Apunt. cienc. soc. 2018; 08(01)
ARTÍCULO ORIGINAL
DOI: http://dx.doi.org/10.18259/acs.2018005

Detección de patrones de éxito en estudios universitarios de la Universidad Continental

Detection of the success patterns in the university studies of the Universidad Continental

Daniel Gamarra Moreno [1,*], Rocio Matos Barzola [1], Miguel Tupac Yupanqui Alanya [1]
[1.] Universidad Continental

RESUMEN

El objetivo del estudio fue detectar los patrones de éxito en los estudios universitarios de los estudiantes de la Universidad Continental a partir de la información académica y socio-demográfica recopilado en el primer ciclo de estudios. La investigación tuvo un diseño descriptivo transversal, los estudiantes fueron de diversas carreras profesionales que desarrollaron sus estudios entre el 2012 y 2017. Para la extracción automatizada de conocimiento se utilizó la metodología de proyectos de minería de datos Cross-Industry Standard Process for Data Mining (CRISP-DM) y el software Clementine, mediante la red neuronal perceptrón multicapa se logró identificar las variables que más impactan en el éxito y abandono de los estudios universitarios, para obtener el árbol de decisión se aplicó el algoritmo C5.0 y para el agrupamiento (tres grupos) los algoritmos k-means y TwoStep, y estos resultados se compararon mediante una matriz de confusión. Los resultados muestran que los estudiantes que abandonaron sus estudios universitarios no pasaron del quinto ciclo; las variables que más influyen en el éxito de los estudios universitarios son: estado civil del estudiante, monto de la pensión, estado civil de los padres y con quien vive el estudiante en Huancayo. Por otro lado, las variables que más influyen en el abandono de los estudios son, con quien vive el estudiante, estado civil del estudiante, satisfacción del desempeño del docente y estado civil de los padres. En conclusión las variables socio demográficas son las que tienen mayor predominancia sobre el éxito y fracaso de los estudios universitarios.

Palabras clave: Minería de datos, deserción universitaria, detección de patrones.

ABSTRACT

The objective of the study was to detect the success patterns in the university studies of Universidad Continental students from the academic and socio-demographic information collected in the first cycle of studies. The research had a cross-sectional descriptive design, the students were from diverse professional careers that developed their students between 2012 and 2017. For the automated extraction of knowledge the methodology of data mining was used (CRISP-DM) and the Clementine software, throught the multilayer perceptron neural network, it was possible to identify the variables that most impact the success and abandonment of university studies, To obtain the decision tree,algorithm C5.0 was applied and for grouping (three groups) the algorithms k-means and TwoStep, and these results were compared using a confusion matrix. The results show that the students that abandoned their university studies did not pass the fifth cycle. The variables that most influence the success of the university studies are: student`s civil status, amount of the pension, marital status of the parents and with whom the student lives in Huancayo. On the other hand, the variables that most influence the abandonment of the studies are , whit whom the student lives, the student marital status, satisfaction with the teacher's performance and the parent`s marital status. In conclusion, the socio demographic variables are those that have the most predominance over the success and failure of university studies.

Keywords: Data mining, university dropout, detect patterns.

Historial del artículo:
Recibido, 04 de marzo 2017; aceptado, 10 de diciembre de 2017; disponible en línea, 05 de enero de 2018

* Doctor en Ingeniería, Jefe de la Oficina de Registros Académicos de la Universidad Continental.
Correo: dgamarra@continental.edu.pe

INTRODUCCIÓN

Los estudiantes que tienen éxito o fracaso en sus estudios universitarios tienen características comunes que se pueden determinar a partir de la información histórica. En el estudio, el éxito se da cuando el estudiante aprueba todas las asignaturas del plan de estudios en los 10 ciclos académicos o menos, y fracaso si abandona sus estudios. La minería de datos aplicada a la información académica y socio-demográficas permite detectar patrones de éxito en los estudios universitarios. La caracterización de los estudiantes permitirá generar políticas de incentivos, programas de mejora y otros; dirigido a los estudiantes de manera preventiva. El problema del estudio es: ¿Cuál son los patrones de éxito y fracaso en sus estudios universitarios de los estudiantes ingresantes de la Universidad Continental? A partir de esta interrogante se formula el siguiente objetivo: detectar los patrones de éxito en los estudios universitarios de los estudiantes de la Universidad Continental a partir de la información académica y socio-demográficas recopilado en el primer ciclo de estudios.

En un estudio realizado por Ocaranza (2006) se logró establecer el perfil de los desertores usando la metodología de descubrimiento de conocimiento de base de datos. Donde, deserción es: "el retiro voluntario u obligatorio de los alumnos de una carrera al finalizar el primer año académico".

Timarán (2009) utilizando la herramienta de descubrimiento del conocimiento TariyKDD. Los algoritmos de minería de datos utilizados fueron C4.5 y EquipAsso para la clasificación y asociación respectivamente.

Spositto, Etcheverry, Ryckeboer y Bossero (2009) aplicaron el proceso de descubrimiento de conocimiento para encontrar el árbol de decisiones del rendimiento académico, y detectar los patrones de deserción estudiantil con los algoritmos de minería de datos J48 y FT, disponible en el Software Weka.

En el estudio se aplicó el proceso de descubrimiento de conocimiento en bases de datos, con datos académicos y socio-demográficas de los estudiantes de la cohorte 2009-1. Con esa información se llegó a determinar el perfil de los estudiantes de que aprueban el plan de estudios y de los que abandonan. El proceso estándar entre industrias para el modelo de proceso de minería de datos, CRISP-DM, es un modelo iterativo que permite el refinamiento de la minería de datos (Chapman, Clinton, Kerber, Khabaza, Reinartz, Shearer y Wirth, 2000). La selección de la información se complica cuando está dispersa y en medios no digitales. En este estudio el esfuerzo de selección, transformación y limpieza de datos demandó el 50% del esfuerzo total.

MATERIAL Y MÉTODOS

Para el estudio se utilizó los datos de 1048 estudiantes que estudiaron entre el 2012 y 2017 y cuya ficha socioeconómica y avance academico estaban almacenados en los sistemas de información de la Universidad. De ellos se realizó un seguimiento de su avance académico hasta el año 2017, carreras profesionales de 10 ciclos académicos.

Se definió como éxito en los estudios universitarios cuando el estudiante aprobó todas las asignaturas de su plan de estudio de su carrera en los 10 ciclos académicos o menos. Si el estudiante desertó antes de culminar su carrera se consideró como fracaso. Para el estudio, deserción es el retiro voluntario del estudiante antes de culminar su carrera universitaria. También, se consideró el estado "estudiando" en el caso que no completo su plan de estudios y no desertó.

Por las características del estudio se aplicó el diseño descriptivo transversal, la metodología CRISP-DM y se utilizó el software Clementine.

La metodología CRISP-DM es un estándar de facto en los proyectos de minería de datos, donde la minería de datos es uno de los pasos del proceso de descubrimiento de conocimiento en bases de datos (Mariscal, Marbán, y Fernández, 2010). Este modelo describe las actividades que deben realizarse para desarrollar un proyecto de minería de datos, sus fases son: entendimiento del negocio, entendimiento de los datos, preparación de datos, modelamiento, evaluación de resultados y despliegue de resultados (Chapman et al., 2000).

El acceso a las universidades latinoamericanas ha aumentado, pero son deficitarios en cuanto a la permanencia e inserción laboral de los jóvenes provenientes de familias con menores ingresos. La repitencia es la acción de cursar reiterativamente una asignatura, sea por mal rendimiento del estudiante o por causas ajenas al ámbito académico produciendo su deserción (CINDA, 2006). La universidad Continental ha creado programas de intervención a partir de la detección de patrones de los estudiantes aprobados y desaprobados en las asignaturas con mayor cantidad de desaprobados. En ese sentido el presente estudio tiene como objetivo: detectar los patrones de éxito y fracaso en los estudios universitarios de la Universidad Continental usando las técnicas de minería de datos.

Para el primer modelo se consideró 35 variables del desempeño académico, nivel de satisfacción de su docente y socio-demográficas. Como primer paso, se extrajo 1255 registros, los cuales correspondían datos de los ingresantes del semestre 2009-1. En el proceso de selección, limpieza y transformación; se eliminaron y corrigieron los datos incorrectos y se decidió la

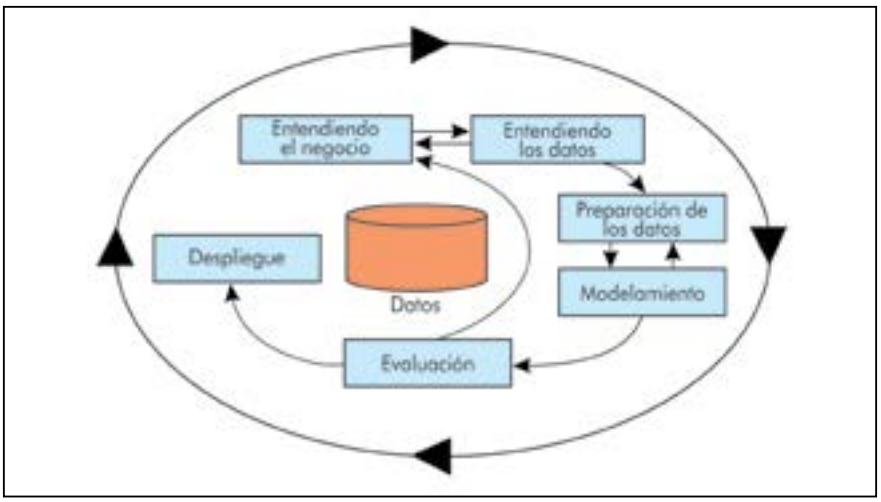

Figura 1. Proceso estándar entre industrias para el modelo de proceso de minería de datos, CRISP-DM (Chapman et al., 2000).

estrategia a seguir con los datos incompletos, los datos atípicos (outliers) y valores desaparecidos (datos missing). Luego de estas consideraciones se obtuvieron 1048 registros de los 1255 que se contaba en un inicio. De estos 1048 registros podemos indicar que 526 son varones, 522 son mujeres, 508 provienen de colegios estatales y 540 provienen de colegios particulares, 36 se encuentran en la etapa de adolescencia (etapa intermedia) que comprende de los 14 años a los 16 años, 792 se encuentran en la etapa de adolescencia (etapa tardía) que comprende de los 17 años a los 19 años y 220 se encentran en la etapa adulta temprana que comprende de los 19 años a los 45 años. Así mismo, 1021 son solteros, 21 son casados y

Tabla 1
Variables del modelo final.

Variable	Campo	tipo	valores
Genero del estudiante	Sexest	Set	H,M
Estado Civil Estudiante	EstCivEst	set	1 Divorciado, 2 Casado, 3 Conviviente, 4 Soltero, 5 Viudo
Estado Civil Padres	EstCivPad	Set	1 Divorciado, 2 Madre Soltera, 3 Padre Soltero, 4 Viudo, 5 Conviviente, 6 Casado
Dependencia económica del estudiante	DepEconEst	Flag	Dependiente, Independiente
¿Con quién vive en Huancayo?	CqViveHyo	Set	1 Otros, 2 Sólo, 3 Familiares, 4 Hermanos, 5 Sólo Mamá, 6 Sólo Papá, 7 Padres
Composición familiar	ComposFam	Range	1,2,3,4,5
Ingreso Familiar	IngFamilia	Ordered Set	1,2,3,4,5,6
Monto que percibe el jefe del hogar	MontoHaber	Ordered Set	1,2,3,4,5,6
Tipo de vivienda en Huancayo	UvivactHyo	Flag	Zona Rural, Zona Urbana
Escala de pensión	Escala	Ordered Set	A, B, C, D, E, F, G, X
Estado del estudiante	EstadoEst	set	1 Abandono, 2 Estudia, 3 Egreso
Tiempo estudiando	TiempoEst	Range	[1,5.5]
Nivel de Satisfacción Docente	SatisDoc	Range	0,1,2,3,4,5,6
Cantidad de asignaturas aprobadas	CantAsigAp	Range	0,1,2,3,4,5,6,7
Cantidad de asignaturas desaprobadas	CantAsigDes	Range	0,1,2,3,4,5,6,7

6 son convivientes. El refinamiento del modelo inicial al modelo final dejo 15 variables de las 35 iniciales (Tabla 1).

Las estadísticas descriptivas para la cohorte 2012 muestran que sólo el 12% de lo ingresantes culminaron sus estudios en 10 periodos académicos o menos (Tabla 2). El porcentaje de abandono en el primer semestre es de 49.4 (19.3% más 30.1%), el más alto de todos (Tabla 3). Los abandonos se dieron hasta el cuarto ciclo de estudios. En el ciclo de estudios igual a cero se contabilizan a los que no aprobaron ninguna asignatura y abandonaron sus estudios.

La figura 2 muestra una separación lineal entre la relación cantidad de asignaturas aprobadas y desaprobadas. Este conocimiento se usó para mejorar el modelo e insertar una nueva variable derivada ratio de aprobación (ratioAprob) igual a la cantidad de asignaturas aprobadas entre la cantidad de asignaturas cursadas.

Tabla 2
Situación académica al 2017 de los estudiantes de la cohorte 2012.

Estudios universitarios	Frecuencia	Porcentaje	Porcentaje acumulado
1. Abandono	502	48,2	48,2
2. Estudia	413	39,7	87,9
3. Egreso	126	12,1	100,0
Total	1041	100,0	

Tabla 3
Abandono al 2017 de los estudiantes de la cohorte 2012.

Ciclo de estudios alcanzado	Frecuencia	Porcentaje	Porcentaje acumulado
0	97	19,3	19,3
1	151	30,1	49,4
2	100	19,9	69,3
3	85	16,9	86,3
4	69	13,7	100,0
Total	502	100,0	

Para determinar la importancia de los atributos (variables) en el abandono y el egreso (quienes aprobaron el plan de estudios) se utilizó la red neuronal perceptrón multicapa con un acierto de 98% en ambos casos (Perez y Santin, 2006). Para el árbol de decisión se aplicó el modelo C5.0 con un 65.97% de acierto. Para comparar el desempeño de los algoritmos k-means y TwoStep, para tres grupos, se utilizó una matriz de confusión que mostró que sólo 19 de los 238 ejemplos tendrían un agrupamiento diferente (Hernández, Ramírez y Ferri, 2008).

Los modelos fueron validados usando la validación cruzada, dividendo los datos en 77% de entrenamiento y 23% de prueba (Hernández et al., 2008).

RESULTADOS

El árbol de decisión se descartó por el bajo acierto y los resultados del estudio muestran en las siguientes figuras.

La figura 3 nos muestra que los estudiantes que terminan su carrera tienen en promedio una ratio de aprobación de más del 0.80 en el primer ciclo de estudios.

De la figura 4, se infiere que los estudiantes que abandonaron su carrera universitaria no pasaron del 5to ciclo y en promedio los estudiantes abandonan sus estudios en 1,76 ciclos de estudio.

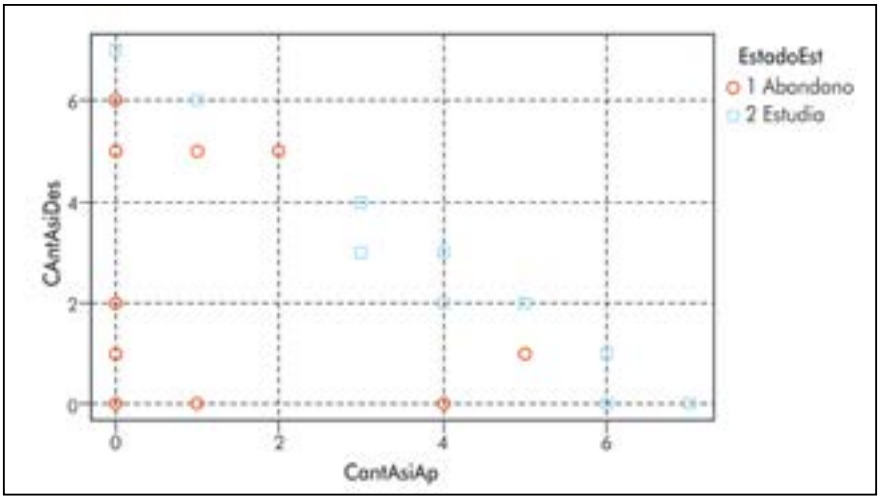

Figura 2. Dispersión entre la cantidad de asignaturas aprobadas versus desaprobadas en el primer ciclo de estudios.

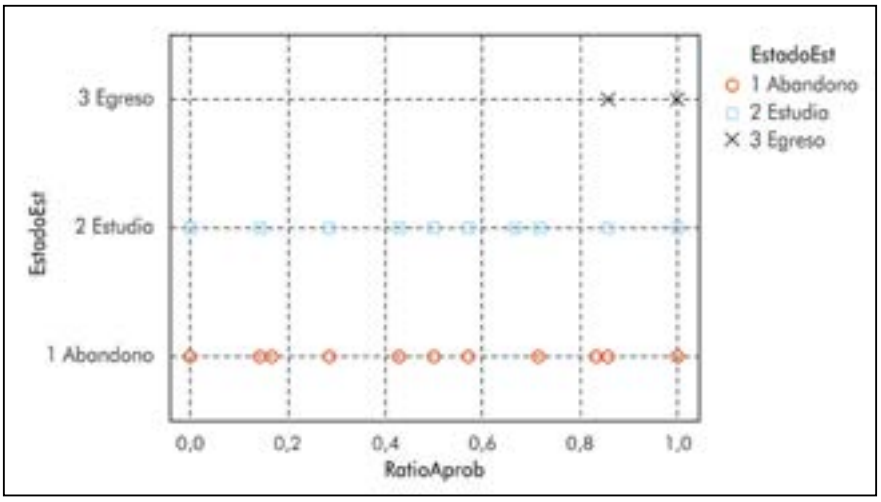

Figura 3. Ratio de aprobación(ratioAprob) vs estado del estudiante (estado Est).

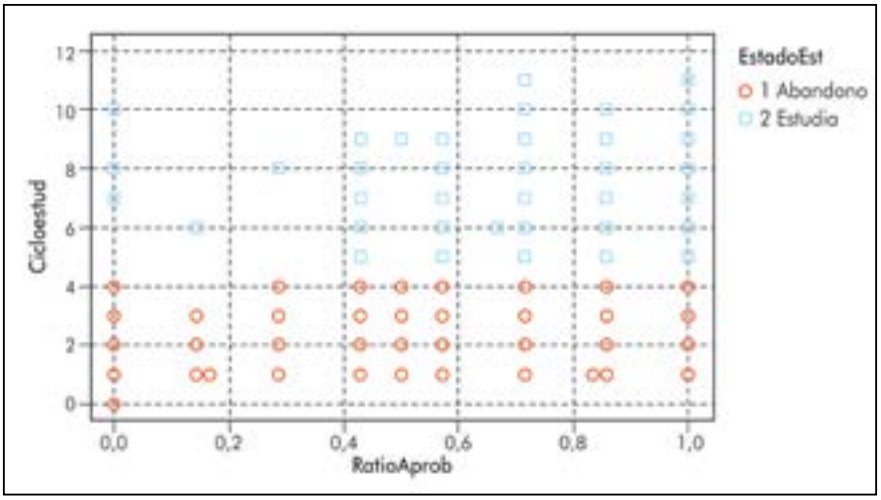

Figura 4. Ratio de aprobación versus ciclo de estudios.

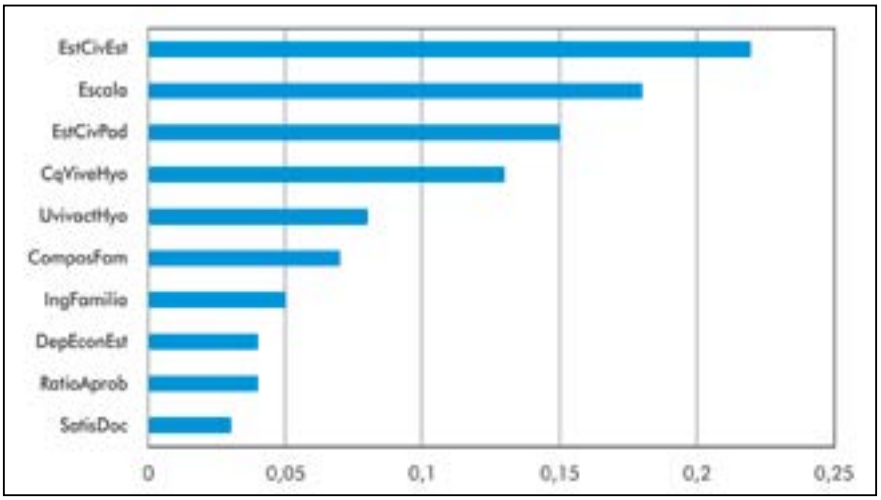

Figura 5. Importancia de los atributos cuando el estudiante termina su carrera.

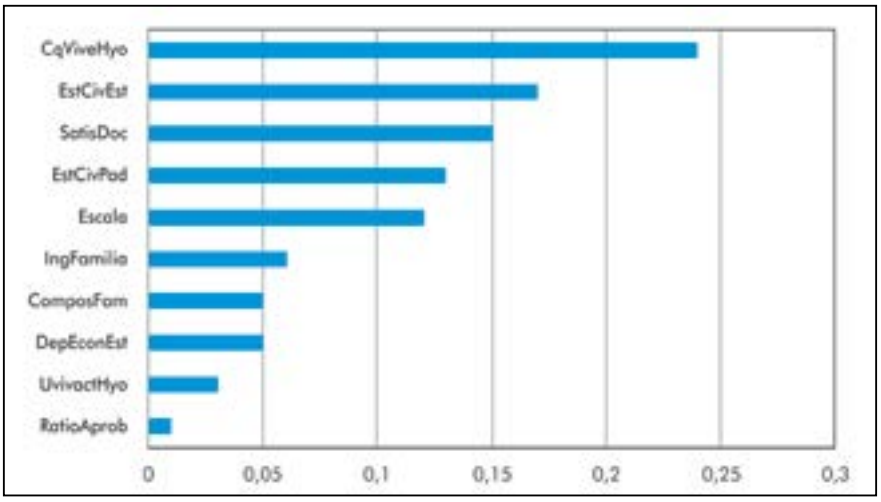

Figura 6. Importancia de los atributos cuando el estudiante abandona la universidad.

Las cuatro variables que más influyen en los estudiantes que terminan su carrera en diez o menos periodos académicos son: estado civil del estudiante, el monto de la pensión, estado civil de los padres y con quien vive en Huancayo (figura 5). Por otro lado, las cuatro variables que más influyen en los estudiantes que abandonan su carrera son: estado civil de los padres, con quien vive en Huancayo, satisfacción del desempeño del docente y estado civil de los padres (Figura 6).

DISCUSIÓN

Los estudios de deserción universitaria realizados en las universidades de Latinoamérica se basaron en datos académicos y socio-demográficas. Estos coinciden en que la mayoría de estudiantes dejan sus estudios universitarios en los primeros ciclos de estudios. Las variables que influyen en la deserción son las condiciones socioeconómicas del estudiante y de la familia, aspectos de orden personal (actitudinales y motivacionales) y los aspectos académicos. Estos se vinculados con la separación familiar y la adaptación a los estudios universitarios.

El porcentaje de estudiantes universitarios que terminan su carrera en el tiempo previsto es bajo: en argentina el 30%, en Bolivia el 5%, en México el 60% y en la universidad Continental para la cohorte estudiada (Perú) 12%.

REFERENCIAS BIBLIOGRÁFICAS

Chapman, P., Clinton, P., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. CRISP-DM 1.0. Step-by-step data mining guide (2000). [Software de computación]. Clementine SPPS (versión 12,0). SPSS Inc.

CINDA. (2006). Repitencia y deserción universitaria en America Latina Colección Gestión Universitaria, L. González (Ed.).

Hernández Orallo, J., Ramírez Quintana, J., & Ferri Ramírez, C. (2008). Introducción a la Minería de Datos. Madrid: Pearson Prentice Hall.

Mariscal G, Marbán Ó, Fernández C. A survey of data mining and knowledge discovery process models and methodologies. The Knowledge Engineering Review.

Ocaranza, O. and M. Quiroz (2005). Deserción estudialtil en el pregrado en la Pontificai Universidad Católica de Valparaíso, Chile. Chile.

Perez López, C., & Santin Gonzalez, D. (2006). Data Mining - Soluciones Con Enterprise Miner Con 1 CD: Alfaomega Grupo Editor.

Spositto, O., Etcheverry, M., Ryckeboer, H., & Bossero, J. (2009). Aplicación de técnicas de minería de datos para la evaluación del rendimiento académico y la deserción estudiantil.

Timarán Pereira, R. (2009). Detección de Patrones de Bajo Rendimiento Académico y Deserción Estudiantil con Técnicas de Minería de Datos. Paper presentado en la Octava Conferencia Iberoamericana en Sistemas, Cibernética e Informática: CISCI 2009, Orlando, Florida ~ EE.UU.