

- David Luis La Red, Corrientes (Argentina)
Carlos Enrique Podestá, Curuzú Cuatiá (Argentina)

Metodología de Estudio del Rendimiento Académico Mediante la Minería de Datos

Methodology Study of Academic Performance Using Data Mining

RESUMEN

La minería de datos orientada a la educación permite predecir determinado tipo de factor o característica de un caso, fenómeno o situación. En este artículo se describen los modelos de minería utilizados y se comentan los principales resultados obtenidos. Se consideran especialmente modelos de minería de agrupamiento, clasificación y asociación. En todos los casos se busca determinar los patrones de éxito y de fracaso académico de los alumnos para, de esta manera, predecir la probabilidad de los mismos de desertar o tener un bajo rendimiento académico, con la ventaja de poder hacerlo tempranamente, permitiendo así encarar acciones tendientes a revertir tal situación. Este trabajo se ha realizado en el año 2013, con información de los años 2009 a 2013, con alumnos de la asignatura Sistemas Operativos de la carrera terciaria de Tecnicatura Superior Analista Programador (TSAP) del Instituto Superior de Curuzú Cuatiá (ISCC), de Corrientes, Argentina.

ABSTRACT

Data mining, education-oriented, predicts certain type of factor or characteristic of a case, phenomenon or situation. Mining models used are described in this article and the main results are discussed. Especially mining models clustering, classification and association are considered. In all cases is to determine patterns of academic success and failure for students, thus predicting the likelihood of dropping them or having poor academic performance, with the advantage of being able to do it early, allowing face action to reverse this situation. This work was carried out in 2013, with information for the years 2009 to 2013, with students of Operating Systems, subject to the Superior Technical Analyst-Programmer (TSAP), who studied at the Higher Institute of Curuzú Cuatiá (ISCC), Corrientes, Argentina.

PALABRAS CLAVE / KEYWORDS

Rendimiento académico, Minería de datos, Perfiles de alumnos, Sistemas operativos, Enseñanza-aprendizaje, Almacenes de datos. Academic performance, Data mining, Student profiles, Operating systems, Teaching and learning, Data warehouse.

SOBRE EL AUTOR/ES

Dr. David Luis La Red Martínez. Dpto. de Informática, Facultad de Ciencias Exactas, Universidad Nacional del Nordeste, Profesor Titular (laredmartinez@gigared.com).

Mgter. Carlos Enrique Podestá Gómez. Rectoría, Instituto Superior Terciario de Curuzú Cuatiá, Profesor (cpodesta@acoanet.com).

1. Introducción

La carrera de Técnico Superior Analista Programador (TSAP) del Instituto Superior de Curuzú Cuatiá (ISCC) ha sido reiteradamente la primera en número de alumnos, considerando la totalidad del ISCC: 36,71%, y la que más egresados produce: 51,57% de dicha institución, porcentajes correspondientes a los respectivos informes elaborados para los años: 2006, 2007, 2008, 2009 y 2010 por el Departamento de Alumnos del ISCC.

Un análisis más detallado permite observar los relativamente bajos porcentajes de egresados respecto de nuevos inscriptos en la TSAP; estos porcentajes varían si se considera solamente el título terminal (Técnico Superior Analista Programador) o si además se considera el título intermedio (Operador en Sistemas de Computación). Sin considerar el título intermedio los datos son: 2006: 10,25%, 2007: 11,55%, 2008: 10,75%, 2009: 11,45%, 2010: 10,45%. Considerando el título intermedio: 2006: 21,81%, 2007: 23,22%, 2008: 21%, 2009: 23%, 2010: 22%.

Las relativamente bajas tasas de egresados respecto de nuevos inscriptos mencionadas en el apartado anterior, que podríamos considerar el “rendimiento académico global” de una carrera, se observan también en numerosas asignaturas de la TSAP, considerando “rendimiento académico particular” o simplemente “rendimiento académico”, a los resultados de las evaluaciones de los alumnos durante el cursado de una asignatura, y la condición final lograda por los mismos en el marco de la Res. N° 1551/01 Reglamento Orgánico Marco (RAM) para los Institutos Superiores (régimen de evaluación y promoción: Art. 85 y 86): promocionado, regular o libre.

Para la asignatura Sistemas Operativos los valores de los últimos años son los siguientes: Alumnos promocionados y regularizados respecto de los que rindieron algún examen parcial: 2006: 16,25%, 2007: 27,45%, 2008: 30,55%, 2009: 28,50%, 2010: 30,39%. Además se ha observado que un considerable porcentaje de alumnos se inscriben para cursar la asignatura, pero luego no completan el cursado (55,39% en el 2010).

Ante la situación mencionada se consideró de gran importancia efectuar una investigación que permitiera determinar las variables que inciden en el relativamente bajo rendimiento académico de los alumnos de Sistemas Operativos de la TSAP del ISCC, perteneciente a la Dirección General de Educación Superior (DGES), identificar los perfiles de alumnos exitosos (los que promocionan o regularizan la asignatura), como así también los perfiles de alumnos que no lo logran (los que quedan en la condición de libres). Una vez determinados los perfiles de alumnos con bajo rendimiento académico, se podrían encarar acciones tendientes a evitar potenciales fracasos académicos. Para la determinación de los perfiles de alumnos se consideró apropiado utilizar técnicas de Almacenes de Datos (Data Warehouse: DW) y Minería de Datos (Data Mining: DM).

En este contexto, caracterizado por la masividad, falta de recursos en las proporciones adecuadas, bajo rendimiento académico, la aplicación de las NTICs sería un importante complemento del proceso tradicional de enseñanza – aprendizaje, constituyéndose a priori en una herramienta efectiva para intentar solucionar la situación antes mencionada (Acosta & La Red Martínez, 2012), (La Red Martínez & Acosta, 2012), (La Red Martínez et al., 2011), (Bolaños Calvo, 2001), (Peiró, 2001), (Wallace & Young, 2010), (IEEE, 2012).

El trabajo realizado partió de la siguiente hipótesis: la utilización de las herramientas de enseñanza-aprendizaje basadas en las Nuevas Tecnologías de la Información y las Comunicaciones (NTICs), incide en el rendimiento académico de los alumnos de Sistemas Operativos en la Tecnicatura Superior Analista Programador del ISCC, pero esta utilización de las NTICs está influenciada por diversas variables socio-económicas y actitudinales.

El objetivo general del trabajo fue: conocer las variables que inciden en el rendimiento académico de los alumnos de Sistemas Operativos con respecto al uso de las NTICs en la Tecnicatura Superior Analista Programador del Instituto Superior de Curuzú Cuatiá. Los objetivos específicos fueron: determinar cómo

influyen las siguientes variables en el uso académico de las NTICs y en el rendimiento académico de los alumnos: a) el nivel educacional de los padres, b) el nivel socio-económico, c) la posesión de una PC, d) el ámbito en el cual los alumnos acceden a las NTICs, e) la actitud general hacia el estudio.

Como consecuencia de lo antes indicado, este artículo se ha estructurado de la siguiente manera: en la Sección 2 se mencionarán los antecedentes considerados más relevantes, describiendo el marco teórico que sustenta este trabajo, en la Sección 3 se expondrá la metodología utilizada, en la Sección 4 se presentarán los principales resultados obtenidos, finalizándose con las conclusiones, las líneas de trabajo futuras, los agradecimientos y las referencias bibliográficas.

Antecedentes y Marco teórico

En (Sancho Gil, 2004) se estudia la evolución del uso de las NTICs en la mejora de la calidad educativa.

En (Carrasco Pradas et al., 2005) se estudia la incorporación de las NTICs a la metodología docente universitaria. Se afirma que la generalización en el uso de las NTICs no garantiza por sí sola la consecución de los objetivos perseguidos y es, por tanto, una condición necesaria pero no suficiente.

En (Acosta & La Red Martínez, 2012) se relata la experiencia realizada por los autores en la utilización de un Aula Virtual no convencional de Álgebra en la FaCENA – UNNE, donde se concluye que el enfoque utilizando b-learning y multimedia, ha sido exitoso.

En (La Red Martínez & Acosta, 2012) se muestran la problemática de integrar la perspectiva tecnológica con la pedagógica, brindando una arquitectura para sistemas de b-learning. Didácticamente, se adoptan principios educativos comprobados del enfoque centrado en la persona (person-centered approach) para impulsar procesos educativos, con utilización de las NTICs de manera segura. Técnicamente, se propone un entorno de trabajo (framework) en capas capaz de brindar soporte basado en Web para estos principios educativos.

En (La Red Martínez et al., 2011) se ha podido comprobar las grandes ventajas de la utilización de las tecnologías y software de última generación que soportan sistemas multiplataforma, se estudia el rendimiento académico con técnicas de DW y DM, considerándose la importancia concedida al estudio y su influencia en el rendimiento académico.

Según (Joyanes Aguilar, 1997), la nueva sociedad de la información o cibersociedad plantea un gran número de interrogantes de orden técnico, económico, sociológico, cultural y político.

Uno de los interrogantes es si los sistemas educativos serán capaces de producir la cantidad y calidad de egresados necesarios para soportar las demandas de personal altamente capacitado de esta sociedad de la información y el conocimiento (SIC) en las diferentes áreas, especialmente en las relacionadas con las NTICs. Es acá donde aparece el problema del rendimiento o desempeño académico. En (Forteza, 1975) se define el rendimiento académico como la productividad del sujeto, matizado por sus actividades, rasgos y la percepción más o menos correcta de los cometidos asignados. En (Marreno Hernández & Orlando Espino, 1988) se analiza el poder predictivo de las distintas aptitudes, mediante regresión múltiple, concluyendo que la más importante predictora del rendimiento académico es la verbal, seguida de la aptitud numérica y del razonamiento.

En un estudio acerca del rendimiento académico en el primer curso universitario (García & San Segundo, 2001) se utilizan indicadores como las tasas de graduación, diferenciando por tipos de centros y analizando el rendimiento académico a partir de datos individuales.

En (Herrera Clavero et al., 2004) se ha considerado que desde las primeras investigaciones sobre el aprendizaje los estudios se centraron exclusivamente en los aspectos cognitivos; luego los investigadores des-

cubrieron la importancia de los componentes afectivos y su influencia decisiva en el aprendizaje; finalmente se conjugaron los aspectos cognitivos y los afectivos, naciendo así el constructo llamado aprendizaje autorregulado (self-regulated learning).

También se ha estudiado el rendimiento académico universitario (Di Gresia, 2007), aplicando el enfoque de función de producción para estimar los determinantes del rendimiento académico.

En (Delfino, 1989) se han analizado los determinantes del aprendizaje mediante un enfoque de función de producción sugiriendo que los rendimientos escolares dependen de factores genéticos y socioeconómicos, de la calidad del docente, de las condiciones de la escuela y del grupo de alumnos (peer effect).

Los resultados publicados en (Maradona & Calderón, 2004) han mostrado que el factor más significativamente relacionado con la calidad educativa es el propio alumno como co-productor, medido a través del nivel socioeconómico del hogar de donde proviene.

En (Porto & Di Gresia, 2000) se ha mostrado que la productividad del estudiante es mayor para las mujeres, para los estudiantes de menor edad y para quienes provienen de hogares con padres más educados.

En (Fazio, 2004) se ha analizado detalladamente la vinculación entre horas trabajadas y rendimiento académico.

En general los estudios empíricos confirman la correlación entre mayores niveles de educación y atributos positivos luego de los estudios (McMahon, 2002).

En California (USA), el Academic Performance Index Reports incluye aspectos relacionados al rendimiento académico (academic performance) (California Department of Education, 2010).

En (García Jiménez et al., 2000) se ha estudiado la capacidad de la regresión lineal y de la regresión logística en la predicción del rendimiento y del éxito/fracaso académico, partiendo de variables como la asistencia y la participación en clase.

En (Marcelo García, 1987) se ha demostrado que las variables planificación del estudio, inteligencia, apoyo del profesor, estudio, tiempo, condiciones ambientales de estudio e implicación formaban parte de la ecuación de predicción de regresión múltiple, explicando un 25,70% de la varianza del rendimiento escolar en cursos de bachillerato.

El problema de encontrar buenos predictores del rendimiento futuro de manera que se reduzca el fracaso académico en los programas de postgrado ha recibido una especial atención en EE.UU. (Wilson & Hardgrave, 1995), habiéndose encontrado que las técnicas de clasificación como el análisis discriminante o la regresión logística son más adecuadas que la regresión lineal múltiple a la hora de predecir el éxito/fracaso académico.

Además de las herramientas tradicionales antes señaladas utilizadas para el estudio del rendimiento académico, existen otras provenientes de la Inteligencia de Negocios (BI: Business Intelligence), tales como los Almacenes de Datos (Data Warehouses: DW) y la Minería de Datos (Data Mining: DM), utilizada para el descubrimiento de conocimiento oculto en grandes volúmenes de datos.

Un DW es una colección de datos orientado a temas, integrado, no volátil, de tiempo variante, que se usa para el soporte del proceso de toma de decisiones gerenciales (Inmon, 1992), (Inmon, 1996), (Simon, 1997).

La DM es la etapa de descubrimiento en el proceso de KDD (Knowledge Discovery from Databases), es el paso consistente en el uso de algoritmos concretos que generan una enumeración de patrones a partir de los datos pre-procesados (Fayyad et al., 2001), (Hand et al., 2000).

La DM está muy ligada a los DW ya que los mismos proporcionan la información histórica con la cual los algoritmos de minería obtienen la información necesaria para la toma de decisiones (IBM Software Group, 2003).

La DM es un conjunto de técnica de análisis de datos que permiten extraer patrones, tendencias y regularidades para describir y comprender mejor los datos y extraer patrones y tendencias para predecir comportamientos futuros (Simon, 1997), (Berson & Smith, 1997), (White, 2001).

La DM genera modelos que pueden ser descriptivos o predictivos (Agrawal & Shafer, 1996); sus técnicas son diversas, una de las más utilizadas es la de clustering (o agrupamiento de datos) (Grabmeier & Rudolph, 1998), (Ballard et al., 2007). El cluster demográfico es un algoritmo desarrollado por IBM, que resuelve automáticamente los problemas de definición de métricas de distancia / similitud, proporcionando criterios para definir una segmentación óptima.

2. Metodología

Se ha utilizado un enfoque desde la lógica cuantitativa, trabajando con medición de variables, producción de hipótesis y utilización de técnicas de minería de datos inteligente, a los efectos de la extracción de conocimiento oculto en los datos.

Se buscó cumplir con los objetivos mencionados anteriormente trabajando con la hipótesis ya señalada en la sección 1.

El universo estuvo constituido por los alumnos en condiciones de cursar la asignatura Sistemas Operativos de la Carrera TSAP del ISCC.

La unidad de análisis estuvo integrada por cada alumno en condiciones de cursar la asignatura Sistemas Operativos. Los casos seleccionados fueron alumnos en condiciones de cursar dicha asignatura (unos 200 alumnos).

Se analizaron los datos cuantitativos que se obtuvieron (integrados en un DW) con las herramientas de DM, con el propósito de investigar relaciones entre las variables con métodos no tradicionales.

Metodología de Definición del DW Utilizado

Es importante recordar que un DW no se puede adquirir, se tiene que construir siguiendo determinada metodología.

La técnica a utilizar en la creación del DW depende de hacia quién se enfoca como punto principal el desarrollo del mismo, puede ser hacia el manejo de datos, de metas o de usuarios (Gutting, 1994). Los modelos propuestos son: "Data-Driven", "Goal-Driven" y "User-Driven". A continuación se describe en forma general en qué consiste cada uno.

Data-Driven: Este modelo considera que en un DW lo que se manejan son datos, a diferencia de los sistemas clásicos, en los que se manejan requerimientos, los cuales son el último aspecto a ser considerado en la toma de decisiones, considerando las necesidades de los usuarios en segundo término (Poe, 1996). El modelo de datos consiste de pocas dimensiones y de grupos de hechos. La dimensión representa la estructura básica del diseño. Los hechos son basados en el tiempo y tienen poco nivel de granularidad.

Goal Driven: Este modelo considera que el proceso de desarrollo gira en torno a los objetivos y metas

establecidas en principio. Al contrario del modelo anterior, este contiene más dimensiones y pocos hechos, los cuales son basados en el tiempo y tienen un bajo nivel de granularidad.

User Driven: Considera que el factor principal a tener en cuenta son las necesidades de los usuarios, pues son quienes utilizarán finalmente el sistema. El modelo consta de pocos hechos, los cuales tienen un nivel moderado de granularidad.

Independientemente de los modelos de desarrollo mencionados, las metodologías a seguir para el desarrollo del DW dependen en gran parte del tamaño del DW a crear y de la prontitud con que se requiera el DW.

A continuación se hace una descripción general de las dos principales metodologías para el desarrollo de un DW, la “Big Bang” y la “Rapid Warehousing”.

Big Bang: Esta metodología trata de resolver todos los problemas conocidos para crear un DW de gran tamaño, antes de liberarlo para su evaluación y prueba (Harinarayan et al., 1996).

Rapid Warehousing: Esta es también conocida como metodología evolutiva o incremental y considera que la construcción e implantación de un DW es un proceso evolutivo, el cual consiste en crear rápidamente una parte de un DW con la integración de data marts (Widom, 1995).

En este trabajo se ha seguido el modelo User Driven y la metodología Big Bang.

Descripción de la Estructura del DW Utilizado

A continuación se describirá el esquema básico resumido del DW con su tabla de hechos y sus tablas de dimensiones (Figura 1), (Figura 2), para luego continuar con la descripción detallada de cada una de las tablas mencionadas.

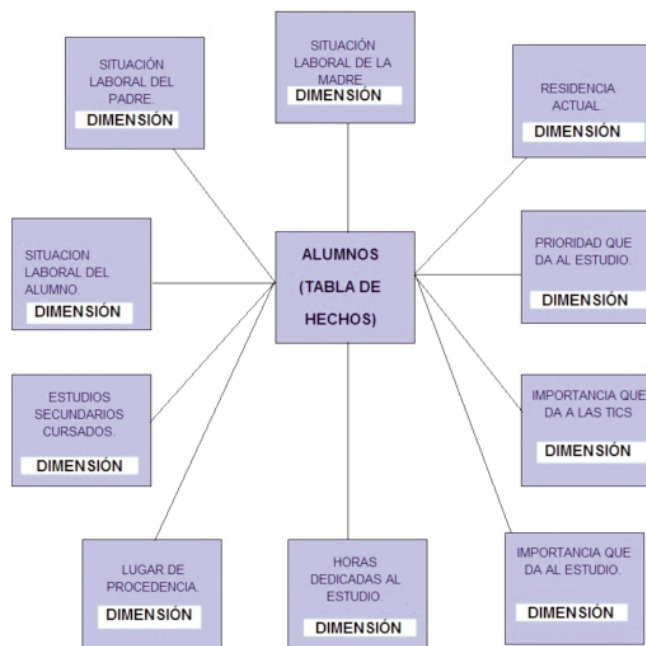


Figura 1. Esquema básico del almacén de datos (fuente propia).

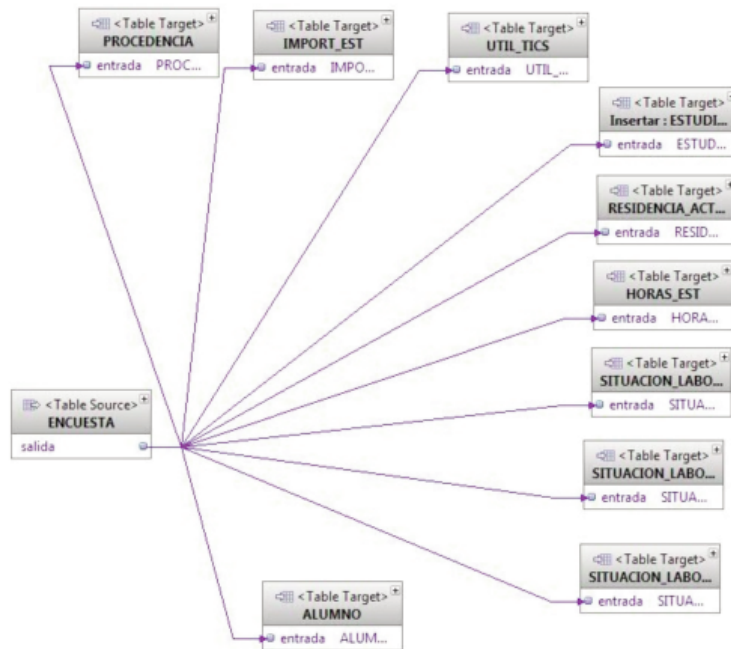


Figura 2. Fuente y almacén de datos (fuente propia).

En la Tabla 1 se puede observar las variables más significativas de la tabla de hechos.

En la Tabla 2 se observan las variables que conforman la dimensión Importancia Otorgada al Estudio.

En la Tabla 3 se puede observar las variables que conforman la dimensión de la Procedencia del Alumno

En la Tabla 4 se puede observar las variables que conforman la dimensión de la Utilización de las NTICs en la consideración del alumno.

En la Tabla 5 se puede observar las variables que conforman la dimensión Estudios Secundarios del Alumno.

En la Tabla 6 se observa las variables que conforman la dimensión Residencia Actual del Alumno.

En la Tabla 7 se observa las variables que conforman la dimensión Horas Dedicadas al Estudio en la apreciación del alumno.

En la Tabla 8 se describen las variables que conforman la dimensión Situación Laboral de la Madre del Alumno.

En la Tabla 9 se puede observar las variables que conforman la dimensión Situación Laboral del Alumno.

En la Tabla 10 se puede observar las variables que conforman la dimensión Situación Laboral del Padre del alumno.

El estudio se realizó sobre datos obtenidos mediante encuestas realizadas al alumnado, considerando además los resultados de las distintas instancias de evaluación previstas durante el cursado de Sistemas Operativos.

Nombre de Variable	Significado
LU_ALUM	Numero de libreta del alumno.
DNI_ALUM	Documento Nacional de Identidad.
CARRERA_ALUM	Carrera.
SEXO_ALUM	Sexo (Género).
EDAD	Edad.
ESTCIVIL_ALUM	Estado civil.
FECHA_NACIM	Fecha de nacimiento.
PAIS_NACIM	País.
PROV_NACIM	Provincia.
CIUDAD_NACIM	Ciudad.
FECHA_ENCUESTA	Fecha de la encuesta.
GRUPO_SANGUINEO	Grupo sanguíneo.
NOTA_PP_ALUM	Primer Parcial.
NOTA_PR_ALUM	Primer Recuperatorio.
NOTA_SP_ALUM	Segundo Parcial.
NOTA_SR_ALUM	Segundo Recuperatorio.
NOTA_EXT_ALUM	Extraordinario.
SF	Situación Final del alumno luego del cursado.
ANIO	Año lectivo.

Tabla 1. Variables y significados de la tabla de hechos.

Nombre de Variable	Significado
LU_IES	Libreta del Estudiante.
DNI_IES	Documento del Estudiante.
IMPORTANCIA_IES	Importancia que le da al estudio.

Tabla 2. Variables y significados de la dimensión importancia otorgada al estudio.

Nombre de Variable	Significado
LU_P	Libreta del Estudiante.
DNI_P	Documento del Estudiante.
PROVINCIA_P	Provincia de residencia.
CIUDAD_P	Ciudad de residencia.

Tabla 3. Variables y significados de la dimensión procedencia del alumno.

Nombre de Variable	Significado
LU_TIC	Libreta del Estudiante.
DNI_TIC	Documento del Estudiante.
UTIL_TIC	Provincia de residencia.

Tabla 4. Dimensión utilización de las NTICs.

Nombre de Variable	Significado
LU_ES	Libreta del Estudiante.
DNI_ES	Documento del Estudiante.
COLEGIO_ES	Nombre del Colegio.
DEPENDENCIA_ES	Dependencia del Colegio.
PROVINCIA_ES	Provincia a la que pertenece el Colegio.
CIUDAD_ES	Ciudad a la que pertenece el Colegio.
TITULO_ES	Título otorgado por el Colegio.
FEGRESO_ES	Fecha de egreso del alumno.

Tabla 5. Variables y significados de la dimensión estudios secundarios.

Nombre de Variable	Significado
LU_RA	Libreta del Estudiante.
DNI_RA	Documento del Estudiante.
TIPO_RA	Tipo de residencia del Estudiante.
CALLE_RA	Dirección de la residencia actual.
PROVINCIA_ES	Provincia de residencia.
CIUDAD_ES	Ciudad de residencia.

Tabla 6. Variables y significados de la dimensión residencia actual.

Nombre de Variable	Significado
LU_ES	Libreta del Estudiante.
DNI_ES	Documento del Estudiante.
HORAS_ES	Horas dedicadas al estudio.

Tabla 7. Variables y significados de la dimensión horas dedicadas al estudio.

Nombre de Variable	Significado
LU_SLM	Libreta del Estudiante.
DNI_SLM	Documento del Estudiante.
ULT_EST_SLM	Grado de escolaridad de la madre.
SIT_LAB_SLM	Situación Laboral de la madre.
HS_TRAB_SLM	Horas Semanales Trabajadas.
REL_EST_SLM	Rama de la Actividad Laboral de la madre.
CATOCUP_SLM	Categoría Ocupacional de la madre.

Tabla 8. Variables y significados de la dimensión situación laboral de la madre.

Nombre de Variable	Significado
LU_SL	Libreta del Estudiante.
DNI_SL	Documento del Estudiante.
TIPO_SL	Categoría Ocupacional.
ACTIV_ECONOM_SL	Rama Actividad Económica.
HS_TRAB_SL	Horas Semanales Trabajadas.
REL_EST_SL	Relación del Trabajo con la carrera elegida.
OBRA_SOCIAL_SL	Tiene Obra Social.
CATOCUP_SL	Categoría ocupacional.
SIT_LAB_SL	Situación Laboral

Tabla 9. Variables y significados de la dimensión situación laboral del alumno.

Nombre de Variable	Significado
LU_SLP	Libreta del Estudiante.
DNI_SLP	Documento del Estudiante.
ULT_EST_SLP	Grado de escolaridad del padre.
SIT_LAB_SLP	Situación Laboral del padre.
HS TRAB_SLP	Horas Semanales Trabajadas.
REL_EST_SLP	Rama de la Actividad Laboral de Padre.
CATOCUP_SLP	Categoría Ocupacional del Padre.

Tabla 10. Variables y significados de la dimensión situación laboral del padre.

Metodología de DM Utilizada

Actualmente existen varias metodologías de DM; las más utilizadas son la SEMMA y la CRISP-DM.

La metodología SEMMA fue desarrollada por el SAS Institute para descubrir patrones de negocio desconocidos. El nombre refiere a las cinco fases básicas del proceso (Figura 3) (Matignon, 2009), (SAS Institute, 2009).



Figura 3. Fases de la metodología SEMMA (SAS Institute, 2009).

La metodología CRISP-DM (Cross-Industry Standard Process for Data Mining: CRISP-DM), se organiza en seis etapas, cada una de ellas a su vez se divide en varias tareas (Figura 4) (Chapman et al., 1999).

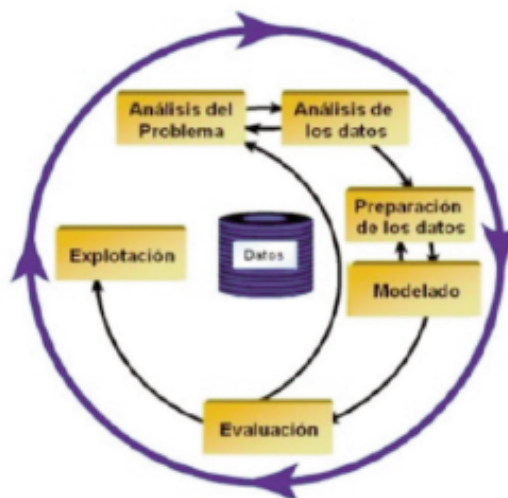


Figura 4. Fases del proceso de modelado de la metodología CRISP-DM (Chapman et al., 1999).

En este trabajo se utilizó la metodología CRISP-DM.

3. Resultados

Seguidamente se mostrarán los principales resultados obtenidos con las distintas técnicas de DM utilizadas: clusterización (segmentación), generadores de asociación (reglas de asociación) y árboles de decisión (predicción clasificación).

Resultados con Clusterización

Influencia del sexo (género) en el uso de las NTICs por parte de los alumnos y su rendimiento académico

El grupo más grande contiene el 16% de la población total. El grupo más pequeño contiene el 4,17% de la población total. La calidad global del modelo, medida de homogeneidad de los clusters, es de 0,749, lo que indica que, en promedio, las tuplas en un mismo cluster tienen en un 74,9% el mismo valor en los atributos activos.

Uno de los clusters, que representa el 16% de la población total, tiene predominantemente alumnos con las siguientes características: masculinos, situación final de 6, nota con la cual aprueban la asignatura, estado civil soltero, su ciudad de origen es Curuzú Cuatiá (84%), su provincia de origen es Corrientes (96%), para la población masculina (predominante en dicho cluster) las NTICs facilitan el proceso de enseñanza de la asignatura en un 58%, mientras que un 27% visualiza la importancia de la misma en su aplicación al campo profesional.

Otro cluster, con el 11,46% de la población, tiene totalmente alumnos femeninos, el 21% han logrado una situación final de 7, 8 y 9; en este grupo se puede observar que si bien no tienen la regularidad del 6, nota común en la población masculina, las mujeres han obtenido mejores calificaciones, el estado civil es soltero en todos los casos, la ciudad de origen es Curuzú Cuatiá para el 86%, la provincia de origen es 100% Corrientes, un 27% opinan que las NTICs son una realidad, mientras que un 64% opina que la importancia de las mismas radica en su aplicación al campo profesional.

Influencia del nivel educacional de los padres en el uso de las NTICs por parte de los alumnos

El grupo más grande contiene el 31% de la población total. El grupo más pequeño contiene el 3,84% de la población total.

Un cluster, correspondiente al 31% de la población total, indica que el 23% de los padres de los alumnos tienen escuela primaria completa, en tanto que el 14% tiene escuela secundaria completa; respecto del grado de utilización de las NTICs por parte de los alumnos, el 56% define el uso de las mismas como facilitadoras del proceso de enseñanza, el 28% considera que serán imprescindibles en el ejercicio profesional, lo que permite aseverar a priori un alto grado de aceptación en relación al uso de estas tecnologías (84%).

Otro cluster, correspondiente al 13% de la población total, muestra que el grado de escolarización de los padres es del 100% de escolaridad secundaria completa; respecto del grado de utilización de las NTICs por parte de los alumnos, se puede observar una respuesta contundente en relación a la importancia que el alumno asigna a la utilización de estas herramientas (98%), vinculándolas fundamentalmente a su proceso de formación académica.

En otro grupo, correspondiente al 11,39% de la población total, se observa que el 95% de los padres de los alumnos tienen escuela primaria completa, en tanto que el 3% tienen estudios universitarios completos y un 2% estudios superiores no universitarios completos; el 59% de los alumnos opinan que las NTICs facilitan el proceso de enseñanza, en tanto que el 26% expresa que serán imprescindibles para el ejercicio profesional.

Considerando lo precedentemente indicado, se puede extraer a modo de comentario que a medida que mejora el grado de escolaridad de los padres, esto influye sin ninguna duda en la opinión que tiene el alumno con respecto a la utilización de las NTICs.

Influencia del tipo de formación obtenido en la escuela secundaria en el uso de las NTICs por parte de los alumnos

El grupo más grande contiene el 38% de la población total. El grupo más pequeño contiene el 3,36% de la población total.

En el cluster correspondiente al 38% de la población total, se observa que el perfil de titulación predominante es el relacionado con la gestión administrativa de las organizaciones (35%); respecto a de la opinión que tiene el alumno con relación a la utilización de las NTICs, se puede observar que el 100% definen a estas herramientas como facilitadoras del proceso de enseñanza; a priori se puede afirmar que sí influye el tipo de titulación obtenida por el alumno al finalizar sus estudios secundarios, dado que el alumno cuyo perfil de titulación está orientado a la gestión administrativa de las empresas, tiene una mejor opinión con relación a la utilización de estas tecnologías.

Influencia del hecho de que los alumnos trabajen además de estudiar en el uso de las NTICs

El grupo más grande contiene el 18,61% de la población total. El grupo más pequeño contiene el 5,12% de la población total.

El cluster correspondiente al 18,61% de la población total, respecto de la situación laboral del alumno, muestra que el 100% de esa población no trabaja; respecto de la utilización de las NTICs, muestra que el 100% de esa población coinciden en que facilitan el proceso de enseñanza.

En otro cluster, correspondiente al 8,54% de la población total, respecto de la cantidad de horas trabajadas por el alumno en la semana, se puede observar que el 100% de dicha población se desempeña en tareas que insumen un promedio de más de 5 horas reloj por día; referido a la situación de la utilización de las NTICs por parte de los alumnos, se puede decir que si bien la importancia asignada al uso de estas herramientas en cuanto a su utilización no indica claramente que existe una influencia en cuanto al alumno que trabaja y el que no lo hace, sin embargo se puede señalar que hay una opinión más concreta en aquel alumno que trabaja y estudia, en base al hecho de que el alumno que trabaja y estudia, manifiesta además su interés por la utilización de estas herramientas en el campo profesional.

Influencia de la actitud general hacia el estudio en el uso de las NTICs por parte de los alumnos

El grupo mayor contiene el 19,72% de la población total. El grupo menor contiene el 5,45% de la población total.

El cluster correspondiente al 19,72% de la población total, tiene predominantemente alumnos que dedican más de 10 y hasta 20 horas inclusive al estudio, que además con relación a la utilización de las NTICs opinan que facilitan el proceso de enseñanza y aprendizaje y que la importancia que asignan al estudio es más que a la diversión; respecto de la cantidad de horas dedicadas al estudio por el alumno, se puede observar que el 100% de dicha población manifiesta una dedicación entre 10 y 20 horas; respecto de la importancia que el alumno asigna al estudio, se puede observar que el 100% de dicha población manifiesta darle una importancia mayor que a la diversión; respecto de la utilización de las NTICs por parte del alumno, se puede observar que el 100% de dicha población manifiesta que facilitan el proceso de enseñanza.

Otro cluster, correspondiente al 10,14% de la población total, respecto de la cantidad de horas semanales dedicadas al estudio por el alumno, se puede observar que el 100% de dicha población manifiesta una dedi-

cación entre 10 y 20 horas; respecto de la importancia que el alumno asigna al estudio, se puede observar que el 98% de dicha población manifiesta darle una importancia mayor que a la diversión, en tanto que el 1% más que a la familia; respecto de la utilización de las NTICs por parte del alumno, se puede observar que el 100% de dicha población manifiesta que serán imprescindibles para el ejercicio profesional.

Otro agrupamiento, correspondiente al 5,45% de la población total, respecto de la cantidad de horas semanales dedicadas al estudio por el alumno, se puede observar que el 88% de dicha población manifiesta una dedicación mayor a 20 horas, en tanto que un 2% hasta 10 horas inclusive; respecto de la importancia que el alumno asigna al estudio, el 77% de dicha población opina que es más importante que la diversión, por otra parte un 1% más que la familia y el 22% más que el trabajo; respecto de la utilización de las NTICs por parte del alumno, se observa que el 70% de dicha población opina que facilitan el proceso de enseñanza, por otra parte el 15% opina que serán imprescindible para el ejercicio profesional y un 11% opina que son una realidad en la actualidad.

Se puede observar que el grado de dedicación e importancia asignada por el alumno a sus estudios tiene una relación directa con la actitud del mismo en cuanto a la utilización de las NTICs.

Resultados con Generadores de Asociación

El objetivo de la función de la minería de asociaciones es encontrar los elementos que se asocian consistentemente con los demás de una manera significativa. Las relaciones descubiertas se expresan como reglas de asociación. La función de la minería de asociación encuentra asociaciones y también asigna probabilidades. La primera parte de una regla de asociación se llama el cuerpo de la regla y la segunda parte la cabeza de la regla.

Las reglas de asociación tienen los siguientes atributos: a) confianza: el valor de confianza representa la validez de la norma (una regla tiene el 70% de confianza si en el 70% de los casos en que el cuerpo de la regla está presente en un grupo, la cabeza de la regla también está presente en el grupo); b) soporte: el valor para el soporte se expresa como un porcentaje del número total de registros o transacciones; c) elevación: el valor de elevación indica hasta qué punto el valor de confianza es más alto de lo esperado; se define como el cociente del valor de confianza y el valor de soporte de la cabeza de la regla; el valor de soporte de la cabeza de regla puede ser considerado como el valor esperado para la confianza e indica la frecuencia relativa de la cabeza de la regla en todo el conjunto de transacciones.

Se obtuvieron 112 reglas, algunas de las cuales se indican a continuación.

Si el género del alumno es masculino, implica un estado civil soltero en un 91% de los casos.

Si el género del alumno es femenino, implica un estado civil soltero en un 85% de los casos.

Si la situación final del alumno es 6, lo cual ocurre en un 31%, implica un estado civil soltero en un 86% de los casos.

Si el género del alumno es femenino, implica que opinará que las NTICs facilitan el proceso de enseñanza en un 56% de los casos.

Si el alumno opina que la utilización de las NTICs será imprescindible para el ejercicio profesional, lo cual ocurre en un 25%, implica que su estado civil es soltero en un 88% de los casos.

Si el sexo (género) del alumno es masculino, implica que su situación final será de 6 en un 37,5% de los casos.

Si el sexo (género) del alumno es femenino, implica que su situación final será de 6 en un 35,44% de los

casos.

Si el alumno opina que la utilización de las NTICs será imprescindible para el ejercicio profesional, lo cual ocurre en un 14%, implica que el género del alumno es femenino en un 49% de los casos.

Si la opinión del alumno es que la utilización de las NTICs facilita el proceso de enseñanza y las horas dedicadas al estudio son hasta 10 horas inclusive, lo que ocurre en un 12,54%, implica que el género del alumno será masculino en el 50,31% de los casos.

Si el estado civil es soltero y el alumno opina que la utilización de las NTICs será imprescindible para el desempeño profesional, lo que ocurre en un 13%, implica que el género del alumno será masculino en el 52% de los casos.

Si la opinión del alumno es que la utilización de las NTICs facilita el proceso de enseñanza y las horas dedicadas al estudio son más de 10 y hasta 20 inclusive, lo que ocurre en un 13,43%, implica que el género del alumno será femenino en el 49,68% de los casos.

Si la opinión del alumno es que la utilización de las NTICs facilita el proceso de enseñanza y las horas dedicadas al estudio son más de 10 y hasta 20 inclusive, lo que ocurre en un 13,60%, implica que el género del alumno será masculino en el 50,31% de los casos.

Si el género del alumno es femenino y la situación final es 6, lo que ocurre en un 14,46%, implica que el estado civil del alumno será soltero en el 82% de los casos.

Si la situación final del alumno es 6 y las horas dedicadas al estudio son más de 10 y hasta 20 inclusive, lo que ocurre en un 15%, implica que el estado civil del alumno será soltero en el 86% de los casos.

Si la situación final es 6 y es masculino, lo que ocurre en un 17%, implica que el estado civil del alumno será soltero en el 90% de los casos.

Si es femenino y dedica hasta 10 horas al estudio inclusive, lo que ocurre en un 19%, implica que el estado civil del alumno será soltero en el 85% de los casos.

Si es soltero y las horas dedicadas al estudio son hasta 10 inclusive, lo que ocurre en un 22% de los casos, implica que la opinión sobre la utilización de las NTICs será que facilitan el proceso de enseñanza en el 56% de los casos.

Si el alumno opina que la utilización de las NTICs facilita el proceso de enseñanza y las horas dedicadas al estudio son más de 10 y hasta 20 inclusive, lo que ocurre en un 24%, implica que el estado civil del alumno será soltero en el 88% de los casos.

Si la utilización de las NTICs facilita el proceso de enseñanza y el género del alumno es masculino, lo que ocurre en un 25,63%, implica que el estado civil del alumno será soltero en el 91,25% de los casos.

Resultados con Árboles de Decisión

Los resultados obtenidos se han resumido y agrupado según calificación final (clase); se ha considerado alto rendimiento académico a las calificaciones finales entre 7 y 10, rendimiento académico medio a la calificación final de 6 y bajo rendimiento académico a las calificaciones finales entre 0 y 5; a modo de ejemplo, los resultados para alto rendimiento académico con nota de 7 se muestran en Tabla 11.

Clase	Atributo	Valor predominante
7 ^a 15,09% de la población	Situación final	7
	Primer parcial	6 - 8 (75%)
	Segundo parcial	6 - 8 (75%)
	Provincia	Corrientes (75%)
	Tipo de residencia	Con familiares (88%)
	Situación laboral alumno	No trabaja (92%)
	Horas semanales trabajadas	Hasta 20 inclusive (8%)
	Relación con la carrera elegida	Parcial (54%)
	Estudios cursados padre	Esc. Primaria Completa (33%) Esc. Secundaria Completa (25%)
	Situación laboral padre	Ocupado (75%)
	Estudios cursados madre	Esc. Primaria Completa (21%) Esc. Secundaria Completa (42%)
	Situación laboral madre	No trabaja (58%) Ocupado (33%)
	Género	Masculino (58%)
	Prioridad otorgada al estudio	Más que a la diversión (71%)
	N° horas semanales dedicadas al estudio	Hasta 10 horas (50%)
	Estudia para	Aprobar la asignatura (38%) Aprender a aprender (28%)
	Utilización de las NTICs	Facilitan el proceso de enseñanza (50%) Indispensables en el ejercicio profesional (21%)

Tabla 11. Alumnos de alto rendimiento académico: calificación final 7.

Por razones de espacio no se incluyen las tablas correspondientes a las demás calificaciones finales.

Los resultados resumidos con relación al perfil de los alumnos considerados de alto rendimiento académico, correspondiente al 25,78% de la población, son los siguientes: a) la mayoría vive con el grupo familiar, b) generalmente no trabajan, c) un grupo minoritario trabaja hasta 20 horas semanales, d) en la mayoría de los casos la relación del trabajo con la carrera elegida es parcial, e) el grado de escolaridad primaria y secundaria de los padres es relativamente bajo, registrándose casos de escolaridad terciaria o universitaria, f) mayoritariamente el porcentaje de ocupación de los padres es relativamente alto, g) en la mayoría de los casos el objetivo de los alumnos es estudiar para aprender a aprender o para aprender integralmente la materia, h) la mayoría considera la utilización de las NTICs asociadas al proceso de enseñanza-aprendizaje y como imprescindibles para el ejercicio profesional, i) la mayoría son solteros, registrándose un buen porcentaje de casados, j) la mayoría corresponden al género masculino, k) un grupo minoritario otorga al estudio más prioridad que al trabajo.

Los resultados resumidos con relación al perfil de los alumnos considerados de rendimiento académico medio, correspondiente al 36,44% de la población, son los siguientes: a) la mayoría vive con el grupo familiar, b) generalmente no trabajan, c) un grupo minoritario trabaja hasta 20 horas semanales, d) en la mayoría de los casos la relación del trabajo con la carrera elegida es parcial, e) el grado de escolaridad primaria y secundaria de los padres es relativamente bajo, no registrándose casos de escolaridad terciaria o universitaria, f) mayoritariamente el porcentaje de ocupación de los padres es relativamente bajo, g) en la mayoría de los casos el objetivo de los alumnos es estudiar para aprobar la materia, h) la mayoría considera la utilización de las NTICs asociadas al proceso de enseñanza-aprendizaje, i) la mayoría son solteros, registrándose un buen porcentaje de casados, j) la mayoría corresponden al género masculino.

Los resultados resumidos con relación al perfil de los alumnos considerados de bajo rendimiento académico, correspondiente al 37,73% de la población, son los siguientes: a) la mayoría vive con el grupo familiar, registrándose un grupo minoritario importante que vive en forma independiente concentrándose especialmente en la clase correspondiente a la calificación de 2, b) generalmente no trabajan, pero un grupo significativo sí lo hace, en esta categoría está la mayor cantidad de alumnos que trabajan, c) un grupo minoritario trabaja hasta 20 horas semanales y otro grupo menor más de 36 horas semanales, d) en la mayoría de los casos la relación del trabajo con la carrera elegida es parcial o no existe relación, e) el grado de escolaridad primaria y secundaria de los padres es relativamente bajo, registrándose casos de escolaridad terciaria o universitaria, f) may-

oritariamente el porcentaje de ocupación de los padres es relativamente alto, registrándose un grupo minoritario importante con un bajo porcentaje de ocupación, g) en la mayoría de los casos el objetivo de los alumnos es estudiar para aprobar la materia y un grupo minoritario lo hace para aprender a aprender o para aprender integralmente la materia, h) la mayoría considera la utilización de las TICs asociadas al proceso de enseñanza-aprendizaje y un grupo minoritario como imprescindibles para el ejercicio profesional, i) la mayoría son solteros, j) la mayoría corresponden al género femenino.

En la Tabla 12 se muestran algunas de las correlaciones que han sido seleccionadas, considerándolas relevantes para el análisis de los objetivos establecidos en esta investigación.

Hay correlaciones interesantes, como por ejemplo, la que muestra la incidencia de la nota del primer parcial en la situación final del alumno, así también la incidencia del tipo de residencia con relación a la situación final del alumno, el grado de escolaridad de los padres en relación a las horas dedicadas al estudio y situación final del alumno, la incidencia de la utilización de las NTICs en relación a la situación final del alumno.

4. Discusión

Conclusiones

Es de fundamental importancia conocer desde el inicio de las actividades académicas cuáles alumnos son candidatos al bajo rendimiento académico y qué factores inciden en el mismo, para encarar tempranamente acciones tendientes a revertir tal situación.

Campo	Campo	Correlación
Tipo de Residencia	Escolaridad del Padre	0,553
Tipo de Residencia	Situación final alumno	0,509
Escolaridad del padre	Importancia asignada al estudio	0,458
Escolaridad del padre	Estudia para	0,446
Escolaridad del padre	Situación final alumno	0,541
Situación laboral Padre	Estudia para	0,406
Escolaridad de la madre	Horas dedicadas al estudio	0,499
Escolaridad de la madre	Utilización de las NTICs	0,478
Situación laboral madre	Situación final alumno	0,484
Estudia para	Utilización de las NTICs	0,524
Utilización de las NTICs	Situación final alumno	0,505
Nota primer parcial	Situación final alumno	0,985

Tabla 12. Correlación e importancia de campo.

En esta investigación sólo se han abarcado algunos métodos de extracción del conocimiento a través de DM. No obstante, existen muchas más posibilidades que ofrecen ésta y otras herramientas.

Con las tres técnicas de DM seleccionadas se han obtenido muy buenos resultados, cumpliéndose los objetivos planteados y verificándose la hipótesis de trabajo. Han quedado evidenciadas las características de los perfiles representativos de alumnos de bajo, medio y alto rendimiento académico.

El modelo de Clasificación a través de Árboles de Decisión superó en calidad a los patrones obtenidos con el método de Generación de Clústeres.

Las técnicas de DM han permitido construir modelos predictivos, de asociación, de segmentación, basados en datos históricos almacenados en distintas fuentes; se considera adecuada la calidad de los modelos obtenidos. Ha sido posible determinar los perfiles de éxito y fracaso académico de los alumnos de S.O. de la TSAP del ISCC, lo que ha permitido definir líneas de acción tendientes a dar un mayor soporte a los alumnos detectados con perfil de riesgo de fracaso académico.

Futuras Líneas de Trabajo

A lo largo del desarrollo del presente trabajo han surgido varias líneas para ser abordadas en futuras investigaciones.

Entre algunas de ellas se pueden mencionar las siguientes: a) integrar los diferentes flujos de minería en flujos de control que permitan automatizar los procesos descritos en este trabajo; b) diseñar los hipercubos de datos incorporando nuevas variables socioeconómicas; c) implementar mecanismos académicos de seguimiento de las acciones que se realicen en base a la información suministrada por los procesos de minería, a los efectos de efectuar ajustes que se consideren pertinentes en cuanto a la ejecución de las acciones antes mencionada; d) aplicar el modelo desarrollado en este trabajo a otras asignaturas de la carrera TSAP del ISCC especialmente las del primer año en las que se registran los mayores porcentaje de fracaso académico.

Agradecimientos

Se ha utilizado el IBM Data Warehouse Edition (DWE) V.9.5, que incluyen al DB2 Enterprise Server Edition (DB2 ESE), al Design Studio (DS) y al Intelligent Miner (IM), que se han obtenido de la empresa IBM Argentina S.A. en el marco de la Iniciativa Académica (Acuerdo del 18/06/04 D, Res. N° 1417/04 D, Res. N° 858/06 CD).

Referencias

- Acosta, J.C., La Red Martínez, (2012). D.L. Un Aula Virtual no convencional de Algebra en la FaCENA – UNNE: un enfoque utilizando b-learning y multimedia. Editorial Académica Española – LAP LAMBERT Academic Publishing GmbH & Co. KG Heinrich – Bocking- Str. 6-8,66121 Saarbrücken, Alemania. ISBN 978-3-659-02034-6. Alemania.
- Agrawal, R.; Shafer, J. C. (1996). Parallel Mining of Association Rules. IEEE Transactions on Knowledge and Data Engineering. USA.
- Ballard, Ch.; Rollins, J.; Ramos, J.; Perkins, A.; Hale, R.; Dorneich, A.; Cas Milner, E. & Chodagam, J. (2007). Dynamic Warehousing: Data Mining Made Easy. IBM International Technical Support Organization. IBM Press. USA.
- Berson, A. & Smith, S. J. (1997). Data Warehouse, Data Mining & OLAP. Mc Graw Hill. USA.
- Bolaños Calvo, B. (2001). Las Nuevas Tecnologías y los Desafíos Teórico – Prácticos en los Sistemas de Educación a Distancia: Caso UNED de Costa Rica. Temática: Universidades Virtuales y Centros de Educación a Distancia. UNED. Costa Rica.
- California Department of Education. (2009). 10 Academic Performance Index Reports. USA.
- Carrasod Pradas, A., Gracia Expósito, E., de la Iglesia Villalobos, C. (2005). Las TIC en la construcción del espacio europeo de educación superior. Dos experiencias docentes en teoría económica. Revista Iberoamericana de Educación, 36, 1-16.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Renartz, T., Shearer, C., Wirth, R. (1999). CRISP-DM 1.0. Step-by-step data mining guide.
- Delfino, J. A. (1989). Los determinantes del aprendizaje. In Petrei, A. H., editor, Ensayos en economía de la educación. Educational Evaluation and Policy Analysis.
- Di Gresia, L. (2007). Rendimiento Académico Universitario. Tesis Doctoral. Universidad Nacional de La Plata. Argentina.
- Fayyad, U.M.; Grinstein, G. & Wierse, A. (2001). Information Visualization in Data Mining and Knowledge Discovery. Morgan Kaufmann. Harcourt Intl.
- Fazio, M. V. (2004). Incidencia de las horas trabajadas en el rendimiento académico de estudiantes universitarios argentinos. Documentos de Trabajo UNLP, 52. Argentina.
- Forteza, J. (1975). Modelo instrumental de las relaciones entre variables motivacionales y rendimiento. Revista de Psicología General y Aplicada, 132, 75-91. España.
- García, M. M.; San Segundo, M. J. (2001). El Rendimiento Académico en el Primer Curso Universitario. X Jornadas de la Asociación de Economía de la Educación. Libro de Actas, págs. 435-445. España.
- García Jiménez, M. V.; Alvarado Izquierdo, J. M.; Jiménez Blanco, (2000). A. La predicción del rendimiento académico: regresión lineal versus regresión logística. Psicothema Vol. 12, Supl. n° 2, pp. 248-252. España.
- Grabmeier, J. & Rudolph, A. (1998). Techniques of Cluster Algorithms in Data Mining version 2.0. IBM Deutschland Informationssysteme GmbH. GBIS (Global Business Intelligence Solutions). Germany.
- Gutting, R. (1994). An Introduction to spatial database systems. VLDB Journal, 3, 357- 399.
- Hand, D.J.; Mannila, H. & Smyth, P. (2000). Principles of Data Mining. The MIT Press. USA.
- Harinarayan V., Rajaraman, A., Ullman, J. (1996). Implementation data cubes efficiently. ACM SIGMOD Record, 25 (2), 205 - 216.
- Herrera Clavero, F. (2004). ¿Cómo Interactúan el Autoconcepto y el Rendimiento Académico en un Contexto Educativo Pluricultural?. Revista Iberoamericana de Educación. España.
- IBM Software Group. (2003). Enterprise Data Warehousing whit DB2: The 10 Terabyte TPC-H Benchmark. IBM Press. USA.

22. IEEE. (2012). Learning Technology Standards Committee. Retrieved Jan 6, 2012, from <http://www.ieeeltsc.org:8080/Plone>.
23. Inmon, W. H. (1992). *Data Warehouse Performance*. John Wiley & Sons. USA.
24. Inmon, W. H. (1996). *Building the Data Warehouse*. John Wiley & Sons. USA.
25. Joyanes Aguilar, L. *Cibersociedad*. (1997). España: Mc Graw Hill.
26. La Red Martínez, D.L., Acosta, J.C. (2012). B-Learning: Una propuesta de Arquitectura Segura Basada en Patrones. *Revista Internacional PEI: Por la Psicología y Educación Integral*, Año 2, 3, 58-95.
27. La Red Martínez, D.L., Acosta, J., Agostini, F., Uribe, V., Rambo, A. (2011). La importancia otorgada al estudio y su relación con el rendimiento académico. *Revista Documentación*. Año IV, 24, 54-62.
28. Maradona, G. & Calderón, M. I. (2004). Una aplicación del enfoque de la función de producción en educación. *Revista de Economía y Estadística*. Universidad Nacional de Córdoba, XLII. Argentina.
29. Marcelo García, C.; Villarín Martínez, M.; Bermejo Campos, B. (1987). Contextualización del rendimiento en bachillerato. *Revista de Educación*, 282, 267-283. España.
30. Marreno Hernández, H.; Orlando Espino, M. (1988). Evaluación comparativa del poder predictor de las aptitudes sobre notas escolares y pruebas objetivas. *Revista de Educación*, 287, 97-112. España.
31. Matignon, R. (2009). *Data Mining Using SAS Enterprise Miner*. U.S.A.: Wiley.
32. McMahon, W. W. (2002). *Education and Development*. Oxford University Press.
33. Peiró, J. M. (2001). *Las competencias en la sociedad de la información: nuevos modelos formativos*. España: Centro Virtual Cervantes.
34. Poe, V. (1996). *Building a Data Warehouse for Decision Support*. New Jersey: Prentice Hall.
35. Porto, A. & Di Gresia, L. (2000). Características y rendimiento de estudiantes universitarios. El caso de la Facultad de Ciencias Económicas de la Universidad Nacional de La Plata. *Documentos de Trabajo UNLP*, 24.
36. Sancho Gil, J. M. (2004). Las Observaciones de la Sociedad de la Información: Evaluación o Política de promoción de las TIC en Educación. *Revista Iberoamericana de Educación*, 36, 37-68.
37. SAS Institute, (2009). Disponible en: <http://www.sas.com/technologies/analytics/datamining/miner/semma.html>.
38. Simon, A. (1997). *Data Warehouse, Data Mining and OLAP*. John Wiley & Sons. USA.
39. Wallace, L. & Young, J. (2010). Implementing Blended Learning: Policy Implications for Universities, *Online Journal of Distance Learning Administration*, Volume XIII, Number IV, winter 2010 University of west Georgia, Distance Education Center.
40. White, C. J. (2001). *IBM Enterprise Analytics for the Intelligent e-Business*. IBM Press. USA.
41. Widom J. (1995) Research Problems in data warehousing. *Conf. Information and Knowledge Management*, Baltimore. U.S.A.
42. Wilson, R. L.; Hardgrave, B. C. (1995). Predicting graduate student success in an MBA program: Regression versus classification. *Educational and Psychological Measurement*, 55, 186-195. USA.

Cómo citar este artículo / How to cite this paper

La Red, D. L. & Podestá, C. E. (2014). Metodología de Estudio del Rendimiento Académico Mediante la Minería de Datos. *Campus virtuales*, 3(1), 56-73.

La Red, D. L. & Podestá, C. E. (2014). Methodology Study of Academic Performance Using Data Mining. *Campus virtuales*, 3(1), 56-73.