

RECYT

Año 22 / N° 34 / 2020 / 45–50

Modelo para la estimación de la deserción estudiantil Awajún y Wampis empleando minería de datos

Awajún and Wampis Student Dropout Estimation Model Using Data Mining

Lenin Quiñones Huatangari^{1,*}, Diomer M. Jara¹, Nicanor Alvarado¹, Manuel E. Milla¹, Oscar A. Gamarra¹

1- Instituto de Ciencia de Datos. Universidad Nacional de Jaén (UNJ), Perú.

* E-mail: lenin.quinones@unj.edu.pe

Recibido: 01/04/2020; Aprobado: 14/10/2020

Resumen

La deserción estudiantil es un problema complejo y crucial en el campo de la educación, que está presente en todos los niveles y modalidades del sistema educativo, por tanto, la detección temprana es una estrategia clave para las instituciones académicas. La Minería de Datos Educativa permite modelar la deserción de estudiantes considerando sus datos socioeconómicos, académicos y personales. El objetivo del trabajo fue emplear la minería de datos para determinar modelos que estimen la deserción de estudiantes Awajún y Wampis de la Universidad Nacional de Jaén. Se empleó la metodología CRISP-DM: Entender el problema de la deserción del 45% de la población en estudio, comprender las variables, construir la matriz de datos de los cuarenta y nueve estudiantes, modelamiento utilizando el software Weka y evaluación del modelo. Se identificaron cinco variables que influyen en la deserción: Los cursos aprobados, créditos aprobados, comunidad de procedencia, promedio y ciclo de ingreso. Además, se proponen tres modelos con porcentaje de instancias bien clasificadas de 87.8, de los que se concluye que si los cursos aprobados, son inferiores a diez asignaturas o los créditos aprobados son menores a 27, el estudiante se retira de la Universidad.

Palabras clave: Pueblos originarios; Deserción estudiantil, Minería de datos educativa.

Abstract

Student dropout is a complex and crucial problem in the field of education, which is present at all levels and modalities of the educational system. Therefore, early detection is a key strategy for academic institutions. Educational Data Mining allows modeling the dropout of registered students with their socioeconomic, academic and personal data. The objective of the work was to use data mining to determine models that estimate the dropout of Awajún and Wampis students from the National University of Jaén. The CRISP-DM methodology was used to understand the dropout problem of 45% of the study population, understand the variables and build the data matrix of the forty-nine students. The Weka software and model evaluation were used for modelling. Five variables that influence dropout were identified: the approved courses, approved credits, community of origin, average and cycle of admission. In addition, three models are proposed with percentages of well classified instances of 87.8, from which it is concluded that if the approved courses are less than ten subjects or the approved credits are less than 27, the student withdraws from the University.

Keywords: Native peoples; Student dropout; Educational data mining.

Introducción

El Perú posee una gran diversidad cultural, sin embargo, se conoce muy poco de esta. Existen dos pueblos originarios que habitan en la Amazonía del Perú, en la frontera con Ecuador, específicamente al norte de los departamentos de Amazonas, Loreto, Cajamarca y San Martín. Estos pueblos están íntimamente relacionados, debido a que comparten una tradición histórica, cultural y sus idiomas pertenecen a la misma familia lingüística (Jíbaro). Son conocidos por su habilidad como guerreros y la fuerte resistencia frente a distintas poblaciones que

incursionaron en el territorio ocupado por ellos. El pueblo Awajún, también conocido con el nombre de “aguaruna”. Asimismo, el pueblo originario Wampis es conocido con los nombres de “Shuar” o “huambisa” [1].

Por otra parte, la minería de datos es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto [2]. Los algoritmos de uso común de minería de datos [3]: Regresión lineal, regresión logística, árbol de decisión, máquinas de vectores de soporte, red

bayesiana, kNN, K-medias, algoritmos de reducción de dimensión y algoritmos de aumento de gradiente. Debido al empleo de la minería de datos a problemas de Educación, se ha creado la Minería de Datos Educacional (MDE), donde una de las áreas de aplicación permite predecir la deserción estudiantil [4].

En efecto, la deserción estudiantil es un problema complejo y crucial en el campo de la educación, que está presente en todos los niveles y modalidades del sistema educativo. Esto genera daños sociales, económicos, políticos, académicos y financieros a todos los involucrados en el proceso educativo [5]. En este sentido, la razón del trabajo de investigación, fue recolectar información y determinar patrones con respecto a la deserción de los estudiantes de la Universidad Nacional de Jaén (UNJ), que provienen de los pueblos originarios Awajún y Wampis. En consecuencia, se planteó el problema: ¿Empleando técnicas de minería de datos en la información histórica de los estudiantes proveniente de los pueblos originarios Awajún y Wampis de la Universidad Nacional de Jaén, será posible determinar modelos que permitan describir deserción en la misma?

Se han realizado estudios en el área de MDE para predecir la deserción estudiantil, diferenciándose entre ellos, el área de estudio, las técnicas empleadas y las propuestas que han originado: Los autores proporcionan una visión general del fenómeno de predicción de deserción de estudiantes de un curso de Massive Open Online Course (MOOC), cursos online masivos y abiertos, además proponen ideas y recomendaciones que podrían conducir al desarrollo de soluciones útiles y efectivas de aprendizaje automático para resolver el problema de la deserción de MOOC [6]. No obstante, otros autores proponen un plan de tutoría que puede ser usado por las instituciones para reducir el porcentaje de deserción de los cursos virtuales [7]. Incluso se han realizado artículos de revisión sobre el tema: Se analizaron múltiples trabajos, identificando las ventajas y desventajas de las metodologías más utilizadas; análisis correlacionales, análisis de regresión logística, análisis de supervivencia y minería de datos. Dichos resultados son consistentes con los obtenidos por diversos autores a lo largo del tiempo, y en base a ellos se proponen dos tipos de medidas; por un lado, acciones encaminadas a facilitar el diagnóstico respecto al problema de la deserción, y por otro lado, medidas encaminadas a su prevención [8]. Se encontraron las brechas existentes en la predicción de la deserción educativa en todos los artículos científicos publicados en revistas científicas de alto impacto y determinaron los atributos que faltan, lo que puede aportar para una mejor predicción [4].

Se ha evidenciado la efectividad de los modelos de árboles de decisión que evalúa a los nuevos posibles desertores [11, 3, 15], han obtenido reglas que fueron programadas y visualizadas en una interfaz web [10]. Utilizaron los algoritmos de Exponential smoothing state (ETS) y Redes

Neuronales, para la construcción del modelo de minería de datos basado en series de tiempo logrando realizar las predicciones de deserción escolar y el uso de la metodología XP para el desarrollo del sistema [11]. Sobre la base de lo anterior y considerando los estudios previos realizados [12], se concluye que mediante técnicas de minería de datos es posible determinar modelos que estimen la deserción estudiantil universitaria, analizando la historia académica del estudiante junto a los factores socio económicos y otros, que determinan su condición de potencial desertor.

En este contexto, el objetivo del trabajo de investigación fue emplear la minería de datos para determinar modelos que estimen la deserción de estudiantes Awajún y Wampis de la UNJ.

Materiales y métodos

Tipo de Investigación

El tipo de investigación fue descriptivo, ya que se buscó especificar las variables socioeconómicas, académicas y personales que influyen en la deserción del grupo de estudiantes universitarios provenientes de pueblos originarios Awajún y Wampis de la UNJ.

Población, muestra y muestreo

Población y muestra

La población es igual a la muestra, cuando la muestra coincide con la población se está en presencia de una muestra censal. Ha sido el elemento de registro de información socioeconómica, académica y personal de los cuarenta y nueve (49) estudiantes Awajún y Wampis en los períodos 2012 – 2019 en la UNJ.

Muestreo

Se ha empleado un muestreo por conveniencia. El muestreo de o por conveniencia es una técnica de muestreo no probabilístico donde los sujetos son seleccionados dada la conveniente accesibilidad y proximidad de los sujetos para los investigadores [13].

Metodología

La metodología empleada, fue la propuesta por la Cross Industry Standard Process for Data Mining, CRISP-DM [19, 11]. Constó de cinco fases:

Entendimiento del problema: Se realizó una revisión de la literatura del tema, se consultó con profesionales expertos sobre pueblos originarios y minería de datos.

Comprensión de los datos: Los datos fueron recolectados de las oficinas de la UNJ (Admisión, Asuntos Académicos, Archivo, Bienestar Universitario e Informática). Fueron alojados en una base de datos, realizado en una hoja de cálculo de Microsoft Excel.

El fin de la base de datos fue determinar los modelos basados en minería de datos, que nos permitieron describir el comportamiento de deserción estudiantil proveniente de los pueblos originarios Awajún y Wampis de la UNJ. Estuvo constituida por 49 instancias concernientes a la misma cantidad de la muestra. Cada instancia constó de diecisiete (17) variables, ver Tabla 1. Las variables fueron determinadas en base al análisis bibliográfico de trabajos relacionados [14], [15].

Tabla 1: Nombre de las diecisiete variables, definición, tipo y rango.

Variables	Definición	Tipo	Rango
Comunidad	Determina el nombre de la comunidad originaria de procedencia.	Nominal	
Distrito	Determina el nombre del distrito que pertenece el pueblo originario de procedencia.	Nominal	
Provincia	Determina la provincia que pertenece el pueblo originario de donde proviene.	Nominal	
Año nacimiento	Se determina en base a la partida de nacimiento para postular a la Universidad.	Discreta	
Genero	Género del alumno.	Nominal	<ul style="list-style-type: none"> • M: Masculino • F: Femenino
Ingreso colegio	Año de ingreso al nivel secundario.	Discreta	
Termino colegio	Año de finalización del nivel secundario.	Discreta	
Modalidad ingreso	Modalidad de ingreso a la Universidad Nacional de Jaén.	Nominal	<ul style="list-style-type: none"> • CI: Examen extraordinario • AD: Examen regular
Código estudiante	Código único de matrícula del estudiante.	Nominal	
Carrera Profesional	Nombre de la carrera profesional de la Universidad Nacional de Jaén a donde el estudiante pertenece.	Nominal	<ul style="list-style-type: none"> • IIA: Ingeniería de Industrias Alimentarias • IC: Ingeniería Civil • IM: Ingeniería Mecánica • IFA: Ingeniería Forestal y Ambiental • TM: Tecnología Médica
Cursos matemática	Número de veces que ha llevado cursos de matemáticas, incluyendo las repeticiones.	Discreta	
Créditos aprobados	Número de créditos aprobados en la carrera profesional donde pertenece.	Discreta	
Cursos aprobados	Número de cursos aprobados en la carrera profesional donde pertenece.	Discreta	
Cursos desaprobados	Número de cursos desaprobados en la carrera profesional donde pertenece.	Discreta	
Promedio	Promedio acumulado del estudiante durante los semestres cursados.	Continua	
Ciclo ingreso	Ciclo de ingreso del estudiante.	Nominal	<ul style="list-style-type: none"> • 2012-I • 2012-II • 2013-I • 2013-II • 2014-I • 2014-II • 2015-I • 2015-II • 2016-I • 2016-II • 2017-I • 2017-II • 2018-I • 2018-II • 2019-I • 2019-II
Deserción Universidad	Identifica si el alumno desertó de los estudios hasta el semestre 2019-I; esta variable es la que se desea predecir.	Nominal	<ul style="list-style-type: none"> • Estudiante • Retirado

Preparación de los datos: Se obtuvieron datos sin valores nulos o anómalos, que permitieron obtener resultados confiables, para ello se ha utilizado técnicas estadísticas como la media, mediana y la moda. Se transformó la base de datos al formato arff (attribute relation file format) requerido por el software empleado. Existen varios métodos, para seleccionar variables independientes que más cercanamente afectan la variable dependiente; los mismos, que permiten hacer una descripción de los datos a un menor costo y por su importancia, son ampliamente usados en procesos asociados a aprendizaje de máquina [16]. En este sentido, se ha empleado el algoritmo InfoGainAttributeEval de Weka. Por esto, de las dieciséis variables independientes descritas en la Tabla 1, solo se emplearon cinco variables para determinar los modelos, las cuales fueron: Cursos aprobados, créditos aprobados, comunidad de origen, promedio acumulado y ciclo de ingreso. En conclusión, siendo estas variables las que mayor inciden en la variable dependiente, que fue la deserción de los estudiantes.

Modelación: La herramienta de software libre de minería de datos que se empleó en este estudio para determinar los modelos que estimaron la deserción, fue Weka [16]. Para la selección de los algoritmos adecuados; se llevó a cabo la experimentación empleando todos los métodos que posee el software, concerniente a árboles y reglas. Esta selección se ha realizado debido a que estos algoritmos, porque se obtienen reglas de clasificación del tipo “Si – Entonces” o árboles de decisión; además son empleados en trabajos previos [7, 3, 19, 4]. Una regla tiene dos partes, el antecedente (la parte del “Si”) y el consecuente (la parte del “Entonces”). De esta manera, una regla asigna una instancia de datos a la variable señalada por el consecuente (estudiante o deserta) si los valores de los atributos de predicción satisfacen las condiciones expresadas en el antecedente.

Evaluación del modelo: Con el fin de evaluar la calidad y precisión de la predicción de las reglas de clasificación obtenidas se utilizó el método de validación cruzada con 10 pliegues (n-fold cross validation). Se evaluó mediante Weka, a través de los estadísticos: Error cuadrático medio (RMSE) que comparó el valor predicho y el valor conocido, porcentaje de aciertos que está bien clasificado mediante las reglas determinadas (% aciertos), el coeficiente de Kappa permitió medir el grado de concordancia de las evaluaciones nominales realizadas, error absoluto medio (MAE) sirvió para cuantificar la precisión de los modelos comparando los valores predichos frente a los recolectados y el área bajo la curva ROC (AUC) que representó la probabilidad de que el modelo clasifique un ejemplo positivo aleatorio más alto que un ejemplo negativo aleatorio.

Además, se eliminaron las variables redundantes o irrelevantes. Se interpretaron los tres modelos descubiertos para determinar su calidad y consolidar el conocimiento

descubierto, para confrontarlo con conocimiento previamente descubierto. De esta manera se obtuvo patrones útiles en términos que sean entendibles para el usuario.

Resultados

En la Tabla 2, se muestran tres modelos que están basados en técnicas de minería de datos que poseen mejor comportamiento, para estimar la deserción estudiantil Awajún y Wampis. Los mismos que han considerado dos variables independientes (cursos aprobados y créditos aprobados) de las cinco que más influyen en la deserción de estudiantes de los pueblos Awajún y Wampis de la UNJ (cursos aprobados, créditos aprobados, comunidad de origen y promedio acumulado y ciclo de ingreso). Además, se menciona a las técnicas empleadas y las reglas que se han originado por cada modelo.

Tabla 2: Algoritmo empleado y reglas obtenidas de los tres modelos.

Identificador	Algoritmo	Reglas
Modelo 1	J48	Si (cursos aprobados <=10), entonces (deserción universidad=retirado) Si (cursos aprobados >10), entonces (deserción universidad = estudiante)
Modelo 2	Ridor	Si (créditos aprobados<=27), entonces (deserción universidad=retirado) Si (créditos aprobados >27), entonces (deserción universidad = estudiante)
Modelo 3	PART	Si (cursos aprobados <=10), entonces (deserción universidad = retirado) Si (cursos aprobados >10), entonces (deserción universidad = estudiante)

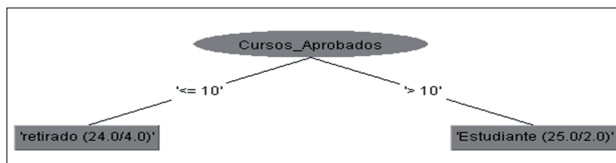


Figura 1: Árbol creado con el algoritmo J48 en el software Weka

En la Tabla 3, se describe el comportamiento de los tres modelos propuestos, para ello se ha empleado la validación cruzada. Los indicadores estadísticos empleados han sido el RMSE, porcentaje de aciertos, Kappa, MAE y AUC.

Tabla 3: Comportamiento de los tres modelos, empleando diversos indicadores estadísticos.

Identificador	RMSE	% aciertos	Kappa	MAE	AUC
Modelo 1	0.3373	87.7551	0.7525	0.2156	0.805
Modelo 2	0.3499	87.7551	0.7525	0.1224	0.876
Modelo 3	0.3373	87.7551	0.7525	0.2156	0.805

Discusión

La población de estudiantes provenientes de pueblos originarios Awajún y Wampis de la UNJ, estuvo conformado por todos los ingresantes desde el semestre 2012-I, semestre en que la universidad inició sus actividades académicas, hasta el 2019-II, conformada por 49 personas.

Los factores que han conllevado a la deserción del 45% (22) de los 49 estudiantes, debe ser analizados a profundidad en trabajos futuros. Se ha podido recabar información de diecisiete variables concerniente a datos personales, demográficos y académicos de la población estudiada, siendo este un estudio piloto. Una tarea importante realizada en este trabajo, fue la recopilación de la información y el pre procesamiento de los datos, ya que la calidad y fiabilidad de la información afecta de manera directa en los resultados. Es una tarea, que ha implicado invertir mucho tiempo y disposición, puesto que se tuvo que realizar la integración de datos de cinco fuentes diferentes para formar la base de datos. En cambio en otros trabajos, las instituciones poseen un data warehouse, el cual les permite tener centralizada la información como en el trabajo de Gallardo [9].

Cada situación de estudio responde a un modelo específico adaptado a la realidad de información de cada región geográfica por sus variables socioeconómicas y educativas. De aquí que el presente estudio ha considerado una población particular y no se debe realizar ninguna inferencia, generalización o expansión de estos tres modelos a otros lugares del sistema universitario nacional. Se realizó un conjunto de experimentos con el objetivo de conseguir predecir con un buen grado de exactitud el estado académico de los estudiantes hasta la fecha mediante la utilización de los algoritmos J48, Ridor, PART y de esta manera obteniendo el modelo 1, modelo 2 y modelo 3. Los algoritmos J48 y Ridor han sido empleados en el trabajo de Vera [17]. De las diferentes variables que Kumar [4] en su estado del arte propone, las que frecuentemente son usadas para predecir la deserción estudiantil. Considera el género, sin embargo, esta variable en el presente estudio, no ha sido considerado por los tres modelos propuestos. Debido a que al emplear el algoritmo InfoGainAttributeEval del software Weka, el mismo que evalúa el valor de un atributo para medir la información consigue con respecto a la clase, determinó que las cinco variables que influyen en la deserción estudiantil, han sido: cursos aprobados, créditos aprobados, comunidad de origen y promedio acumulado y ciclo de ingreso.

El algoritmo J48 descubre un árbol con dos hojas y de tamaño tres, ver Figura 1. De la cual se desprenden las dos reglas del modelo 1. La primera regla nos dice que, si la cantidad de los cursos aprobados es menor o igual que diez, entonces el alumno estaría retirado o propenso a retirarse de la UNJ. La segunda regla menciona que, si los cursos aprobados son mayores que diez, ese alumno sigue siendo estudiante de la UNJ. Interpretando estas dos reglas, el patrón que se desprende, es que los estudiantes de los pueblos originarios que han desertado, no aprobaron más de diez cursos, es decir ni dos ciclos seguidos, debido a que un ciclo tiene siete cursos. Las reglas obtenidas del modelo 3, son lógicamente equivalentes con las dos reglas obtenidas del modelo 1 y que se originan del árbol de decisión respectivo. En las dos reglas que componen el modelo 2, la variable que influye es la cantidad de créditos

aprobados; la primera regla menciona que, si la cantidad de créditos aprobados es menor o igual a veinte y siete, ese alumno ya está retirado o se va a retirar de la UNJ y la segunda regla es que, si el alumno tiene más de veinte y siete créditos aprobados, esta persona va a seguir siendo estudiante de la UNJ.

La evaluación del comportamiento de los tres modelos se ha obtenido mediante diversos estadísticos, ver Tabla 3: Se ha obtenido un porcentaje de instancias bien clasificadas de 87.8, el coeficiente de Kappa 0.7525 que se interpreta como la fuerza de concordancia buena, los valores de RMSE y MAE cercanos a cero indican cuán cerca están los puntos de datos observados de los valores predichos del modelo, finalmente el valor de AUC es la probabilidad del 0.81 que el modelo clasifique un ejemplo positivo aleatorio más alto que un ejemplo negativo aleatorio.

Conclusiones

Se realizó una base de datos, constituida por 49 instancias y cada instancia constó de diecisiete (17) variables concernientes a información socioeconómica, académica y personal de los estudiantes provenientes de los pueblos originarios Awajún y Wampis de la Universidad Nacional de Jaén. Empleando el algoritmo de selección de variables, InfoGainAttributeEval del software Weka, se redujo a cinco variables independientes: cursos aprobados, créditos aprobados, comunidad de origen, promedio acumulado y ciclo de ingreso. La variable dependiente fue la variable deserción de la UNJ.

El empleo de los algoritmos J48, Ridor y PART de clasificación han permitido obtener tres modelos basados en dos reglas por cada uno con un porcentaje de instancias bien clasificadas de 87.8, de esta manera siendo los que tienen mejor comportamiento. El modelo 1, es equivalente al modelo 3 que nos menciona que, si los cursos aprobados de los estudiantes son menores o iguales a diez, este alumno proveniente de pueblos originarios se retirará o ya ha abandonado la UNJ; sin embargo, si el estudiante ha aprobado más de diez cursos, entonces el alumno sigue siendo estudiante de la UNJ. Del modelo 2, se interpreta que, si un alumno va a ser estudiante excepto cuando sus créditos aprobados sean menores o iguales a veinte y siete, en este caso el alumno proveniente de pueblos originarios se retiraría o ya ha desertado de la vida universitaria.

Se puede usar las reglas generadas por los algoritmos de minería de datos en la implementación de un software que permita alertar sobre los estudiantes provenientes de los pueblos originarios Awajún y Wampis que potencialmente se encuentren en riesgo de desertar. De esta manera, la universidad pueda realizar acciones que puedan servir de ayuda a los estudiantes en riesgo.

Finalmente, se incluye algunas líneas futuro de trabajo derivado del desarrollo logrado: (a) Proponer métodos para ayudar a los estudiantes provenientes de los pueblos

originarios Awajún y Wampis detectados dentro del grupo de riesgo de desertar. (b) Posteriormente comprobar que porcentaje de las veces fue posible evitar que el estudiante detectado a tiempo desertara. (c) Desarrollar algoritmos basados en la combinación de técnicas de minería de datos para poder comparar con los resultados de algoritmos clásicos y obtener mejores resultados de predicción.

Bibliografía

1. «Inicio | BDPI». <https://bdpi.cultura.gob.pe/> (accedido oct. 02, 2019).
2. «Datamining (Minería de datos)». https://www.sinnexus.com/business_intelligence/datamining.aspx (accedido oct. 02, 2019).
3. P. F. Alania Ricaldi, «Aplicación de técnicas de minería de datos para predecir la deserción estudiantil de la facultad de ingeniería de la Universidad Nacional Daniel Alcides Carrión», 2019. Master's Thesis, Universidad Nacional Daniel Alcides Carrión. Escuela de Posgrado, Maestría en Ingeniería de Sistemas y Computación, 2019.
4. M. Kumar, A. J. Singh, y D. Handa, «Literature survey on educational dropout prediction», *Int. J. Educ. Manag. Eng.*, vol. 7, n.o 2, p. 8, 2017.
5. V. R. D. C. Martinho, C. Nunes, y C. R. Minussi, «An intelligent system for prediction of school dropout risk group in higher education classroom based on artificial neural networks», en 2013 IEEE 25th International Conference on Tools with Artificial Intelligence, 2013, pp. 159–166.
6. F. Dalipi, A. S. Imran, y Z. Kastrati, «MOOC dropout prediction using machine learning techniques: Review and research challenges», en 2018 IEEE Global Engineering Education Conference (EDUCON), 2018, pp. 1007–1014.
7. C. Burgos, M. L. Campanario, D. de la Peña, J. A. Lara, D. Lizcano, y M. A. Martínez, «Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout», *Comput. Electr. Eng.*, vol. 66, pp. 541–556, 2018.
8. A. B. B. Gutiérrez, R. C. Menéndez, L. J. Rodríguez-Muñiz, J. C. N. Pérez, E. T. Herrero, y M. E. García, «Predicción del abandono universitario: variables explicativas y medidas de prevención», *Rev. Fuentes*, n.o 16, pp. 63–84, 2015.
9. D. E. Gallardo Corrales, «Análisis de patrones de deserción estudiantil de la unidad educativa Lenin School aplicando minería de datos», Master's Thesis, Universidad de las Fuerzas Armadas ESPE. Maestría en Gestión de Sistemas de Información e Inteligencia de Negocios, 2017.
10. B. R. Cuji Chacha, «Las Técnicas de Predicción y su incidencia en la detección de patrones de Deserción Estudiantil en la Carrera de Docencia en Informática de la Facultad de Ciencias Humanas y de la Educación de

- la Universidad Técnica de Ambato», Master's Thesis, Universidad Técnica de Ambato. Facultad de Ingeniería en Sistemas, Electrónica e Industrial, 2016.
11. **L. E. Piscoya Ordoñez**, «*Aplicación de técnicas de minería de datos para predecir la deserción estudiantil en la educación básica regular en la Región de Lambayeque*», Tesis de Licenciatura, Universidad Señor de Sipán, Lambayeque, Perú 2017.
 12. **E. Pineda, E. Leal, y C. Barrera**, «*Los Sistemas Expertos como alternativa de Solución a la Deserción Universitaria*», Congreso Académico UDI, Octubre 2010.
 13. «*¿Qué es el muestreo por conveniencia?*», *QuestionPro*, abr. 28, 2018. <https://www.questionpro.com/blog/es/muestreo-por-conveniencia/> (accedido oct. 02, 2019).
 14. **S. Pal**, «*Mining educational data to reduce dropout rates of engineering students*», *Int. J. Inf. Eng. Electron. Bus.*, vol. 4, n.o 2, p. 1, 2012.
 15. **A. Torrado, Y. Karina, E. Barrientos Avendaño, y D. J. Heredia Vizcaíno**, «*Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos*», Cuarta Conferencia de Directores de Tecnología de Información, TICAL2014, pp 553-562, Mayo 2014.
 16. **J. Arroyo Hernández**, «*Métodos de reducción de dimensionalidad: Análisis comparativo de los métodos APC, ACPK y ACPK*», *UNICIENCIA*, vol. 30, n.o 1, pp. 115-122, 2016, doi: <http://dx.doi.org/10.15359/ru.30-1.7>.
 17. **K. Azoumana**, «*Análisis de la deserción estudiantil en la Universidad Simón Bolívar, facultad Ingeniería de Sistemas, con técnicas de minería de datos*», *Pensam. Am.*, vol. 6, n.o 10, 2013.
 18. **C. M. Vera, C. R. Morales, y S. V. Soto**, «*Predicción del fracaso escolar mediante técnicas de minería de datos*», *Rev. Iberoam. Tecnol. Delda AprendizajeAprendizagem*, vol. 109, 2012.