



[Cierre de edición 01 de setiembre del 2016]

doi: <http://dx.doi.org/10.15359/ree.20-3.11>

URL: <http://www.una.ac.cr/educare>

CORREO: educare@una.cr

Aportaciones desde la minería de datos al proceso de captación de matrícula en instituciones de educación superior particulares

Contributions to the Enrollment Process with Data Mining in Private Higher Education Institutions



*Rafael Isaac Estrada-Danell*¹

Instituto de Estudios Superiores de Tamaulipas

Tamaulipas, México

correo: rafael.estrada@iest.edu.mx

orcid: orcid.org/0000-0001-7534-8456

*Roman Alberto Zamarripa-Franco*²

Instituto de Estudios Superiores de Tamaulipas

Tamaulipas, México

correo: roman.zamarripa@iest.edu.mx

orcid: orcid.org/0000-0001-5791-221X

*Pilar Giselle Zúñiga-Garay*³

Fundación Teletón México

Tamaulipas, México

correo: garay@teleton.org.mx

orcid: orcid.org/0000-0002-5246-3168

*Isaías Martínez-Trejo*⁴

Universidad Autónoma de Tamaulipas

Tamaulipas, México

correo: ysayas55@hotmail.com

orcid: orcid.org/0000-0002-5462-0542

Recibido 23 de noviembre de 2015 • Corregido 1 de julio de 2016 • Aceptado 16 de agosto de 2016

¹ Licenciatura en Ingeniería Industrial para la Dirección, Maestro en Gestión del Capital Humano, ambos por el IEST – Anáhuac. Candidato a Doctor en Educación Internacional por la Universidad Autónoma de Tamaulipas. Actualmente es coordinador del Repensar de la Universidad en el Instituto de Estudios Superiores de Tamaulipas. Miembro de la Sociedad Mexicana de Educación Comparada. Árbitro de diversas revistas científicas. Sus temas de investigación son: minería de datos, ciencia de datos, ética y deontología de la educación superior.

² Ingeniero en Sistemas Computacionales por el Instituto Tecnológico de Cd. Madero, Maestro en Educación y Maestro en Calidad por el IEST – Anáhuac, Especialidad en Entornos Virtuales de Aprendizaje por Virtual Educa y la OEI, Doctor en Educación Internacional por la Universidad Autónoma de Tamaulipas. Profesor, investigador, coordinador de Tecnologías para la Educación en el IEST – Anáhuac. Coautor de diversos capítulos de libros, artículos científicos y ensayos.

³ Licenciada en Psicología por el IEST – Anáhuac, Maestra en Neuropsicología por el Instituto Europeo de Formación y Consultoría (INEFOC) de Madrid, España. Actualmente se desempeña como psicóloga familiar en el Centro de Rehabilitación Infantil Teletón (CRIT), Tamaulipas. Es además, Orientadora en CH & R Consultores. Sus temas de investigación son: educación, discapacidad, minería de datos y rehabilitación neuropsicológica.

⁴ Contador Público Auditor, Master en Comercio Exterior por la Universidad Autónoma de Tamaulipas, Dr. en Ciencias de la Educación y doctorante en Gestión Estratégica de Negocios. Maestro de tiempo completo en la UAT e integrante del Cuerpo Académico Consolidado, Investigaciones Jurídicas y Sociales, participante en proyectos de investigación.



doi: <http://dx.doi.org/10.15359/ree.20-3.11>

URL: <http://www.una.ac.cr/educare>

CORREO: educare@una.cr

Resumen. El presente artículo científico de investigación tiene por objetivo analizar la forma en que la minería de datos (MD) permite optimizar el proceso de captación de matrícula. Esto, con la intención de diseñar un modelo predictivo de gestión de matrícula para las IES particulares de México. Se analiza la situación actual de las instituciones de educación superior (IES) particulares en relación con su proceso de captación de matrícula y la aplicación de la MD en este. Con un método correlacional, se utilizó un conjunto de datos de prospectos ficticios para crear un árbol de decisión con enfoque de criterio de disminución de entropía con el software *Rapid Miner*. Los resultados muestran que es posible construir y probar un modelo predictivo de gestión de matrícula como el ZAM&EST propuesto por quienes escriben, para que las IES particulares puedan mejorar sus procesos de captación.

Palabras claves. Gestión educacional, planificación de la educación, administración de la educación, instituto de enseñanza superior, universidad, gestión de la información.

Abstract. This article aims to analyze how data mining (DM) optimizes the enrollment process, with the intention of designing a predictive model to manage private enrollment for higher education institutions of Mexico. It analyzes the current status of the higher education institutions in relation to its enrollment process and the application of the DM. With a correlational method, a dataset (DS) was used to model an entropy decision tree with the help of Rapid Miner software. The results show that it is possible to build and test a predictive model management of private enrollment for higher education institutions of Mexico as the ZAM&EST model proposed by the authors.

Keywords. Educational management, educational planning; educational administration, higher education institutions, universities, information management.

El acelerado crecimiento de la educación superior en el mundo es una realidad evidente. En algunos países el porcentaje de población que ingresa a la educación superior se elevó explosivamente en las últimas décadas (Álvarez, 2011); no obstante, parece que el ritmo se ha estabilizado. De acuerdo con datos de la Organización para la Cooperación Económica y el Desarrollo (OCDE, 2014), en 2005 el 24% de la población mundial⁵ estaba matriculada en educación terciaria⁶, cifra que para el 2012 ascendió al 32%, lo que representó un aumento de 8% en un periodo de siete años.

La explosiva expansión de la educación superior iniciada en los años 60 (Álvarez, 2011) presenta una serie de retos para los países del mundo. Uno de ellos consiste en incrementar la capacidad de los sistemas de educación en aras de responder a la creciente demanda.

⁵ La población se conforma por todos los sujetos que se encuentran en el rango de los cinco años posteriores a lo que corresponde como el fin de los estudios de educación secundaria.

⁶ La educación terciaria corresponde a la clasificación ISCED 5 y 6 de la UNESCO, por lo que equivale a lo que referimos como educación superior en México.

doi: <http://dx.doi.org/10.15359/ree.20-3.11>URL: <http://www.una.ac.cr/educare>CORREO: educare@una.cr

De acuerdo con la [Secretaría de Educación Pública, SEP y Sistema Interactivo de Consulta Estadística Educativa \(s. f.\)](#), según información obtenida del Sistema Nacional de Información Estadística Educativa (SINEP), en 2008 había en México⁷ un total de seis mil seiscientos veintiséis instituciones de educación superior (IES)⁸, que para 2014 aumentaron a 9 mil 939; lo que significa un 50% de incremento. Entre el 2008 y el 2012, tanto el sector público como el particular incrementaron, en 50%, el número de instituciones de educación superior.

A pesar de que hay 20% más IES particulares que públicas, la brecha no parece haberse abierto en los últimos siete años. La oferta del sector público en México ha aumentado a la par que la oferta del sector privado en el periodo de 2008-2014⁹. En el mismo periodo la matrícula total de estudiantes en educación superior se incrementó de 2 millones 705 mil 190 a 3 millones 419 mil 331, un aumento de 26%.

En 2008, del total de estudiantes matriculados en educación superior, el 33% era atendido por IES particulares; mientras que el sistema público captaba al 67%. En 2014, se registró un total de 1 millón 046 mil 754 estudiantes matriculados en IES particulares, mientras que en públicas se registró un total de 2 millones 272 mil 637, lo cual significa que, del total de estudiantes en nivel superior en 2014, el 31% fue atendido por IES particulares mientras que el 69% por el sistema público.

Lo anterior permite esbozar el escenario actual de alta competencia entre las IES particulares. En este contexto, la captación de matrícula se posiciona como un tema fundamental. A esto se suma el crecimiento en la oferta y captación del sistema público. La MD constituye, ante la situación descrita, una importante oportunidad para hacer más eficientes los procesos de captación.

La MD es un tema relativamente nuevo en México. Debido principalmente a la complejidad en el manejo de grandes bases de datos, *big data* y al uso de software especializado para su análisis. No obstante, en los últimos años, dicho software se ha aproximado a los entornos académicos –ejemplo de ello es la iniciativa académica de *Rapid Miner*–, que pone al alcance de IES el uso de técnicas avanzadas para generar modelos predictivos y descriptivos de diversa índole.

De ahí el interés de este trabajo por responder a la pregunta: ¿De qué manera la MD puede aportar a la optimización del proceso de captación de matrícula en las IES particulares? El objetivo es, por tanto, analizar de qué forma la MD representa una oportunidad de optimización del proceso de captación de matrícula en IES particulares, así como diseñar un modelo predictivo de gestión de matrícula, cuya aplicación aportaría a la subsistencia de instituciones, de manera que estas puedan seguir concentrando esfuerzos en la formación académica, tarea que constituye el auténtico corazón de la universidad.

⁷ Se agrupan los datos del rubro de escuelas e instituciones bajo el título de instituciones.

⁸ Los datos incluyen los siguientes niveles educativos: técnico superior, normal, licenciatura, licenciatura universitaria tecnológica y posgrado.

⁹ Para el cálculo del PROPHE se toma en cuenta exclusivamente el total de instituciones, sin considerar las escuelas. Los cálculos presentados a partir de los datos del SINEP consideran el total de instituciones más escuelas.



doi: <http://dx.doi.org/10.15359/ree.20-3.11>URL: <http://www.una.ac.cr/educare>CORREO: educare@una.cr

Panorama de la educación superior particular en México

En el 2012, 43% de la población de América Latina se encontraba matriculada en educación superior; México, en el mismo año, registraba 29% de participación en este rubro. Como se aprecia en la [figura 1](#), países como Chile, Argentina y Cuba¹⁰ constituían la cabeza del bloque latinoamericano en 2012, con porcentajes de participación de 74.4%, 78.6% y 62.5% respectivamente (OCDE, 2014).

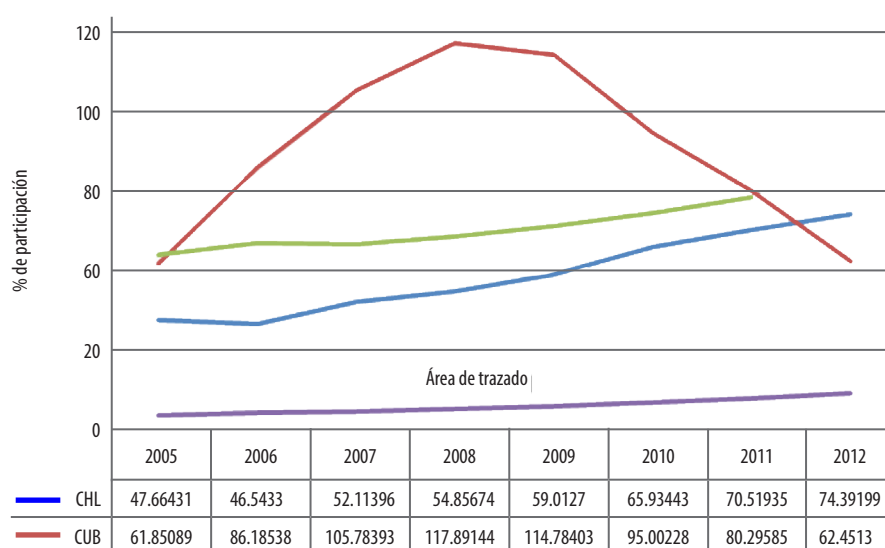


Figura 1. Educación terciaria en América Latina (2005-2012). Elaboración propia a partir de los datos de [The World Bank \(2000\)](#).

Ante el incremento de la demanda en la educación superior, los rumbos de acción de los gobiernos son variados, pero en general es posible distinguir dos aproximaciones principales: el fortalecimiento del sistema público de educación superior o bien el impulso del sector privado para que este atienda a las masas que el sistema público no está en condiciones de recibir. Si bien se habla de una aproximación mixta al problema porque combina los dos caminos mencionados, parece que en América Latina se ha optado, principalmente, por la vía de la apertura al sector privado.

¹⁰ Cuba ha experimentado una dramática caída a partir de 2008, donde contaba con prácticamente el total de su población en educación superior. En 2012 se registra una participación aún alta con respecto al continente de 65%.

En la [tabla 1](#), se observa que, en los principales países de América Latina, se viene presentando un crecimiento acelerado de la participación del sector privado en la oferta de servicios de educación superior. Las IES particulares han ganado cada vez más terreno y matrícula, la expansión es incluso abrumadora en algunos países como Chile, donde en el año 2005 el 92% de las IES eran particulares y de 753 mil 543 sujetos estudiantes activos en educación superior, 584 mil 722 se encontraban matriculados en el sector privado, es decir, 77% ([Bjarnason, Cheg, Fielden, Lemaitre, Levy y Varghese, 2009](#))¹¹ y en Brasil con un incremento del 74% en su matrícula y el 89% en las instituciones educativas.

Tabla 1

Crecimiento de la educación superior particular en América Latina 2002-2007

País	Matricula	Instituciones
Argentina	24%	53%
Brasil	74%	89%
Chile	77%	92%
México	33%	72%
Total	48%	71%

Nota: Elaboración propia a partir de los datos de [Bjarnason et al. \(2009\)](#).

En la misma [tabla 1](#), se observa que el crecimiento de las instituciones educativas no ha estado a la par del incremento de matrícula. Se muestra una brecha importante en los países de México y Argentina, en donde se tiene un alto crecimiento de instituciones (72% y 53% respectivamente) y un bajo crecimiento de matrícula (33% y 24%). A continuación, se profundiza en la explicación de esta problemática, comparándola con el crecimiento en el sector público.

En México, el número de IES particulares registró un explosivo crecimiento en los años 60. Producto de esta explosión, en 2014 el sector particular participa del 60% del total de las IES del país. No obstante, en últimos años se registra un crecimiento a la par de ambos sectores. Entre 2008 y 2014, el número total de IES aumentó en 50%, las IES públicas y particulares aumentaron en este mismo periodo en 50%, ambas en la misma proporción; por su parte el incremento de estudiantes en el periodo mencionado es de 26%, en este periodo; la captación de las IES particulares pasó de representar el 33% al 31%, mientras que el sector público pasó del 67% al 69%. En la [tabla 2](#) se puede consultar a detalle el crecimiento y la participación de los sectores públicos y particulares en lo que respecta a educación superior en México.

¹¹ PROPHE Program for Research on Private Higher Education. Este programa es dirigido por Daniel Levy de la Universidad Estatal de Nueva York. Sus datos son utilizados por la UNESCO (2009) en su documento "A new dynamic: private higher education".

doi: <http://dx.doi.org/10.15359/ree.20-3.11>URL: <http://www.una.ac.cr/educare>CORREO: educare@una.cr

Aunque las IES particulares son mayoría con un 60%, su número no crece más rápido que lo que lo hacen las IES públicas. Si a esto se le suma que las IES particulares están perdiendo participación en la matrícula total, el escenario actual queda más claramente dibujado. El mercado para la captación de matrícula, si bien aumenta, se reparte con preferencia hacia el sistema público. Es, entonces, necesario que las IES particulares realicen esfuerzos cada vez mayores para ganar estudiantes en el ya reducido mercado que al parecer se constriñe más cada año.

Tabla 2

Comportamiento de IES y matrícula en México 2008-2014

Categoría	2008-2009	2013-2014	Crecimiento	% Crecimiento	% Participación 2008	% Participación 2014
IES particulares	3,957	5,941	1,984	50%	60%	60%
IES públicas	2,668	3,998	1,330	50%	40%	40%
Total IES	6,625	9,939	3,314			
Matrícula particular	895,783	1,046,754	150,971	17%	33%	31%
Matrícula pública	1,809,407	2,372,637	563,230	31%	67%	69%
Total matrícula	2,705,190	3,419,391	714,201			

Nota: Elaboración propia a partir de datos de la [SEP](#) y [Sistema Interactivo de Consulta de Estadística Educativa \(s. f.\)](#).

El proceso de expansión acelerado del sector privado en el ramo educativo, de las últimas décadas, no ha estado exento de críticas. Numerosos estudios denuncian que el crecimiento desproporcional y desregulado de IES particulares ha puesto en riesgo la calidad educativa e incluso la esencia de la educación superior ([Altbach y Levy, 2005](#)). [Altbach \(2006\)](#) afirma que este escenario da lugar al surgimiento del nuevo modelo de la “pseudo universidad”, la cual se caracteriza por ser una entidad con fines de lucro centrada en la ganancia económica y donde no se realiza verdadera docencia, verdadera investigación ni verdadera extensión.

Lo cierto es que el enorme aumento de IES particulares en México en décadas pasadas fue tan volátil que es fácil suponer que muchas de las ofertas emergentes carezcan de la capacidad para brindar educación superior de calidad. Es en buena medida la preocupación por la calidad de la educación frente a la desmesura del crecimiento en la oferta la que se constituye como motor para el surgimiento de organismos acreditadores de calidad educativa en diversos países.

En México, el [Consejo para la Acreditación de la Educación Superior \(COPAES, 2015\)](#) es la instancia autorizada por el gobierno para dar reconocimiento formal a las diferentes organizaciones –comités técnicos– cuya finalidad es acreditar programas educativos de tipo



doi: <http://dx.doi.org/10.15359/ree.20-3.11>

URL: <http://www.una.ac.cr/educare>

CORREO: educare@una.cr

superior. Según la BD de la COPAES, existen actualmente 3 mil 307 programas acreditados, de los cuales solo 883 son programas de IES particulares, lo cual representa apenas el 27%. Estos programas pertenecen a 64 IES particulares del total de 5 941, es decir, solo el 1% cuenta con al menos un programa acreditado por COPAES.

La falta de presencia de IES particulares en los registros de COPAES puede ser considerada un indicio de la existencia de «pseudo universidades» que carecen de la capacidad para brindar educación de calidad. Entonces, ¿qué panorama enfrenta la educación superior particular en la actualidad? A continuación, se muestran dos elementos fundamentales para describir esta situación.

1. Mercado saturado. El entorno de la educación superior particular es un entorno con muchas universidades participantes. Cada año se registran en promedio 396 nuevas IES particulares que aspiran a captar un mercado que se reduce cada vez más. A esto hay que sumar el crecimiento de la oferta del sector público y el aumento que registran en su captación de matrícula en los últimos años.
2. Bajo nivel de calidad: Muchas de las nuevas IES no son capaces de ofrecer programas de calidad y ofertan sus servicios a costos menores que los entes competidores que invierten en la calidad educativa. Esto no solo impacta a las IES particulares comprometidas con la calidad educativa, las cuales deben enfrentar una competencia desleal, sino que también genera un problema social al país, ya que egresan generaciones con un bajo nivel profesional.

Es evidente que la mercadotecnia en la captación de matrícula juega un papel fundamental. Resulta indispensable generar estrategias que permitan a las IES particulares atraer prospectos del reducido y saturado mercado de la actualidad, por lo que las investigaciones que aporten a la mejora del proceso de captación de matrícula son fundamentales para la sobrevivencia de IES particulares.

El proceso de captación de matrícula de las IES particulares

El proceso de captación de matrícula es aquel que tiene por objetivo captar la mayor cantidad de alumnado para que se inscriba en la IES. Considera diversas actividades, tales como la invitación de estudiantes de bachillerato a las instalaciones, la visita a los bachilleratos, la participación en ferias de universidades, promoción, publicidad y difusión, entre otras. En cada interacción con alumnado de bachillerato, se capturan sus datos y se almacenan en la BD de prospectos de la IES particular.



doi: <http://dx.doi.org/10.15359/ree.20-3.11>

URL: <http://www.una.ac.cr/educare>

CORREO: educare@una.cr

Las personas responsables de la captación en las IES deben dar seguimiento a estos prospectos, para lograr que se inscriban. Este seguimiento se basa fundamentalmente en la consulta a la BD que se ha generado en todo el proceso. Existen retos que implica el captar matrícula en la actualidad, a continuación, se muestran los principales:

1. La recolección de los datos sobre los prospectos debe ser realizada en apego a la ley de protección de datos individuales, por lo que es preciso tratar con sumo cuidado los datos personales que se utilizan en la modelación.
2. La recolección de los datos sobre los prospectos debe ser realizada con eficacia, ya que se pueden registrar datos que no son válidos, lo cual genera inconsistencias en la BD.
3. Los datos de los prospectos son dinámicos, en cualquier momento pueden cambiar sus datos personales y los relacionados con sus intereses, por lo que se deben actualizar constantemente.
4. Las BD generalmente contienen mucha información de pocos prospectos a inscripción, lo cual no es significativo y requiere de un eficiente trabajo de la persona responsable que las utiliza.
5. Los reportes que se extraen de las BD generalmente muestran información que apoya principalmente la identificación y el seguimiento de contactos con el prospecto. No se realizan análisis profundos de los datos, para generar información útil para apoyar en la captación. Por ejemplo, las segmentaciones del mercado son generalmente realizadas de manera subjetiva y en función de un solo atributo (escuela de procedencia o nivel socioeconómico), cuando en realidad los sujetos que conforman las BD presentan más atributos que pueden impactar en los resultados. Tampoco es posible realizar predicciones sobre el comportamiento de los prospectos, lo cual ayudaría a enfocar los esfuerzos para obtener mejores resultados.

Minería de datos

Para afrontar los retos de la captación de matrícula, las IES particulares se pueden apoyar en el uso de las tecnologías de información, ya que el desarrollo tecnológico de los últimos años ha revolucionado la manera en que nos relacionamos con el mundo. La realidad virtual se traslapa con la realidad física y lo hace a un nivel tal que los sistemas tecnológicos se han vuelto componentes principales de sociedades y organizaciones contemporáneas.



doi: <http://dx.doi.org/10.15359/ree.20-3.11>

URL: <http://www.una.ac.cr/educare>

CORREO: educare@una.cr

El uso de la tecnología en los diferentes aspectos de la vida ha devenido en las denominadas sociedades de la información, donde cada minuto se generan cantidades descomunales de datos, los cuales se organizan en BD, generando grandes silos de información almacenada. A este gran volumen se le conoce como el *big data*.

La organización en la BD se realiza a través de agrupación de caracteres que, a su vez, forman campos, que al juntarse forman registros. Este conjunto de registros forma un archivo y cuando estos se relacionan con otros, se forman las BD. Por ejemplo, una BD de estudiantes puede tener dos archivos, uno de datos personales y otro de grupos en los que tiene clases. Este archivo tiene campos que pueden ser el número de identificación de cada estudiante, el nombre, su dirección y teléfono. Estos campos están formados por caracteres, que pudieran ser la J la U la A y la N, en el caso de que el campo nombre sea JUAN.

Manejar y aprovechar tanta información se vuelve complicado, las organizaciones corren el riesgo de ahogarse en el mar de datos que inunda las áreas de trabajo. De ahí que la extracción de conocimiento valioso a partir de grandes BD sea una de las habilidades claves de la actualidad (Davenport y Patil, 2012).

La MD es justamente ese conjunto de herramientas que facilitan la obtención de conocimiento valioso; se puede entender la MD como una especie de ariete que rompe la impenetrable fortaleza de los silos de la información. El objetivo es generar modelos que permitan comprender y predecir el comportamiento de los procesos a partir de los registros almacenados en las BD. La MD inicia con los datos, a los cuales se les aplican métodos computacionales especializados para descubrir estructuras y patrones inicialmente ocultos (Kotu y Deshpande, 2015).

Los métodos son conocidos como algoritmos. Uno de estos son los árboles de decisión que permiten crear modelos de predicción. Consisten en una técnica que clasifica casos basados en las variables conocidas como predictoras. Cada variable y su posible valor se muestran en un nodo del árbol (Van Der Aalst, 2012).

Los árboles de decisión son un tipo de aprendizaje supervisado en analítica predictiva. Un aprendizaje supervisado significa que uno de sus atributos es definido como variable de respuesta, por lo que el modelo generado ofrece un esquema predictivo de dicha variable de respuesta en términos de las variables de entrada. Los árboles de decisión se utilizan para hacer predicciones sobre un atributo en particular. Una característica importante de resaltar es que la variable de respuesta debe ser de tipo categórica para que el algoritmo pueda funcionar.



doi: <http://dx.doi.org/10.15359/ree.20-3.11>

URL: <http://www.una.ac.cr/educare>

CORREO: educare@una.cr

Como muchos modelos predictivos, el árbol de decisión requiere trabajar con un porcentaje de la BD para la construcción del modelo, este porcentaje se conoce como data de entrenamiento y el otro porcentaje para la evaluación del modelo, que se conoce como data de prueba. Al finalizar el algoritmo se genera un modelo tipo árbol invertido que permite predecir cuándo un sujeto en particular cumplirá con el atributo categórico indicado como variable de respuesta.

Por ejemplo, para crear un modelo predictivo de gestión de matrícula, se fija como variable de respuesta el atributo de inscripción de los *data sets* de los ciclos pasados y se genera, a partir de este un modelo que exprese bajo qué condiciones de atributos predictivos, un sujeto en particular realiza o no su inscripción. La variable de respuesta en este caso deberá ser declarada binominal y los demás atributos podrán ser categóricos o numéricos.

Minería de datos y captación de matrícula en las IES

Las técnicas de MD se han aplicado con éxito en el comercio electrónico, la salud, el transporte, la industria farmacéutica, así como en la educación. En la actualidad, las IES cuentan con sistemas informáticos que permiten el registro de considerables volúmenes de datos que surgen de la medición de sus procesos. En este sentido, la aplicación de las técnicas de la MD en la educación constituye un tema relativamente nuevo que se ha enfocado en diferentes problemáticas educativas y ha servido a todos los actores del proceso de enseñanz-aprendizaje, esto es, a estudiantes, docentes y al personal administrativo de la educación.

En la [tabla 3](#), de acuerdo con [Romero y Ventura \(2007\)](#), se observa cómo la MD ha ayudado a cada uno de los sujetos del proceso, con un gran apoyo al sector administrativo de la educación, es decir, a las personas que gestionan el proceso educativo, que definen estrategias y toman decisiones para marcar el rumbo de esta.

La MD ha ayudado a tomar decisiones a las personas relacionadas con la gestión eficiente de los recursos, efectividad de los cursos, retención escolar, eficiencia terminal, desempeño académico y métodos de evaluación; aspectos todos cruciales para el éxito académico de una institución educativa.

Hacia el año 2010, según [Baker y Yacef \(2009\)](#), se ha observado que la MD, aplicada a la educación, se ha expandido a los países que se encuentran en Norteamérica, Europa Occidental, así como a Australia y Nueva Zelanda, principalmente con el uso de modelos de predicción, lo cual ha logrado impactar en la educación.

Tabla 3

La MD en el proceso educativo hasta el año 2007

Sujeto beneficiado	Área investigada
Estudiantes	<ul style="list-style-type: none">• Aprendizaje personalizado
Docentes	<ul style="list-style-type: none">• Seguimiento de estudiantes• Contenido• Diseño instruccional• Actividades de aprendizaje
	<ul style="list-style-type: none">• Interacción
	<ul style="list-style-type: none">• Necesidades específicas de instrucción• Usos del portafolio del estudiante
Personal administrativo	<ul style="list-style-type: none">• Eficiencia de recursos tecnológicos, materiales y humanos• Efectividad de cursos de educación a distancia
	<ul style="list-style-type: none">• Métodos de evaluación
	<ul style="list-style-type: none">• Desempeño académico
	<ul style="list-style-type: none">• Retención• Eficiencia terminal

Nota: Elaboración propia con base en [Romero y Ventura \(2007\)](#).

Corresponde ahora plantear la pregunta: ¿Qué puede aportar la MD al proceso de captación de matrícula en las IES particulares? Las posibles aplicaciones del árbol de decisión a las BD en el proceso de captación son diversas, pero especialmente un árbol de decisión puede ofrecer un modelo para predecir la inscripción o la no inscripción de un sujeto prospecto en particular.

En otras palabras, el árbol de decisión permitiría predecir si un sujeto prospecto en particular se inscribirá en función de conocer su escuela de procedencia, edad, promedio de preparatoria, dirección de vivienda, carrera de preferencia y opciones de estudio. Esto representa una importante aportación a la optimización de la captación de matrícula, dado que permite centrar esfuerzos en los sujetos que el modelo predice tienen más probabilidades de realizar su inscripción, identificándolos desde el inicio del proceso y dándoles seguimiento a lo largo de este mismo. La MD ofrece una oportunidad clave para mejorar los procesos de captación de matrícula, así como para generar estrategias de mercadotecnia adecuadas para atender segmentos particulares.

doi: <http://dx.doi.org/10.15359/ree.20-3.11>

URL: <http://www.una.ac.cr/educare>

CORREO: educare@una.cr

Chang (2006) investigó sobre las ventajas de analizar la problemática de la matrícula a través de técnicas de MD en lugar de los tradicionales análisis estadísticos. Los métodos tradicionales para interpretar los datos de sujetos prospectos durante el proceso de captación de estudiantes a la universidad, consideran los aspectos económicos, sociológicos y psicológicos, los cuales ya no son suficientes para comprender el proceso. Los avances tecnológicos como el MD hacen posible la identificación de patrones ocultos que permiten predecir el comportamiento de los sujetos prospectos considerando su interacción social (Sigillo, 2013).

Nandeshwar y Chaudhari (2009) utilizaron la técnica de árboles de decisión y encontraron que el apoyo financiero es el factor más importante que atrae a estudiantes a inscribirse, en ese sentido, Antons y Maltz (2006) encontraron que estos apoyos se deben analizar muy bien, por las implicaciones fiscales que conllevan. También, González y DesJardins (2002) utilizaron la técnica de redes neuronales para predecir el comportamiento de los sujetos prospectos.

González (2009) analizó los factores que más inciden en la decisión para inscribirse en carreras de ingeniería y encontró que los principales son los siguientes: las buenas experiencias en matemáticas, la autovaloración del rendimiento académico como sobresaliente en esta área, el hecho de ser hombre que no se haya matriculado anteriormente en otra carrera y que el padre sea quien sugiera su elección. Kumar y Saurabh (2012) encontraron que estudiantes del área de sistemas computacionales con un buen nivel de matemáticas, son quienes se inscriben en un programa de maestría en computación.

Método

Con fines de presentar una posible aplicación y analizar los resultados, este trabajo implementa el árbol de decisión a una BD aleatoria diseñada por el grupo autor, que simula los registros de un sistema electrónico de captación de matrícula de una IES. Esta BD consiste en una matriz donde los atributos de los sujetos prospectos a realizar su matrícula son organizados a manera de columnas. Para cada sujeto se generó un ID aleatorio para fines de seguimiento.

Las variables elegidas para esta prueba son: Promedio de bachillerato, segmento de escuela de procedencia, porcentaje de beca asignado, primera opción de carrera indicada, primera opción de universidad indicada. El diseño del estudio se considera correlacional, ya que a través del algoritmo del árbol de decisión se establecen relaciones entre unas variables, para predecir el comportamiento de otra. A continuación, se define cada variable.

Estatus: Valor categórico nominal dicotómico, que representa si el sujeto realizó o no realizó su matrícula en la IES. Puede tomar valores de "inscrito" o "no inscrito". En el árbol de decisión esta se define como variable de respuesta.

Promedio de bachillerato: Valor numérico entre 6 y 10 que representa el promedio general de las calificaciones obtenidas por el sujeto prospecto durante todo el curso del bachillerato o equivalente.

Segmento de escuela de procedencia: Valor categórico nominal que representa el segmento al que pertenece la escuela de procedencia del sujeto prospecto. Para fines de este ejercicio se asume la existencia de dos segmentos denominados A y B, conformados por bachilleratos con características específicas.

Porcentaje de beca: Valor numérico que representa el porcentaje de beca asignado a un sujeto prospecto previo a su inscripción. Toma valores desde el 0% al 90%.

Primera opción de carrera: Valor categórico nominal que representa la carrera considerada como primera opción para realizar la matrícula por parte del sujeto prospecto. Para fines de este ejercicio, se limitan los valores que puede tomar a cuatro carreras de ingeniería: Ingeniería Industrial (II), Ingeniería Mecatrónica (IM), Ingeniería Petrolera (IP) e Ingeniería Química (IQ).

Primera opción de universidad: Valor categórico nominal que representa la universidad considerada como primera opción para realizar la matrícula por parte del prospecto. Para fines de este ejercicio se limitan los valores que pueden tomar a 4 opciones denominadas A, B, C, D.

El arreglo matricial de esta BD se compone, para casos de esta prueba, por 80 sujetos sobre los cuales se predicen los atributos presentados previamente. Para cada sujeto se genera un ID de valor número que toma valores desde el 1 al 80.

La técnica se aplicó utilizando el software *Rapid Miner Studio 6.2 Starter Edition*¹². Se utilizaron los siguientes criterios para el algoritmo del *árbol de decisión*:

- Criterio general de desempeño: ganancia de información (reducción de entropía)
- Máxima profundidad del árbol: 20
- Pre-Pruning: Aplicado
- Confianza: .25
- Ganancia mínima: 1
- Split: 60% para data de entrenamiento, 40% para data de prueba
- Tipo de muestreo: Estratificado

¹² Versión gratuita disponible en <http://www.rapidminer.com>

doi: <http://dx.doi.org/10.15359/ree.20-3.11>

URL: <http://www.una.ac.cr/educare>

CORREO: educare@una.cr

Este trabajo está limitado en función de que la BD utilizada ha sido generada en forma aleatoria y es ficticia, de ninguna manera representa los registros auténticos de una IES particular. El análisis se dirige a exponer las posibles aplicaciones y evidenciar las ventajas del uso de las técnicas. Por lo que los resultados no pueden ser considerados como una expresión auténtica del comportamiento de los sujetos, sino que se presentan con fines de demostración del diseño del modelo.

El número de sujetos de la BD utilizada es considerablemente menor al que se utilizaría para un análisis de los registros reales de prospectos. Esto influye en el resultado del modelo; no obstante, no afecta el diseño de este mismo.

Debido a que la finalidad de este trabajo es presentar la aplicación de la técnica para construir un modelo, no se exponen aquí los análisis descriptivos de la data que corresponderían –promedios, varianzas, identificación de *outliers*– sino que se implementa directamente el algoritmo. La metodología que se siguió para la construcción del árbol de decisión fue la siguiente:

1. Generación del conjunto de datos
2. División de la BD para los datos de generación de árbol y prueba.
3. Elección de la variable que funcionará como primer nodo del árbol en función del cálculo y comparación de la reducción del nivel de entropía que implicaría la división en términos de las diferentes variables
4. Construcción del árbol
5. Prueba del árbol con la BD

Resultados: Modelo predictivo de gestión de matrícula

En la [figura 2](#) se observa el modelo generado al introducir la BD y aplicar los operadores correspondientes del árbol de decisión. En este caso se seleccionó la variable de segmento de procedencia como el primer nodo del árbol.

En este árbol de decisión se observan colores en cada cuadro de predicción. Estos colores representan el número de sujetos clasificados en forma correcta en la data de prueba. Cuanto más azul la barra, más confiable la predicción. Para esta BD se observó un 56% de precisión en la predicción de “inscritos” y un 44% para la predicción de “no inscritos”. El bajo nivel de los porcentajes de precisión se puede explicar por la aleatoriedad de la BD generada, y por su tamaño considerablemente menor a las BD reales de registros.

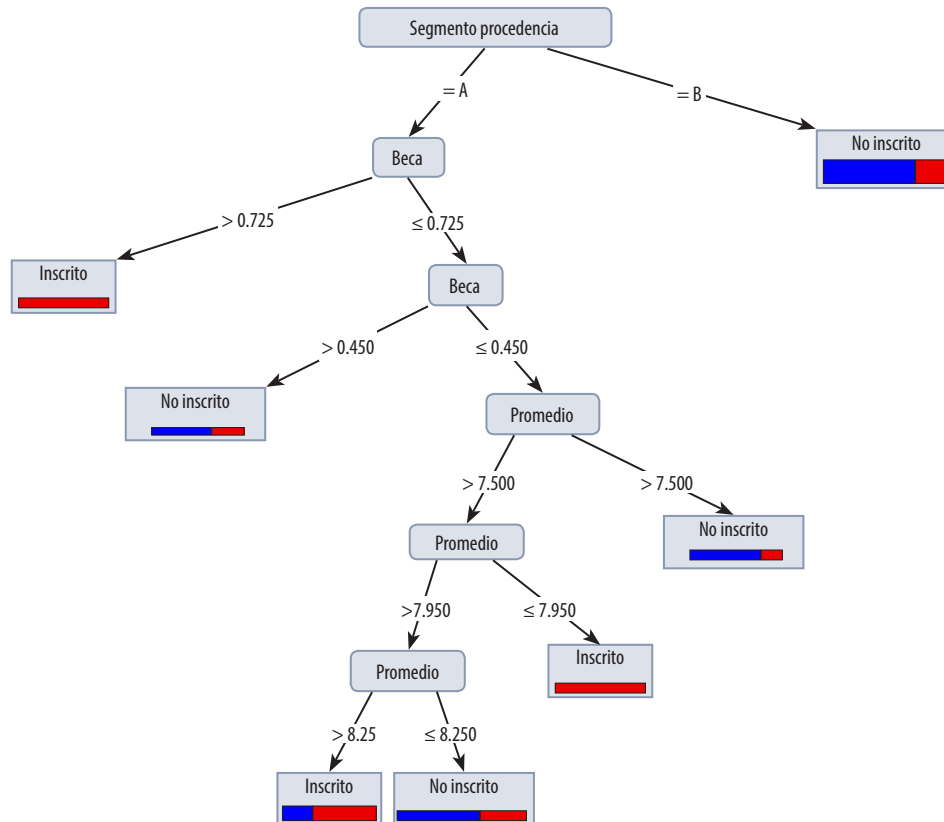


Figura 2. Modelo de árbol de decisión. Elaboración propia a través del software *Rapid Miner*.

El árbol fue generado con el 60% de la BD y, posteriormente, se prueba su capacidad predictiva en el restante 40%. Para aplicar este modelo a un conjunto de datos nuevos y predecir la inscripción o no inscripción de un sujeto, se deben extraer el conjunto de reglas de inducción que se encuentran presentes en el árbol de decisión.

La primera regla que el modelo nos permite observar consiste en dividir a los sujetos de la BD en términos del segmento de procedencia. La regla se lee de arriba hacia abajo.

1. Si el segmento de procedencia del prospecto es B entonces el valor que tomará es de no inscrito.

El resto de las reglas son las siguientes:

2. Si el segmento de procedencia es A, y tiene más de 72% de beca, el valor que tomará es de inscrito.

doi: <http://dx.doi.org/10.15359/ree.20-3.11>

URL: <http://www.una.ac.cr/educare>

CORREO: educare@una.cr

3. Si el segmento de procedencia es A, y tiene menos de 72% de beca, pero más de 45% de beca, el valor que tomará es de no inscrito.
4. Si el segmento de procedencia es A, tiene menos de 45% de beca y un promedio menor de 7.5, el valor que tomará es de no inscrito.
5. Si el segmento de procedencia es A, y tiene menos de 45% de beca y menor o igual 7.9 de promedio, el valor que tomará es de inscrito.
6. Si el segmento de procedencia es A, y tiene menos de 45 % de beca y tiene más de 8.5 de promedio, el valor que tomará es de inscrito.
7. Si el segmento de procedencia es A, y tiene menos de 45% de beca y su promedio es mayor que 7.9, pero menor que 8.2, tomará valor de no inscrito.

La regla generalizada para este modelo es la siguiente: Se inscribirán los prospectos que provengan del segmento A que tengan beca mayor o igual a 45% o con beca menor al 45% y promedio menor o igual a 7.9 o mayor de 8.5.

Estas reglas generadas a partir del árbol de decisión permiten realizar predicciones concretas para nuevas BD en términos del aprendizaje de la base previa, es decir, se actualizará en cada ciclo del proceso y cada vez será más exacta. En este sentido, es factible proponer el desarrollo de un sistema de información automatizado que permita desplegar, de manera visual, las predicciones sobre prospectos de las IES con la finalidad de tomar decisiones en cuanto a la implementación de estrategias específicas.

En la [figura 3](#) se muestra el modelo ZAM&EST de gestión de matrícula que resulta de la aplicación del árbol de decisión y del análisis del proceso de matrícula de una IES particular. Este modelo integra todos los elementos generados a lo largo del proceso, los cuales son los siguientes: captura de datos de sujetos prospectos, almacenamiento de los datos, filtrado y refinación de datos, construcción del árbol de decisión, generación de las reglas del modelo, el sistema de información predictivo, la estrategia a la medida y estudiante que se han inscrito. A continuación, se analiza cada uno de ellos.

Captura de datos de prospectos. Este elemento es fundamental para el resultado exitoso del modelo, ya que la calidad de los datos debe estar garantizada. Es decir, la información que se genere debe ser válida y confiable, no se deben almacenar datos erróneos o incongruentes. Por eso adquiere mucha importancia la tecnología que apoya la captura de los datos, para lo cual se deben utilizar dispositivos móviles, tales como: laptops, tabletas, asistentes personales y teléfonos inteligentes. Otro aspecto muy importante es la ampliación de las fuentes de datos. No solo se deben capturar los datos personales y de intereses de los sujetos prospectos, sino

que también se deben capturar datos de entrevistas, correos electrónicos y comentarios de sujetos prospectos en redes sociales.

Almacenamiento de los datos. Los datos se deben almacenar y organizar en una BD. Es necesario un software manejador de BD que permita la aplicación de operaciones sobre estos datos, tales como altas, bajas, cambios y consultas.

Filtrado y refinación de datos. Es necesario ejecutar un proceso que limpie la BD para eliminar inconsistencias y ajustar la BD al modelo propio de la IES.

Construcción del árbol de decisión. Se debe construir el árbol de decisión basado en la BD refinada. Una vez generado el árbol, este se debe probar con la BD inicial para medir su efectividad. Se utiliza el software *Rapid Miner* para realizar este proceso en forma automática. Aunque esto genera la utilización de dos tipos de datos diferentes e incompatibles (la BD y árbol de decisión), se considera que es la mejor opción por los costos. Esta situación se podría subsanar utilizando un único software integrado que permita cubrir todas las fases del modelo, pero son costosos.

Generación de reglas del modelo. Se interpreta el árbol de decisión para generar las reglas del modelo. Se debe enunciar una regla generalizada que resuma en una expresión qué sujetos prospectos tienen mayor probabilidad de que se inscriban. Estas reglas servirán, además, para retroalimentar la BD y mejorar su nivel de predicción.

Sistema de información predictivo. Se debe contar con un sistema de información que esté basado en las reglas del modelo y en la BD, de tal manera que permita generar la lista de sujetos prospectos que se predice se inscribirán en la IES particular. Este sistema debe volver a estandarizar los datos que se disgregaron en el árbol de decisión.

Estrategia a la medida. Con el conocimiento sobre los sujetos prospectos que provee el sistema de información, se deben diseñar estrategias específicas enfocadas a ellos para lograr su inscripción. Esta estrategia debe derivar en objetivos, metas y planes de trabajo, ya que de lo contrario no se lograrán resultados exitosos, porque el conocimiento sobre los sujetos prospectos no es suficiente para su logro.

Inscritos. Las alumnas y alumnos inscritos ahora forman parte de los datos de entrada para la BD.

El flujo y la interacción de este modelo son secuenciales, cíclicos, pueden ser en ambos sentidos y consideran la retroalimentación a la BD en dos momentos distintos. El primero se debe realizar una vez generadas las reglas y el segundo, al finalizar un ciclo. Esto incrementará el tamaño de la BD y la exactitud de las predicciones.

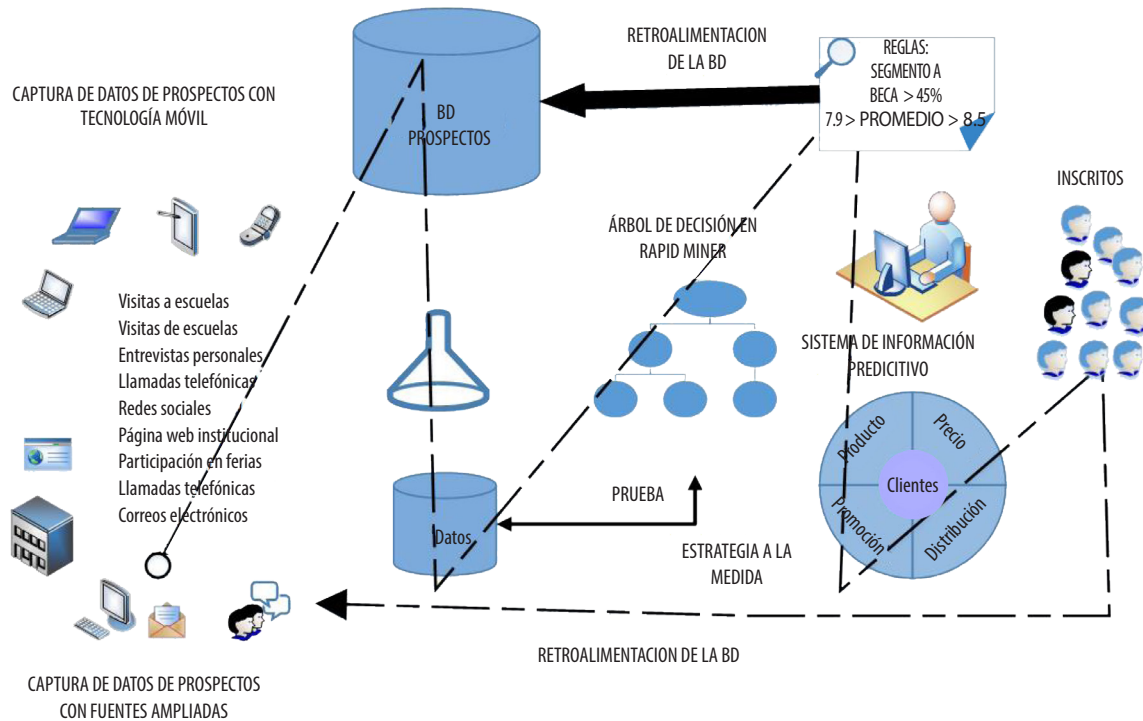
doi: <http://dx.doi.org/10.15359/ree.20-3.11>URL: <http://www.una.ac.cr/educare>CORREO: educare@una.cr

Figura 3. Modelo ZAM&EST de Gestión de Matrícula de una IES particular. Elaboración propia.

Conclusiones

Con respecto al análisis teórico, se concluye que el desarrollo de nuevas tecnologías de información para mejorar el proceso de captación de matrícula es indispensable, esto debido a la reducción del mercado que obliga a las IES a dirigir su mirada hacia técnicas de minería de datos para apoyar este proceso.

En relación con el objetivo de investigación propuesto, se concluye que la técnica de árbol de decisión es adecuada para generar un modelo de predicción de gestión de matrícula, que apoye en el seguimiento de prospectos con respecto a su inscripción en una IES particular, esto, combinado con una estrategia adecuada y enfocada, constituye una herramienta importante para centrar esfuerzos y recursos en prospectos con mayor probabilidad de realizar su inscripción. A su vez, el modelo puede ser actualizado con cada ciclo escolar, aumentando así la BD y permitiendo mejorar su capacidad predictiva con el paso del tiempo.

El modelo de gestión de matrícula ZAM&EST es susceptible de utilizarse en las IES particulares con el objetivo de hacer más eficiente su proceso de captación.

doi: <http://dx.doi.org/10.15359/ree.20-3.11>

URL: <http://www.una.ac.cr/educare>

CORREO: educare@una.cr

Se recomienda a las IES particulares construir su modelo de gestión de captación de matrícula, comenzando con la aplicación del Árbol de Decisión en poblaciones relativamente pequeñas por ejemplo una división, facultad o carrera, evaluar el desempeño del modelo y repetir. Existen otras técnicas de MD que pueden ser aplicadas, se sugiere comenzar la implementación de clústers, mapas auto organizados, redes neuronales, reglas de asociación y regresiones lineales múltiples.

Es muy importante mencionar la necesidad de una adecuada aplicación del modelo propuesto, ya que la MD es una herramienta fundamental que debe usarse en los ambientes universitarios con la finalidad de hacer más eficiente sus procesos, aunque es indispensable recordar que no sustituye en ningún momento la experiencia y el contacto humano que es el centro de los procesos de promoción y captación, ya que tan sólo brindan una base más objetiva para la toma de decisiones.

Referencias

- Altbach, P. G. (2006). *International higher education: Reflections on policy and practice [Educación superior internacional: Reflexiones sobre política y práctica]*. Boston: College Center for International Higher Education. Recuperado de https://www.bc.edu/content/dam/files/research_sites/cihe/pubs/Altbach_2006_Intl_HigherEd.pdf
- Altbach, P. y Levy, D. C. (2005). *Private higher education. A global revolution [Educación superior privada. Una revolución global]*. Rotterdam: Sense Publishers. Recuperado de <https://www.sensepublishers.com/media/787-private-higher-educationa.pdf>
- Álvarez, G. (Abril, 2011). El fin de la bonanza. La educación superior privada en México en la primera década del siglo XXI. *Reencuentro: Educación superior privada*, 60, 10-29. Recuperado de http://148.206.107.15/biblioteca_digital/estadistica.php?id_host=6&tipo=ARTICULO&id=7755&archivo=3-545-7755wke.pdf&titulo=El%20fin%20de%20la%20bonanza:%20La%20educaci%C3%B3n%20superior%20privada%20en%20M%C3%A9xico%20en%20la%20primera%20d%C3%A9cada%20del%20siglo%20XXI
- Antons, C. M., y Maltz, E. N. (2006, otoño). Expanding the role of institutional research at small private universities: A case study in enrollment management using data mining [La expansión de la función de la investigación institucional hacia universidades privadas pequeñas: Un caso de estudio de gestión de matrícula utilizando minería de datos]. *New Directions for Institutional Research*, 131, 69-81. doi: <http://dx.doi.org/10.1002/ir.188>
- Baker, R. y Yacef, K. (2009, otoño). The state of educational data mining in 2009: A review and future visions [El estado de la minería de datos educacional en 2009: Revisión y visiones futuras]. *Journal of Educational Data Mining*, 1(1). Recuperado de <http://www.educationaldatamining.org/JEDM/index.php/JEDM/article/view/8/2>

doi: <http://dx.doi.org/10.15359/ree.20-3.11>

URL: <http://www.una.ac.cr/educare>

CORREO: educare@una.cr

- Bjarnason, S., Cheg, K.-M., Fielden, J., Lemaitre, M.-J., Levy, D. y Varghese, N. V. (2009). *A new dynamic: Private higher education [Una nueva dinámica: Educación superior privada]*. París: Unesco. Recuperado de <http://unesdoc.unesco.org/images/0018/001831/183174e.pdf>
- Chang, L. (2006, otoño). Applying data mining to predict college admissions yield: A case study [La aplicación de la minería de datos para predecir la eficiencia de la admisión a la universidad]. *New Directions for Institutional Research*, 131, 53-68. doi <http://dx.doi.org/10.1002/ir.187>
- Consejo para la Acreditación de la Educación Superior (COPAES). (2015). *Sistema de información de programas acreditados*. Recuperado de <http://sieduca.com/copaes/>
- Davenport, T. H. y Patil, D. J. (Octubre, 2012). Data scientist: Tshe sexiest job of the 21st Century [Científico de datos: El trabajo más sexy del siglo XXI]. *Harvard Bussines Review*. Recuperado de <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>
- González, D. E. (Julio-diciembre, 2009). Factores individuales que afectan la demanda de educación superior de ingenierías: Caso de la Pontificia Universidad Javeriana de Cali. *Cuadernos de Administración*, 22(39), 307-333. Recuperado de <http://www.redalyc.org/pdf/205/20511993014.pdf>
- González, J. M. y DesJardins, S. L. (2002). Artificial neural networks: A new approach to predicting application behavior [Redes neuronales artificiales: Un nuevo enfoque para predecir el comportamiento de solicitudes]. *Research in Higher Education*, 43(2), 235-258. doi: <http://dx.doi.org/10.1023/A:1014423925000>
- Kotu, V. y Deshpande, B. (2015). *Predictive analytics and data mining. Concepts and practice with rapidMiner [Análisis predictivo y minería de datos. Conceptos y práctica con rapidMiner]*. Waltham: Morgan Kaufmann.
- Kumar, S. y Saurabh, P. (2012). Data mining application in enrollment management: A case study [Aplicación de la minería de datos en la gestión de la matrícula: Un estudio de caso]. *International Journal of Computer Applications*, 41(5), 1-6. doi: <http://dx.doi.org/10.5120/5534-7581>
- Nandeshwar, A. y Chaudhari, S. (2009). *Enrollment prediction models using data mining [Modelos de predicción de matrícula utilizando minería de datos]*. Recuperado de http://nandeshwar.info/wp-content/uploads/2008/11/DMWVU_Project.pdf
- Organización para la Cooperación Económica y el Desarrollo (OCDE). (2014). *Education at a Glance 2014 OECD Indicators [Panorama de la educación. Indicadores de la OCDE 2014]*. Recuperado de <http://www.oecd.org/edu/Education-at-a-Glance-2014.pdf>



doi: <http://dx.doi.org/10.15359/ree.20-3.11>

URL: <http://www.una.ac.cr/educare>

CORREO: educare@una.cr

- Romero, C., y Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005 [Minería de datos educativa: Un estudio de 1995 a 2005]. *Expert Systems with Applications* 33(1), 135-146. doi: <http://dx.doi.org/10.1016/j.eswa.2006.04.005>
- Secretaría de Educación Pública (SEP) y Sistema Interactivo de Consulta de Estadística Educativa. (s. f.). *Total de instituciones en periodo 2008-2009*. México: Autor. Recuperado de <http://www.planeacion.sep.gob.mx/principalescifras/>
- Sigillo, A. (2013). *Predictive modeling in enrollment management: New insights and techniques [El modelo predictivo en gestión de matrícula: Nuevas perspectivas y técnicas]*. Recuperado de [http://www.uversity.com/downloads/research/El Whitepaper R6.pdf](http://www.uversity.com/downloads/research/El%20Whitepaper%20R6.pdf)
- The World Bank (2000). *Higher education in developing countries. Peril and Promise*. Recuperado de http://siteresources.worldbank.org/EDUCATION/Resources/278200-1099079877269/547664-1099079956815/peril_promise_en.pdf
- Van Der Aalst, W. (2012). Process mining: Overview and opportunities [Minería de procesos: Descripción y oportunidades]. *ACM Transactions on Management Information Systems*, 99(99). doi: <http://dx.doi.org/10.1145/2229156.2229157>



Cómo citar este artículo en APA:

Estrada-Danell, R. I., Zamarripa-Franco, R. A., Zúñiga-Garay, P. G. y Martínez-Trejo, I. (Setiembre-diciembre, 2016). Aportaciones desde la minería de datos al proceso de captación de matrícula en instituciones de educación superior particulares. *Revista Electrónica Educare*, 20(3), 1-21. doi: <http://dx.doi.org/10.15359/ree.20-3.11>

Nota: Para citar este artículo en otros sistemas puede consultar el hipervínculo "Como citar el artículo" en la barra derecha de nuestro sitio web: <http://www.revistas.una.ac.cr/index.php/EDUCARE/index>

