

ESTRATEGIAS Y PROCESOS RELACIONADOS CON LA ADQUISICIÓN Y RETENCIÓN DE ESTUDIANTES EN UNA UNIVERSIDAD PRIVADA UTILIZANDO HERRAMIENTAS Y TECNOLOGÍAS DE INTELIGENCIA DE NEGOCIOS

“La única condición para hacer minería de datos es tener un problema relacionado con el negocio y los datos pertinentes”

RAFAEL CASTILLO SANTOS¹

Resumen

La educación universitaria en nuestro país, dada la necesidad que tienen las universidades de carácter privado de competir por la vinculación de nuevos estudiantes, retener a los que ya están y evitar las deserciones, convierten cada vez más este servicio, en un producto comercial por el cual competir, y por lo tanto permanecer o no en el mercado.

Por la razón anterior, el objetivo principal del presente artículo, es mostrar las bondades de la aplicación de la minería de datos en el proceso de adquisición y retención de estudiantes en una universidad privada, relacionando las técnicas de modelamiento predictivo y modelamiento descriptivo, análisis de retención y sus algoritmos respectivos, cuando se formulan preguntas como ¿Cómo lograr un mayor número de estudiantes matriculados? ¿Cuántos estudiantes que ingresan continuaran después del Primer semestre? ¿Cuántos continuaran después del segundo semestre? ¿Cuál es el perfil de estudiante que probablemente continuara? ¿Cuál el perfil del estudiante que más probablemente desertara?

Se hace énfasis en la deserción universitaria debido a que es un problema crítico que se presenta actualmente en la mayoría de las universidades no solamente de Colombia, sino de la mayoría de países latinoamericanos.

Se describen las etapas de la metodología propuesta por Michael Berry y Gordon Linoff y las cuatro fases en las cuales se agrupan: filtrado de datos; Selección de variables; extracción del conocimiento e interpretación y evaluación. Además se proporcionan algunas guías que permitirán a una entidad universitaria privada tomar las acciones necesarias tendientes a mejorar la adquisición de sus estudiantes y retener a los que ya están matriculados.

Palabras clave: Retención y Deserción Universitaria; Minería de datos; modelamiento predictivo; modelamiento descriptivo; análisis de retención.

Abstract

The university education in our country, given the necessity that private universities have of competing for the linking of new students, to retain those that are already and to avoid the desertions, it convert more and more this service, in a commercial product for the one which to compete, and therefore to remain or not in the market.

For the previous reason, is the main objective of the present article, to show the kindnesses of the application of data mining in the process of acquisition and students' retention in a private university, relating the techniques of predictive and descriptive modeling, retention analysis and its respective algorithms, when ones formulate questions like How to achieve a bigger number of registered students? How many did students that enter continue after the First semester? How many did

¹ Docente T.C. Universidad Autónoma de Colombia. rcastill@correo.fuac.edu.co

they continue after the second semester? Which is student's profile that probably continued? Which is the student's profile that more probably deserted?

Emphasis is made in the university desertion because it is a critical problem that shows up at the moment not only in most of the universities of Colombia, but of most of Latin American countries. The stages of the methodology described are proposed by Michael Berry and Gordon Linoff and the four phases in which its groups: filtrate of data; Selection of variables; extraction of the knowledge and interpretation and evaluation.

Some guides are also provided that will allow to a private university to take the necessary spread actions to improve the acquisition of their students and to retain those that are already

Words key: Retention and University Desertion; Data Mining; Predictive and descriptive modeling; Retention analysis.

INTRODUCCIÓN Y ANTECEDENTES

Durante el tiempo transcurrido desde la fundación de la universidad, la toma de decisiones la han realizado los directivos fundamentalmente con base en la experiencia, después de realizar largos y tediosos análisis del comportamiento de la información, obtenida ésta, con herramientas que aunque facilitan el proceso, lo hacen de manera muy lenta y luego de interminables cruces.

Actualmente el proceso de acreditación por la calidad de la educación superior, condiciona a las universidades y establece algunos lineamientos orientados a mejorar la calidad. Algunos de ellos están relacionados con calidad académica, bienestar universitario y contratación de docentes. Para el manejo de estas variables, es importante contar con un sistema de información que permita determinar comportamientos, hallar proyecciones y relacionar diferentes variables. Por lo tanto la minería de datos sobre información histórica se constituye en una herramienta útil al permitir facilidades de generación de reportes y descubrir y entender patrones ocultos en bodegas de datos. Estos reportes y patrones permiten entender comportamientos y tendencias con muy alta precisión por medio del uso de algoritmos supervisados o no supervisados. ¿El Resultado? Las instituciones universitarias podrán de una manera más adecuada realizar todos sus procesos de proyección, planeación y gestión de los recursos.

En enero de 2004 SPSS, compañía dedicada a elaborar software para análisis estadístico y CLEMENTINE, software para minería de datos, publicó a través de JING LUAN, PHD, Jefe de planeación y registro, el artículo "Data mining application in Higher education" [1] en el cual se plantean los beneficios del uso de la minería de datos, en entidades universitarias. En él se hace una comparación de los procesos llevados a cabo en entidades de carácter comercial que también utilizan herramientas de minería de datos. Luego de presentar tres casos de estudio, el autor concluye "estas herramientas le permiten a una universidad un mejor uso y planeación de los recursos y la predicción de algunos comportamientos tales como procesos de aprendizaje, transferencia entre carreras y relaciones de mercadeo con los estudiantes" [1].

En el IV Coloquio Internacional sobre Gestión Universitaria en América del Sur ALIANZAS ESTRATÉGICAS, INTEGRACIÓN Y GESTIÓN UNIVERSITARIA 8-10 de diciembre de 2004, Florianópolis, Santa Catarina – Brasil, el Ing. Diego D. Gregoraz y el Dr. Julio C. Durand, presentaron la ponencia "Explotación de la información académica para la mejora continua", donde entre otros aspectos orientados a mejorar los procesos administrativos de la universidad, incluyeron la minería de datos como herramienta "que está en la intersección de varias disciplinas como estadística, aprendizaje automático (machine learning) y manejo de bases de datos, entre otras". A través de un ejemplo mostraron cómo se pudo establecer la correlación de diferentes variables (por ejemplo ubicación geográfica y resultados en diferentes asignaturas) para encontrar comportamientos ocultos de la deserción de los estudiantes.

En la Argentina, Universidad Nacional de la Patagonia, UNPA, un grupo de docentes realizó un estudio debido a la preocupación existente ante la deserción de estudiantes. "Inicialmente se realizó un estudio cuantitativo que en una primera etapa estableció el perfil del alumno que ingresaba y si esto tenía relación con el abandono. La segunda etapa fue de comprensión de las razones y en esta instancia, se fue directamente a la búsqueda de los alumnos desertores para que explicaran los motivos. Otra instancia se fundó en la percepción de esos motivos, donde descubrieron por ejemplo, entre otros, el aspecto metodológico de estudio, la relación con los profesores, las exigencias y falencias de aprendizaje" [2].

En la universidad de Río, provincia de Córdoba, Argentina, también se ha llevado a cabo un proceso similar, teniendo en cuenta que el proceso de deserción universitaria comprende tres términos "proceso de selección, medida del rendimiento académico, y eficacia del sistema educativo" [3]. El primero se enmarca en el enfoque sociológico, según el cual la selección que se opera en la Enseñanza Superior, se constituye en un filtro social que frena la movilidad. El segundo, en la Universidad se debiera abordar tres dimensiones: éxito, retraso y abandono. Y el tercero, la deserción sólo da cuenta que en

algunos estudios se registra una mayor tendencia al abandono en las instituciones que no tienen examen de ingreso [3]. En Uruguay [4], y en Europa [5] las preocupaciones son similares y los estudios realizados pretenden explicar y encontrar soluciones al tema.

SITUACIÓN DE LAS UNIVERSIDADES PRIVADAS FRENTE AL INGRESO DE NUEVOS ESTUDIANTES Y A LA DESERCIÓN ESTUDIANTIL

La situación presentada actualmente en las universidades a nivel latinoamericano incluyendo a México es preocupante. Debido por un lado al no ingreso o niveles bajos de ingreso de nuevos estudiantes y por otro lado a los altos índices de deserción universitaria. Según la revista Estadísticas e indicadores de la universidad Nacional de Colombia, basada en La información estadística del Departamento Nacional de Planeación, Boletín No 27, Educación y fuerza de trabajo, Bogotá, septiembre de

2000, y el Informe de Desarrollo Humano para Colombia, 1999 [6], muestra que tal tendencia para Colombia se ve reflejada en indicadores tales como el estrato social, tipo de educación impartida, ubicación geográfica, actividad económica y en menor medida el sexo.

Para el caso de los estratos 1, 2 y 3 se ve una tendencia significativa a la deserción debido fundamentalmente a que la edad de estos estudiantes oscila entre 17 y 20 años, en la cual ya están aptos para el ingreso a la población económicamente productiva; a la disminución de los ingresos familiares y por otra parte a la mayor valoración del trabajo material antes que el intelectual.

Quizá el mayor impacto de estas estadísticas, tal como lo muestra la figura 1, esta en el ingreso familiar, el cual para ingresos de estratos 5, cerca del 99% de los estudiantes ingresan y permanecen en la educación superior, mientras que para estrato 4 la cifra es de apenas el 37% y 18% para estrato 3.

Quintil de Ingreso per cápita	Cobertura de educación Superior		Diferencia TB-TN	Extraedad	
	Tasa bruta de cobertura	Tasa neta de cobertura		< 18 años	> 24 años
1	3.6	2.3	14	0.3	1.1
2	9.7	6.0	3.6	0.7	3.0
3	18.0	10.1	7.9	1.2	6.7
4	37.8	18.6	19.2	1.9	17.3
5	99.0	53.1	45.9	2.9	43.0
Total	30.0	16.1	13.9	1.3	12.6

Fuente: Cálculos de la Misión Social del DNP con base en Encuesta de Calidad de Vida 1997.

Figura 1. Cuadro comparativo de ingreso a la educación superior con relación a los ingresos familiares

[6].

Las cifras presentadas también reflejan de alguna manera la falta de apoyo del estado a la educación superior pública, la cual ha mostrado una tendencia a la baja.

Sin embargo un factor positivo en esta dinámica lo muestran las estadísticas de empleo. Tal como lo muestra el cuadro de la figura 2, Los salarios de los ocupados con educación superior crecieron entre 1991 y 1998 un 26% en las zonas urbanas, mientras que

para las personas con sólo estudios básicos de primaria decrecieron en un 18%. El aumento de salarios ha favorecido a los que tienen educación superior y la ocupación para personas con secundaria y universitaria aumentó en un 18% y 41% respectivamente, mientras que la tasa de empleo de las personas sin educación o con educación primaria bajó. Una condición que afecta el ingreso a los niveles salariales altos o al ingreso a determinados cargos es la exigencia de un mayor nivel educativo (figura 3.)

Zona	1991		1994		1998	
	Ocupados	Salario medio	Ocupados	Salario medio	Ocupados	Salario medio
Urbano						
Ninguna		0.84	195.468	1.13	267.692	0.90
Primaria	2.539.926	1.53	2.568.186	1.59	2.436.034	1.28
Secundaria	3.743.908	1.76	4.323.085	2.29	4.497.936	1.78
Superior	1.295.356	3.83	1.614.467	5.77	1.961.838	4.97
Subtotal*	7.869.025	2.00	8.723.186	2.70	9.201.094	2.30
Rural						
Ninguna		0.89	816.626	0.88	1.107.505	0.92
Primaria	3.688.823	1.21	3.554.629	1.03	3.324.455	0.93
Secundaria	1.183.237	2.08	1.282.235	1.65	1.223.566	1.60
Superior		4.26	149.902	3.64	213.110	3.84
Subtotal*	5.887.266	1.39	5.803.392	1.21	5.878.769	1.18
Total						
Ninguna	1.160.038	0.88	1.012.094	0.93	1.375.195	0.91
Primaria	6.228.749	1.34	6.122.814	1.27	5.760.489	1.08
Secundaria	4.927.145	1.84	5.605.320	2.14	5.721.502	1.74
Superior	1.402.644	3.87	1.764.369	5.59	2.174.948	4.86
Subtotal*	13.756.291	1.74	14.526.578	2.11	15.079.863	1.86

* El número restante para alcanzar el total corresponde a las personas que no informan
Fuente: DANE. Encuesta de hogares para septiembre de cada año

Figura 2. Empleo y salarios según nivel educativo. [6]

Nivel educativo/ocupación	1978 %	1991 %	1993 %	1995 %	1997 %	1999 %
Alguna universitaria y más						
Profesionales, técnicos y directivos	27.9	70.9	72.1	75.5	77.6	79.4
Personal administrativo	2.3	26.1	26.2	28.7	30.4	30.2
Comerciantes, vendedores	0.6	8.0	9.7	9.9	10.9+	10.1
Trabajadores de los servicios	0.3	2.1	2.4	2.8	4.0	3.9
Trabajadores agrícolas	0.8	0.6	0.5	0.7	0.9	0.9
Trabajadores operarios no agrícolas	0.4	3.1	3.6	3.7	3.8	3.9

Fuente: Cálculos Departamento Nacional de Planeación-UDS-DIOGS con base en Dane. Encuesta de Hogares, Septiembre.

Figura 3. Población ocupada, para educación universitaria o superior. [6]

Desde otro contexto el problema puede ser visto teniendo en cuenta las consideraciones presentadas por el diario la Prensa en su edición de Junio 20 de 2005 [7]. En este artículo se comenta que el porcentaje de estudiantes que no terminan su carrera universitaria esta alrededor del 52%, es decir de cada dos estudiantes que ingresan a la educación superior solamente uno termina. Entre las razones que se manifiestan para la deserción están:

- Razones económicas.
- Mala calidad de los programas. “51 programas están a punto de ser cerrados por el incumplimiento de requisitos de calidad” [7].
- Dificultades en el aprendizaje y en la lectura.
- Falta de información sobre la vida universitaria.
- Ausencia de una orientación vocacional.

CONTEXTO DE LA DESERCIÓN UNIVERSITARIA

Algunas universidades, como la Universidad Tecnológica de Pereira, han tomado algunas medidas tendientes a contrarrestar el problema. En este caso concreto se modifico el reglamento estudiantil, el cual esta vigente desde septiembre 15 de 2005, para permitir transferencias de estudiantes de no solamente carreras tecnológicas, sino de cualquier estudiante de cualquier carrera afín a las Ingenierías. También en esta reforma se hacen cambios relacionados con la inscripción de materias de semestres anteriores que el estudiante tenga pendientes (ahora puede verlas cuando quiera, siempre que cumpla con los prerrequisitos). Igualmente se puede cancelar una materia hasta una semana antes de exámenes finales; la inasistencia del 20% a una asignatura, siempre y cuando esta no sea práctica no implicara su perdida. Todas estas medidas pretenden disminuir la deserción universitaria y dar mayores posibilidades a los estudiantes de clases menos favorecidas.[8]

En la Universidad Tecnológica de Bolívar, el problema de la deserción universitaria esta siendo enfrentado con estrategias de tipo financiero tales como dar mayor flexibilidad al pago de la matrícula, apoyado por la diversificación del portafolio de ayuda financiera (convenio con bancos, tarjetas de crédito, crédito directo, crédito ICETEX, monitorías, prácticas administrativas, becas, descuentos), y estrategias de tipo académico tales como la revisión del reglamento académico, programa de recuperación y asistencia académica, nuevo modelo pedagógico, programa de formación de docentes en habilidades pedagógicas, entre otros.[9]

BODEGAS DE DATOS Y MINERÍA DE DATOS

En la edición anterior de la revista, Pág. 33 [10], el autor menciona las condiciones bajo las cuales se aplica la minería de datos, con sus correspondientes definiciones y marcos teóricos de aplicación.

Las técnicas de minería de datos, permiten la clasificación de datos según un contexto o comportamiento, por ejemplo de ingresos o retiros de estudiantes. Adicionalmente permite obtener comportamientos por medio del análisis secuencial; permite la totalización de datos y la visualización de los mismos por medio de diferentes representaciones (árboles de decisión, funciones lineales o no lineales, modelos de probabilidad, por mencionar solamente algunos).

Los resultados que se esperan tener de la aplicación de un proceso a través de minería de datos, son uno o varios modelos de decisión que deben haber sido probados suficientemente, para que puedan ser aplicados por uno o varios departamentos o secciones de la universidad. Estos modelos se comportaran como agentes inteligentes, donde la inteligencia estará dada por la aplicación de las políticas y reglas de la universidad. Dos modelos usados frecuentemente en minería de datos son el descriptivo y el predictivo.

Modelamiento descriptivo. Este modelamiento presenta las características fundamentales de los datos a ser estudiados, sin que para ello se creen trabas a la cantidad de los datos. Como resultado se obtendrá un entendimiento mas concreto de las relaciones de los datos y sus correspondientes estructuras. Un modelo obtenido de esta manera se considera que es Generativo, debido a que los datos generados tendrán las mismas características de los datos reales de los cuales se generan. [12]

Modelamiento predictivo (Ver figura 4). El objetivo principal al aplicar un modelo predictivo, es el de encontrar el valor no conocido de una variable, dados los valores o comportamientos de otras variables. Por ejemplo en el caso planteado como titulo de este artículo, el de encontrar cuantos estudiantes, nuevos en la universidad desertaran de su carrera y de la universidad. Este tipo de modelamiento se puede presentar como la relación un grupo de valores en un vector y sus resultados esperados, obtenidos por medio de una función $y = f(x; \square)$ donde x son los valores medidos y \square los parámetros del modelo.

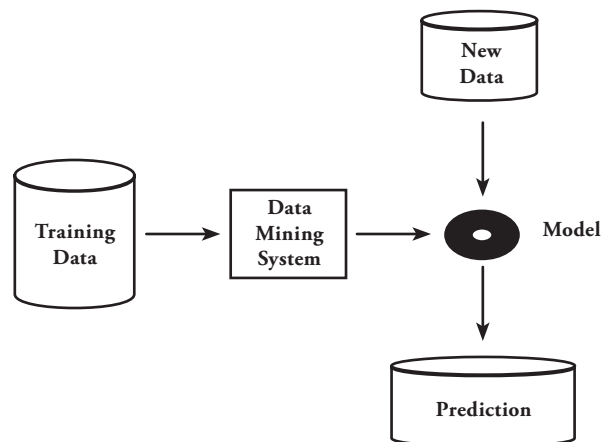


Figura 4. Proceso para realizar el modelamiento predictivo en minería de datos. [13]

La minería de datos, también tiene como característica que es un proceso que tiende a ser orientado a hallar comportamientos a nivel de detalle, como por ejemplo encontrar a aquellos estudiantes que prefieren un grupo particular de asignaturas orientadas por algún docente; notas obtenidas por estudiantes que en el último periodo fueron superiores al promedio, en un grupo particular, sin que esto signifique que no se puedan tener resultados a nivel totalizado como por ejemplo cuantos estudiantes de los que desertan tienen promedio académico inferior al mínimo establecido.

Otra técnica recientemente empleada en mercadeo utilizando herramientas de minería de datos, muy relacionada con la retención de clientes, es el Análisis de supervivencia o análisis de eventos en el tiempo (time to event analysis). Ésta está orientada a entender a los clientes cuando hay motivos por los cuales preocuparse, relacionados estos con la ocupación que están haciendo de los servicios de la organización y determinar cuáles factores de los que tienen que ver con una relación comercial tienen el mayor efecto en la permanencia de un cliente.

Las características de esta última técnica es la de proveer el entendimiento de eventos en el tiempo tales como [11]:

- Cuándo será que un cliente se retirará.
- Cuál será la siguiente ocasión que un cliente se cambiara a otro segmento del mercado.
- Cuál la siguiente ocasión que un cliente ampliará o disminuirá la relación comercial.
- Cuáles son los factores en la relación comercial que incrementan o disminuyen el uso de los recursos por parte de los clientes.

Hay dos valores muy importantes que soportan la información de retención o supervivencia de los clientes: la fecha de inicio y la fecha de finalización de la relación comercial. En algunas aplicaciones estos son bastante evidentes, como por ejemplo el inicio y finalización de un semestre académico, un curso de capacitación o inscripción a alguna revista o periódico. Pero en otras aplicaciones, como las transaccionales, estos datos no son muy evidentes. Con los datos proporcionados por los periodos de retención de clientes, se puede hacer una grafica o curva de retención, la cual mostrara el porcentaje de clientes que se retienen para un periodo de tiempo particular (ver figura 5). En esta curva los clientes que están en los mayores periodos de retención también lo están en los menores, es decir tiene un comportamiento de Histograma y permitirá establecer cuando la mitad de los clientes saldrá o dejará de tener vínculos con la organización (aproximadamente 20 meses, en la grafica) y también a partir de ella, el tiempo promedio de retención de un cliente (ya que el área bajo la curva es la cantidad de clientes dividido por la cantidad total de clientes (aproximadamente 45 meses en la grafica).

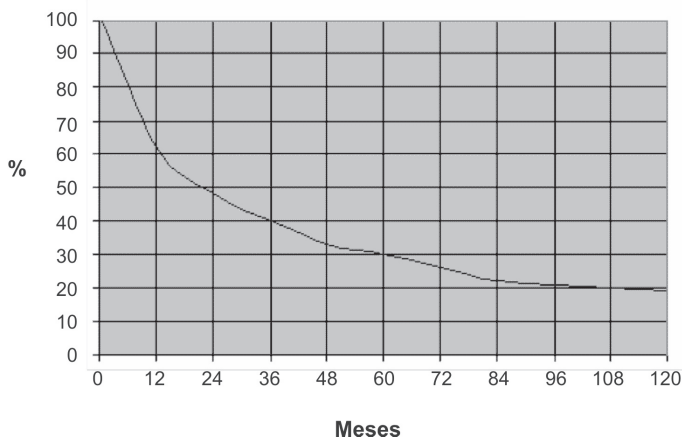


Figura 5. Curva de retención de clientes por periodos de tiempo con duración total de 10 años. [11]

Un análisis de retención se complementara en forma mas precisa con el análisis de amenazas (Hazard analysis) el cual posteriormente podrá servir de base para el análisis de supervivencia. Mientras que el primero permite responder a preguntas como ¿cuántos clientes que han sobrevivido a una cantidad de tiempo tx, saldrán o se retiraran en el tiempo tx +1? la segunda determina la probabilidad que hay que un cliente permanezca después de ese periodo de tiempo (los valores de supervivencia normalmente se calculan a partir de los valores dados por el análisis de amenazas).

LA MINERÍA DE DATOS Y LOS PROCESOS DE ADQUISICIÓN Y RETENCIÓN DE ESTUDIANTES

El proceso a realizar se encuentra enmarcado en las condiciones necesarias para las cuales la minería de datos se aplica. Un buen punto base para el inicio, será tomar el plan estratégico que la universidad tenga y las proyecciones planeadas para el logro del mismo. Todo proyecto de minería de datos es factible si se tiene la necesidad y los datos para realizarlo. Visto de otra manera, quiere decir que al interior de la Universidad se puede realizar cualquier proyecto relacionado con minería de datos, siempre que se tenga una necesidad real que conlleve a una mejora de las condiciones de la organización, mejores condiciones económicas, aumento de los índices de participación en el mercado y se cuente con la suficiente cantidad de datos sobre los cuales se puedan entrenar los diferentes algoritmos o establecer las condiciones necesarias para su aplicación. El análisis de factibilidad económico permitirá determinar si los costos que conlleva el montar todo el proceso de minería de datos, resultará en un producto que genere beneficios económicos a la organización.

El plan estratégico tiene varios objetivos institucionales (como por ejemplo proyección social, calidad administrativa y académica) los cuales a su vez tienen objetivos estratégicos (Generar ofertas

educativas que se adapten a las necesidades de la comunidad, desarrollar e innovar programas y procesos curriculares, promover el desarrollo de líneas y grupos de investigación tanto en pregrado como en posgrados, entre otros) basados a su vez en proyectos necesarios para el logro de los mismos. Algunos de estos proyectos son: Establecer relaciones con entidades gubernamentales y educativas; ofrecimiento de nuevos programas académicos de pregrado y posgrado; virtualización de componentes curriculares y de educación; capacitación continua de docentes; adquisición de nuevos estudiantes y disminución del índice de deserción universitaria. Cada uno de estos proyectos puede tener condiciones similares de factibilidad por lo que habrá que determinarse de alguna manera cual de ellos será el que inicialmente se lleve a cabo. Una forma de realizar esta selección puede ser a través de la asignación de prioridades establecidas con lo que la universidad espera obtener en corto y mediano plazo para el logro de sus objetivos.

Siendo los estudiantes quienes aportan la mayor cantidad de dinero con el cual se financia la Universidad, no es difícil determinar que uno de los proyectos con el cual se puede iniciar un proceso de minería de datos es el de la adquisición de nuevos estudiantes y evitar el retiro o deserción de los mismos.

Cuando se trata de establecer las estrategias de adquisición de nuevos estudiantes, la fuente de la información esta en los datos históricos que se tienen en el datamart de estudiantes. Las variables a identificar serán aquellas que relacionen a los estudiantes que hayan terminado sus estudios graduados o no y los que aun están cursando algún semestre, con el estrato social del que provienen, el área geográfica de la ciudad donde residen, la institución donde realizaron sus estudios primarios y secundarios, el valor de la matrícula que pagan en la universidad, la carrera que cursaron, ubicación de los estudiantes egresados en cargos oficiales o privados, facilidad de ubicación laboral durante o después de cursar la carrera, etc. Del análisis de cada una de estas variables puede generarse un plan de mercadeo orientado a fortalecer aún más las áreas donde se es fuerte y a aumentar la participación en aquellas donde se tiene debilidad o ninguna participación.

Igualmente según lo planteado en los numerales 2 y 3 de este artículo, también se pueden desarrollar modelos de minería de datos para el retiro (o deserción) voluntaria o no de estudiantes de la universidad. Es necesario hacer la diferenciación, pues las causas suelen ser diferentes. Las razones involuntarias pueden ser salir de la universidad porque el promedio académico exigido no le permite continuar o por sanciones disciplinarias y voluntarias por falta de recursos económicos, exigencia laboral en la entidad donde trabaja o va a trabajar, a que encontró otras opciones dependiendo de su visión de calidad académica o costos de los estudios mas baratos en otra universidad con igual o similares proyecciones profesionales, por mencionar algunos.

En cualquiera de los casos planteados habrá necesidad de desarrollar varios modelos de minería de datos basados en algoritmos de segmentación, o de predicción, o de clasificación, realizar el análisis de sobrevivencia (survival análisis) [11] explicados en la sección 4 de este artículo y en las graficas obtenidas como por ejemplo curvas de retención (ver figura 5), o curvas de amenazas, curvas de sobrevivencia. Los resultados permitirán a la universidad crear las acciones necesarias para encontrar su nicho en el mercado, mejorar y flexibilizar los programas académicos y proporcionar planes de financiación que le permitan al estudiante continuar sus estudios basados en algún plan especial de pagos, financiación, becas, planes alternos de estudio, programas especiales de nivelación, etc.

LA MINERÍA DE DATOS Y LOS PROCESOS RELACIONADOS CON EL CICLO DE VIDA DEL ESTUDIANTE

Existen datos suficientes recopilados a través del ciclo de vida que un estudiante realiza en la Universidad para llevar a cabo los procesos relacionados en la sección anterior y en cada uno de ellos, se puede realizar la gestión de manera más detallada, relacionada con la minería de datos. Estos pasos secuenciales o no pueden ser:

Inscripción para el primer semestre. Se hace a través del formulario de inscripción, el cual contiene los datos básicos del nuevo prospecto a ser estudiante. Los datos suministrados alimentan la base de datos de estudiantes (y/o el datamart de estudiantes). La ayuda que Las técnicas de clasificación, predicción, agrupamiento o estimación pueden dar a la universidad, consiste en conocer cuántos estudiantes se inscribieron, de qué colegio o institución salieron, en qué año y como se enteraron de la universidad. Adicionalmente la minería de datos, establecerá qué comportamiento tiene la inscripción dependiendo de si es el primer o segundo semestre académico.

Matricula estudiantes primer semestre. En este paso se completa toda la información que la universidad requiere del estudiante, complementada con los resultados que al final del semestre el estudiante haya obtenido (por ejemplo los resultados de las notas obtenidas). Las técnicas a usar son muy similares a las del paso descrito anteriormente, las cuales permiten clasificar al estudiante en algún grupo en particular, por medio de patrones encontrados luego de entrenar algún algoritmo de estimación y/o predicción. Esto permitirá posteriormente, establecer con mayor exactitud, cuáles son las características o comportamientos del grupo de estudiantes que continuara sus estudios en otros semestres académicos o también determinar al grupo de estudiantes que deserto y cuales sus posibles causas.

Inscripción y matrícula de estudiantes en semestres posteriores al primer semestre. Los comportamientos de los estudiantes para semestres posteriores al tercer semestre tienden, según las estadísticas a ser muy estables. En este paso los estudiantes que continúan sus estudios, semestre a semestre proporcionan información con los resultados obtenidos en las diferentes actividades académicas y de participación de la universidad. Por lo tanto la ayuda de la minería de datos puede ser bastante restrictiva, y orientada fundamentalmente a la inscripción de asignaturas que son electivas y orientadas por algún docente en particular. El perfil del estudiante será más completo al igual que sus correspondientes patrones de comportamiento. La ayuda en este caso será para las directivas de la universidad, al poder gestionar los recursos, como planta física, número de docentes, asignación de salones, número de cursos, horarios, etc.

Retiro (deserción) de estudiantes por diversas causas. Cuando un estudiante se retira de la universidad, se le proporciona un formato donde entre otros datos expone las razones por las cuales toma la decisión y cual es la probabilidad de regreso. Estos datos digitados en la base de datos de estudiantes (y/o en el datamart de estudiantes), permitirá establecer las características de los grupos de estudiantes que mas frecuentemente toman esta decisión; las causas comunes por las cuales la toman; permitirá establecer la correlación de variables y determinar algún patrón que modele las condiciones para realizar predicciones para la toma de decisiones.

En otras actividades relacionadas con el desarrollo del periodo lectivo tales como cancelación o adición de asignaturas, habilitaciones y validaciones, suspensión de estudiantes por promedio inferior al reglamento, suspensión de estudiantes por faltas graves, sustentaciones de tesis y graduación, suspensión temporal de estudios, reintegros de estudiantes, transferencias externas de otras universidades o transferencia interna a otra carrera son procesos que alimentan la base de datos de estudiantes, con datos que permitirán correlacionar las variables del sistema y hallar comportamientos y patrones ocultos.

CONCLUSIONES

Un factor clave de éxito en la aplicación de la minería de datos, es conocer o tener establecido cual es el objetivo de su aplicación, es decir saber cual es el problema que se pretende resolver. Tenerlo claro permite orientar y definir con un alto grado de éxito el desarrollo y resultados del proyecto a desarrollar.

También es fundamental establecer cómo se hará la aplicación del ciclo de vida de la minería de datos, pues ésta constituye la forma organizada y detallada de llevar a cabo el proceso y la obtención de un modelo que permita realizar el logro de los objetivos propuestos.

La minería de datos permite establecer comportamientos futuros y encontrar patrones de comportamiento ocultos, basados en datos históricos obtenidos en el transcurso de la vida de una universidad. No todos los modelos y algoritmos desarrollados dan solución a los problemas planteados o satisfacen sus necesidades. Por lo tanto es prioritario determinar cuales son los modelos de minería de datos que más se ajustan a los tipos de búsqueda y análisis de información para una institución de carácter universitario y en el caso del presente artículo a la deserción y/o retención de estudiantes o reclutamiento de nuevos estudiantes.

BIBLIOGRAFÍA

[1] Data Mining Applications in Higher Education. Jing Luan, PhD Chief Planning and Research Officer, Cabrillo College Founder, Knowledge Discovery Laboratories. SPSS

[2] La UNPA encabeza un estudio sobre las causas de la deserción educativa. Periódico Austral. Julio 1 de 2005.

[3] El factor educacional como causa potencial de la deserción en primer año de la universidad. Universidad Nacional de Río, Córdoba, Argentina. <http://www.unrc.edu.ar/publicar/cde/h21.htm>. Artículo bajado de internet el 25 de octubre de 2005.

[4] BOADO, MARCELO. Una aproximación a la deserción estudiantil universitaria en Uruguay. Publicado por Instituto Internacional para la Educación Superior en América Latina y el Caribe. UNESCO.

[5] Latiesa, Margarita. Tipología y causas de la deserción universitaria y el retraso en los estudios.

<http://dialogo.ugr.es/anteriores/dial05/11-5.htm>. Artículo bajado de Internet el 25 de octubre de 2005.

[6] Estadísticas e indicadores de la universidad Nacional de Colombia, basada en La información estadística del Departamento Nacional de Planeación, Boletín No 27, Educación y fuerza de trabajo, Bogotá, septiembre de 2000, y el Informe de Desarrollo Humano para Colombia, 1999.

[7] Caicedo C., Guarino (2005, Junio 20). COLOMBIA; Alto porcentaje de deserción universitaria; Estadística alcanza el 52% y es considerada como una verdadera 'tragedia nacional'. El Diario La Prensa, p. 14.

[8] Apoyando la formación superior de los estratos I, II y III. http://www.colombiaaprende.edu.co/html/docentes/1596/article75219.html#h2_6. Bajado de Internet el 22 de octubre de 2005.

[9] 25. Estrategias para disminuir la deserción en la Universidad Tecnológica de Bolívar.

[10] Castillo Santos, Rafael, Detección de intrusos mediante técnicas de minería de datos. Revista Clepsidra, Numero 2, 2006.

[11] M. Berry, G. Linoff, “Data mining techniques for marketing, sales and customer relationship management”, Wiley Publishing, Inc. Indianapolis, 2004, pp. 43 – 86.

[12] Hand, David; Mannila, Heikki and Smyth, Padhraic Principles of Data Mining. The MIT Press © 2001 (546 pages)

[13] Thearling, Kurt Ph.D. An Introduction to Data Mining.
www.thearling.com

