos con la finalidad de disminuir las tasas de deserción. La gran problemática es que muchas de estas estrategias propuestas adolecen de efectividad y certeza en su diseño y aplicación debido a que no tienen en cuenta el origen y diversidad de las causas presentadas. Desde otro punto de vista, se hace necesario que la información presentada sea confiable y precisa de forma tal que permita caracterizar a la población objeto de estudio para identificar los posibles casos de deserción antes de que estos ocurran, esto permite el poder diseñar, planear y ejecutar verdaderas acciones preventivas que realmente permitan disminuir las tasas de deserción en forma eficaz (Felizzola, 2018).

Según el Banco Mundial, el 42% de los jóvenes que logran ingresar a la universidad o a una institución de enfoque técnico o tecnológico terminan por abandonar sus estudios. Es tan preocupante el panorama que esta situación ha llevado a nuestro país a ocupar el para nada honroso segundo lugar en el ranking de deserción en la región de Latinoamérica, superado por Bolivia y seguido muy de cerca por Ecuador y Panamá, en tercer y cuarto lugar respectivamente. (Semana, 2018)

Se presentan propuestas que tratan de dar u ofrecer una solución al problema de la deserción, una de las propuestas consiste en aplicar una evaluación estandarizada y homogeneizada que permita identificar el grado de aprendizaje de sus alumnos. Para poder evaluar los resultados, el 70% de la calificación estaría enfocado en el formato mismo de la prueba y el otro 30% estaría a cargo de cada uno de los docentes mediante el desarrollo de pruebas cortas, trabajos en clase, investigaciones fuera del aula de clase o simplemente el mecanismo de

evaluación que el docente considere más pertinente (Ministerio de Educación Nacional, 2016). Otra de las herramientas propuestas es el crear un mercado de competencias, de modo tal que los docentes puedan demostrar sus capacidades y habilidades, así como su experiencia en la orientación de las temáticas propuestas para desarrollar y orientar en forma idónea un área o asignatura específica. En la figura 2 podemos apreciar las tasas de deserción a nivel nacional por ciclos pedagógicos.

Figura 1: Tasa de deserción por nivel de formación



Fuente: Ministerio de Educación Nacional 2016

El aprendizaje de máquina

El aprendizaje automático o Machine Learning se puede considerar como una disciplina que a través del uso de dispositivos computacionales, puedan aprender a extraer un conjunto de patrones y similitudes que hay en nuestros datos por sí solos, es decir que se plantea que estos modelos permiten al computador a aprender de los mismos datos que se encuentra procesando. La idea es que dichos patrones puedan ser usados posteriormente para predecir comportamientos y puedan tener gran validez en el proceso de toma de decisiones (Telefónica educación digital, 2019).

En resumen, podríamos concluir que el aprendizaje de máquina es una técnica para recopilar y analizar datos que para su funcionamiento se basa en el estudio y forma en que se han comportado datos históricos de transacciones o procesos y su posterior comparación con una serie de patrones o incidencias que permitan establecer o predecir el comportamiento de estos datos en el futuro.

Los autores Russo, Ramón, Alonso, Cicerchia, Esnaola y Tessore Introdusen el concepto de aprendizaje de máquina como una de las ramas

de la inteligencia artificial que abarca una serie de técnicas que posibilitan el aprendizaje automático mediante el entrenamiento a través de grandes volúmenes de datos (Russo et al., 2016). Es por ello que en la actualidad se plantea el uso de diferentes modelos que hacen uso de esta técnica y permiten lograr una precisión en sus resultados muy por encima de los estándares que los seres humanos obtienen en las tareas similares, uno de los ejemplos más dicientes es el del reconocimiento de

objetos en una imagen. La construcción de modelos de aprendizaje automático requiere sus propias adaptaciones debido a situaciones como el origen, formato y naturaleza de los datos o los problemas a los que se aplica el modelo para resolver una problemática en especial. Por tanto, se plantea la necesidad de investigar las diferentes técnicas que permitan obtener resultados precisos y fiables en un tiempo razonable y bajo condiciones similares.

Modelos predictivos

Delen (2010) presentó un enfoque apoyado en los factores de retención tomando como punto de referencia los estudios de marketing y minería de datos "churn analysis", en este análisis se analizaron datos mediante analítica descriptiva en relación con un lapso de cinco años de datos históricos. Para dicho proceso se hizo uso de herramientas de minería de datos tales como las máquinas de soporte vectorial, redes neuronales artificiales, árboles de decisión (C5) y regresión logística, Lo que permitió obtener como resultado una capacidad predictiva promedio del 80% y, consecuentemente, permitió demostrar que factores como los académicos y financieros son los más relevantes en la aplicación de este modelo.

Zhang, Oussena, Clark y Hyeonsook (2010) Llevaron a cabo un sistema de minería de datos que contemplaba aspectos académicos y el análisis de deserción con las variables que presenta Tinto (1982). Para ello recolectaron datos de 3 años de registros universitarios sobre una población de 5,458 estudiantes. En dicho estudio hicieron uso de herramientas de minería de datos, como el algoritmo de Naive Bayes, la máquina de soporte vectorial y el árbol de decisión; posteriormente se pudo concluir que el algoritmo de Naive Bayes fue el que ofreció una mejor precisión con una efectividad del (89,5%).

Heredia, Amaya y Barrientos (2015) Construyeron un modelo predictivo relacionado con la deserción

escolar de un grupo de estudiantes en la universidad Simón Bolívar. El objetivo de dicho modelamiento permitía predecir la probabilidad de que un alumno termina desertando de un programa de estudio. Para el estudio en cuestión los autores utilizaron herramientas relacionadas con minería de datos como los algoritmos ID3, el algoritmo C4.5 y los árboles de decisión, logrando una precisión hasta del 92.9%.

Ji-Wu y Mareboyana (2013) se dieron a la tarea de proponer algoritmos que pudieran ser empleados para monitorear la retención en estudiantes de programas de pregrado. El estudio también permitió realizar un conjunto de mejoras a algoritmos de clasificación ya conocidos como las máquinas de soporte vectorial (SVM), los árboles de decisión y las redes neuronales. De las tres técnicas empleadas en el estudio se pudo obtener un mejor conjunto de resultados con las redes neuronales con una efectividad del 94.16%.

Formia, Lanzarini y Hasperue (2013) presentaron un modelo que se encargaba de predecir la probabilidad de un grupo de estudiantes en la Universidad Nacional de Rio Negro (UNRN), En este modelo se tomó como referente del análisis la sede atlántica de dicha institución y como programa objeto de estudio la licenciatura en sistemas. Para lograr su fin los autores emplearon herramientas de

minería de datos como el algoritmo C4.5 con una tasa de precisión del 71.65%.

Fischer y Ángulo (2012) nos presentan la propuesta de una metodología para identificar mediante modelos de aprendizaje automático los estudiantes que presentaban el riesgo más elevado de deserción en los programas de Ingeniería de la Universidad de las Américas. Para poder llevar a cabo este proyecto se tomó como referente la metodología CRISP-DM, la cual es una metodología especializada en proyectos de ciencia de datos y permite estructurar el proceso en seis fases que interactúan entre ellas mediante un modelo iterativo. En cuanto al modelamiento como

tal se aplicaron redes neuronales, árboles de decisión y clúster K-medianas, los cuales eran los encargados de analizar la conducta de los estudiantes y así mismo permitían la evaluación de factores como el promedio obtenido en la prueba inicial de selección universitaria conocida como PSU. También se tuvieron en cuenta otros datos como el promedio de notas obtenidas en la formación de enseñanza media, la fecha de ingreso a la universidad y el género de los estudiantes. La precisión de los modelos es calculada teniendo como punto de partida un conjunto de datos de prueba, lo que finalmente indicó que ninguno de los modelos predictivos arrojó resultados positivos.

Conclusiones

La deserción estudiantil es una de las grandes problemáticas de las universidades y demás instituciones de educación superior, debido a que de la permanencia de sus estudiantes depende en cierto grado de su propia sustentabilidad económica.

Encontramos una gran diversidad de factores en contextos sociales, familiares y de tipo individual que inciden en la deserción escolar, lastimosamente estos contextos implican un abandono parcial o definitivo de la institución y, lógicamente, pérdida de tiempo y recursos económicos por parte de la familia y el estudiante.

Se pudo evidenciar que muchas de las políticas, planes de acción o estrategias desarrollados por las instituciones de educación superior para erradicar la deserción estudiantil no tienen éxito ya que se enfocan solo en uno o unos de los posibles factores, pero no tienen en cuenta el contexto individual de cada uno de los estudiantes.

El aprendizaje de máquina es una disciplina o área que forma parte de la inteligencia artificial y que busca que los dispositivos que posean algún tipo de capacidad computacional puedan aprender a prede-

cir futuras situaciones o comportamientos mediante el análisis de patrones e información histórica.

El uso de modelos predictivos ha demostrado ser una herramienta invaluable en el proceso de análisis de la deserción mediante la aplicación de diferentes algoritmos, unos con mayor precisión en los resultados que otros. En los diferentes estudios presentados utilizando modelos predictivos se encuentran la aplicación de algoritmos como el de Bayes, o el uso de árboles de decisión, redes neuronales y máquinas de soporte vectorial, con capacidades predictivas entre el 71,65% y el 94.16%.

En cada uno de los modelos y algoritmos presentados, se presenta la posibilidad de que al inicio del estudio los resultados o predicciones arrojadas no sean las ideales, se mejoran estos escenarios dependiendo de los modelos y algoritmos utilizados en cada situación hipotética. Es decir que no siempre un algoritmo o modelo funciona de forma adecuada con un conjunto de datos, es por ello que se deben generar y aplicar diferentes modelamientos y algoritmos de aprendizaje para establecer cuál de ellos es el más ideal en cada contexto específico.

Referencias

- Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems* 49(4), 498–506. doi:10.1016/j.dss.2010.06.003
- Felizzola, H., Jaime, A., Pedroza, F., y Castillo, A. (2018). Modelo de predicción para la deserción temprana en la facultad de Ingeniería de la Universidad de la Salle. En C. Lozano (Presidencia), *Gestión, calidad y desarrollo en las facultades de ingeniería*. Conferencia llevada a cabo en el congreso Encuentro Internacional de Educación Superior ACOFI (págs. 21-29). Cartagena de Indias, Colombia.
- Fischer, E. (2012). *Modelo para la automatización del proceso de determinación de riesgo de deserción en estudiantes universitarios* (Tesis de maestría). Universidad de Chile, Santiago de Chile, Chile.
- Formia, S., Lanzarini, L. C., & Hasperue, W. (2013). *Caracterización de la deserción universitaria en la UNRN utilizando minería de datos*. *Revista Iberoamericana de Tecnología En Educación y Educación En Tecnología*, 11, 92-98.
- Guadalupe, A., Dorantes, A., y Can Valle, A.R. (2019). Factors that influence in school dropout in higher education students from a perspective. *International journal of Advanced Research*, 7(12), 23-29. doi: 10.21474/IJAR01/10119
- Heredia, D., Amaya, Y., y Barrientos, E. (2015). Student Dropout Predictive Model Using Data Mining Techniques. *IEEE Latin America Transactions*, 13(9), 3127-3134.
- Ji-Wu, J., y Mareboyana, M. (2013). Machine Learning Algorithms and Predictive Models for Undergraduate Student Retention. *Proceedings of the World Congress on Engineering and Computer Science*. Conferencia llevada a cabo en el World congress on Engineering & Computer science 2013. San Francisco, Eu.
- Ministerio de Educación Nacional de Colombia. (2009). *Deserción estudiantil en la educación superior colombiana*. Recuperado de https://www.mineduacion.gov.co/sistemas-deinformacion/1735/articles-254702_libro_desercion.pdf
- Ministerio de Educación Nacional. (2016). *Reporte sobre deserción y graduación en educación superior año 2016*. Recuperado de <https://bit.ly/3HbqVm9>
- Russo, C.C., Ramón, H.D., Alonso, N., Cicerchia, B., Esnaola, L., y Tessore, J. P. (2016). Tratamiento masivo de datos utilizando técnicas de machine learning En *XVIII Workshop de Investigadores en Ciencias de la Computación; WICC 2016* (pp. 131-134). Entre Rios, Argentina: Facultad de Ciencias de la Administración.
- Semana. (2018, abril, 11). ¿Cómo puede Colombia reducir la deserción universitaria?. Semana. Recuperado de <https://bit.ly/3oU1wXs>

Telefónica educación digital. (14 de 03 de 2019). MiriadaX. Recuperado de: <https://www.telefonica.com/es/sala-comunicacion/blog/educacion-y-competencias-para-la-era-digital-la-necesidad-de-mejorar-nuestras-habilidades-y-volver-a-capacitarnos/>

Tinto, V. (1982). Limits of Theory and Practice in Student Attrition. *The Journal of Higher Education*, 53(6), 687-700.

Zhang, Y., Oussena, S., Clark, T., y Hyeonsook, K. (2010). Using data mining to improve student retention in higher education. En J. Cordeiro (Presidencia). Ponencia realizada en *International Conference on Enterprise Information Systems ICEIS*. Lisboa