

PROCEDIMIENTO PARA IDENTIFICAR FACTORES DE RIESGO EN LOS DATOS EDUCACIONALES DEL PRIMER AÑO DE LA UCI.

Hugo Arnaldo Martínez Noriega, MSc.

*Máster en Matemáticas, mención estadística (Cuba).
Docente de la Universidad de las Ciencias Informáticas (UCI), Cuba.
hugomn@uci.cu*

Bolívar Ernesto Medrano Broche, MSc.

*Máster en Tecnologías de Apoyo a la Toma de Decisiones (Cuba).
Docente de la Facultad de Sistemas Computacionales y
Telecomunicaciones de la Universidad Tecnológica ECOTEC, Ecuador.
boly07@gmail.com*

Gladis Cabanas Gómez, MSc.

*Máster en Matemáticas, mención estadística (Cuba).
Docente del Departamento de Matemática Aplicada
de la Universidad de La Habana (UH), Cuba.
gladis@matcom.uh.cu*

ARTÍCULO DE INVESTIGACIÓN

Recibido: 9 de marzo de 2017.

Aceptado: 15 de mayo de 2017.

RESUMEN

El propósito principal de este artículo es mostrar la efectividad de las técnicas de aprendizaje computarizado supervisado con los datos que se recogen para conformar la caracterización individual del estudiante en el primer año de la UCI. Los datos que se acopian para el diseño de la caracterización individual son: los diagnósticos pedagógicos que se aplican a los estudiantes a su entrada al centro, los datos generales de la ficha de matrícula del estudiante y el primer corte evaluativo que se emite de cada estudiante; a partir de la caracterización individual se tiene una percepción de las aptitudes cognoscitivas y características personales

de cada alumno. Sin embargo, no hay una utilización adecuada de estos datos. En este estudio, se reflejan las relaciones entre estos datos con los resultados de promoción limpia cuando se culmina el primer año de la carrera. Por otra parte, otro de los objetivos del estudio es detectar los factores más influyentes en la promoción final, debido a que la detección de estos será de vital importancia para la toma de decisiones por parte de los directivos docentes. En la modelación se emplearon: los árboles de decisión y la regresión logística binaria.

Palabras clave: análisis multivariado, descubrimiento de conocimiento, factores de riesgo, caracterización individual, medida de eficiencia.

ABSTRACT

The main purpose of this article is to show the effectiveness of computerized supervised learning techniques with the data collected to form the individual characterization of the student in the first year of the UCI. The data that are collected for the design of the individual characterization are: the pedagogical diagnoses that are applied to students at their entrance to the center, the general data of the student's enrollment form and the first evaluation cut that is issued from each student; from the individual characterization we have a perception of the cognitive skills and personal characteristics of each student. However, there is no adequate use of this data. In this study, the relationships between these data are reflected with clean promotion results when the first year of the race ends. On the other hand, another of the objectives of the study is to detect the most influential factors in the final promotion, since the detection of these will be of vital importance for the decision making by the teaching directors. In the modeling were used: decision trees and binary logistic regression.

Keywords: multivariate analysis, knowledge of discovery, risk of factors, individual characterization, efficiency measure.

INTRODUCCIÓN

El vertiginoso desarrollo de las nuevas tecnologías de la informática y las telecomunicaciones requieren un perfeccionamiento continuo del sistema educacional. Además, en la actualidad la necesidad de elevar la eficiencia en la educación superior es de gran importancia para el avance de nuestra sociedad. El mundo actual se enfrenta a condiciones que lo hacen cada vez más competitivo, es decir, es necesario conocer cuáles son las características de nuestros estudiantes, que deficiencias presentan, como se agrupan los

mismos y qué estrategias trazar para limar las anteriores deficiencias. Por consiguiente, para lograr ingenieros de alto nivel, es esencial prepararlos en la modelación matemática asociada a los procesos que analizan en sí; por lo tanto, significa que estas personas deben ser capaces de determinar lo que tienen que hacer, volver a instruirse, volver a aprender en cómo hacer las nuevas tareas, a través de algoritmos y procedimientos eficaces.

En la Universidad de las Ciencias Informáticas (UCI) se necesita contar con métodos para apoyar el proceso de toma de decisiones en el proceso de formación profesional. La UCI cuenta con 7 facultades, en todas ellas se conforma una caracterización individual de cada estudiante cuando arriban a la universidad. La caracterización individual se basa en los siguientes datos: datos de la ficha de matrícula, resultados de los diagnósticos cognoscitivos (que se le aplican a cada estudiante al arribar a la universidad) y el primer corte evaluativo. Sin embargo, la principal deficiencia de este enfoque desde el punto de vista práctico, está condicionado al hecho, que la universidad presenta grandes matrículas y diseñar para cada estudiante una estrategia es realmente complejo. Además, la caracterización individual no permite identificar a los factores de riesgo que inciden en los resultados de promoción limpia al finalizar el semestre.

Es por ello que en este sentido las técnicas de descubrimiento de conocimiento que se emplean en el presente trabajo, tienen como objetivo determinar los posibles factores de riesgo que influyen en los resultados finales de promoción limpia al finalizar el semestre.

REVISIÓN TEÓRICA

La primera etapa de este trabajo se centrará en la realización de un análisis de los trabajos de descubrimiento de conocimiento en los datos, para el apoyo a la toma de decisiones docentes-pedagógicas; como ejemplos de estos tipos de trabajos se pueden citar los siguientes:

- La evaluación del rendimiento académico de los estudiantes en varias asignaturas en un determinado período, ver (Garnica, 1997), (Zúñica, 2005) y (Moral de la Rubia, 2006).
- Evaluación del rendimiento académico en una asignatura, a partir del rendimiento de otras, ver (Rodríguez, 2009).
- Evaluación del rendimiento académico, a partir de su relación con variables demográficas y resultados docentes de la enseñanza

previa a la universidad, ver (Reyes, 2007), (García, 2012), (Gallander, 2011).

- Determinar perfiles de estudiantes como base para el establecimiento de estrategias educativas diferenciadas (Delavari, 2008).
- Predicción del promedio de cada año académico en función de las características y los resultados docentes previos de los educandos, ver (Espinosa, 2007) y (Brito, 2008).
- Predicción del éxito o fracaso académico de los estudiantes al finalizar el primer año (King, 2008) y (Remón, 2009).

En las universidades, el desarrollo de trabajos de descubrimiento de conocimiento en los datos ha sido posible por la presencia de dos factores importantes: la existencia de trabajadores con una amplia experiencia en el análisis de datos y el desarrollo de los sistemas de información universitarios (Rico, 2011). Los análisis a priori o posteriori mostrados anteriormente han contribuido a desarrollar indicadores de eficiencia universitaria (García, 2012). Estos análisis se basan usualmente en estudios estadísticos transversales o longitudinales, con el objetivo de descubrir conocimiento en la información disponible y por otra parte, sirven de soporte para la toma de decisiones en el ámbito universitario.

Cuando se decide realizar un trabajo de descubrimiento en los datos se tienen en cuenta tres elementos fundamentales que se interrelacionan: los objetivos, el tipo de datos que se pretende procesar y el grado de conocimiento sobre el tema que tengan los investigadores para obtener e interpretar el modelo (Rodríguez, 2009).

Por ejemplo, García pretende, predecir la medida de eficiencia o función objetivo: “Se gradúa en tiempo” (Sí/No), la cual como se evidencia es una variable dicotómica. Para ello emplea como variables predictivas datos de los estudiantes individuales previos al comienzo de los estudios universitarios, la facultad y la carrera donde están matriculados ahora. Observe que se utilizan solamente datos predictores “epidemiológicos” y no “clínicos”, como podría ser el desempeño de esos estudiantes durante el primer y segundo año de la carrera.

En la Tabla 1 se muestran representaciones de los formatos que recogen los resultados de evaluaciones finales y de un corte evaluativo generados por la

aplicación del sistema informático utilizado en la Universidad para apoyar el proceso docente (Akademos), (Gestor de información de la UCI).

Tabla 1: Resultados de las evaluaciones finales de asignaturas.

Nº	Nombre y Apellidos	EF.I.		IP.		M.I.		MD.I.		PP.I.		SN.	
		A	NF	A	NF	A	NF	A	NF	A	NF	A	NF
1	Nombre1		4	1	3	2	4	1	5		4		4
2	Nombre2		5		5				5		5		5
3	Nombre3		4		3	1	2		2		4		3
...
N1	NombreN1		5		3	1	2		5		5		5

Fuente: elaboración propia.

Tabla 2: Corte evaluativo.

No.	Nombre y Apellidos	AE.		GC.		IS.II.		P.IV.		PE.		PP.III.		TL.II.	
		N	A	N	A	N	A	N	A	N	A	N	A	N	A
1	Nombre 1	B	2	B	1	B	1	B	3	B	2	B		B	1
2	Nombre 2	B	8	B	6	B	6	M	8	R	2	B		R	5
3	Nombre 2	B	5	B	2	B	12	M	10	R	3	B		R	3
...
N2	Nombre N2	B	1	B		B		B	1	B		B		B	

Fuente: elaboración propia.

En los dos ejemplos, se recogen resultados docentes que pueden ser tomados simultáneamente como evaluaciones y como diagnósticos del aprendizaje estudiantil. Ambos formatos concentran toda la información oficial lo que es muy positivo; sin embargo, su extensión y contenido puede acarrear las siguientes dificultades:

- Los resultados académicos en forma de notas no siempre reflejan el estado real de los conocimientos, habilidades y valores de los estudiantes ni otros aspectos de interés.
- Debido a que la cantidad de estudiantes es muy grande, se pierden de vista detalles importantes, pues muchas veces se presta atención a los individuos con más dificultades y se supone que el resto de los educandos no presentan deficiencias que superar.

- Se dificulta la gestión del conocimiento dentro de la junta de año y a los diferentes niveles de decisores, ya que no se utiliza toda la información disponible, la que debe ser cruzada para dar seguimiento a la evolución de los procesos, cosa que no se realiza con facilidad con las técnicas tradicionales de tratamiento de la información.
- El procesamiento de la información docente se realiza primordialmente a través de estadísticas descriptivas, las que pueden reducir la dimensión y estratificar en la misma medida en que destruyen la información particular, lo que no siempre es deseable.

Cuando los estudiantes arriban a la UCI, el Centro de Investigación de la Calidad Educativa (CICE) le aplica unos diagnósticos cognoscitivos, con el objetivo de constatar las aptitudes cognoscitivas con que estos comienzan sus estudios universitarios. Los diagnósticos se dividen en tres baterías de encuestas relacionadas con los perfiles de inteligencia, estrategias de aprendizaje y estilos de aprendizaje (Álvarez, 2012). Luego, a partir de los anteriores diagnósticos, la información que se recoge en el expediente de cada estudiante y el primer corte evaluativo se conforma una caracterización individual de cada estudiante.

Para el primer año de Ingeniería en Ciencias Informáticas, la caracterización individual se consideró que sería de utilidad para el trazado de estrategias pedagógicas, las cuales tenían como propósito evitar el éxodo de estudiantes al finalizar su primer año de estudios universitarios, reforzar las debilidades de bases y métodos estudios de los estudiantes. No obstante, hasta la actualidad los resultados obtenidos no han sido los esperados, debido a que no se ha explotado correctamente la caracterización y el conocimiento que se puede extraer de los datos que la conforman.

Teniendo en cuenta lo expresado anteriormente en la introducción, un aspecto importante a resolver se encuentra relacionado con la forma en que se procesan los datos que se acopian en las primeras semanas del curso (la ficha de matrícula, los diagnósticos cognoscitivos y el primer corte evaluativo) para su posterior interpretación y socialización. Es por ello que el procedimiento que se propone permite detectar los factores que modifican a la variable respuesta: “Promueve en tiempo” (Sí/No).

MATERIALES Y MÉTODOS

Datos

Los datos utilizados fueron la ficha de matrícula del estudiante, los diagnósticos pedagógicos, el primer corte evaluativo de los estudiantes

de primer año de la Facultad 1 de la (UCI) y las calificaciones finales del semestre, correspondientes al curso 2009-2010. Es importante destacar que en las calificaciones finales del semestre no se incluyeron las correspondientes al examen extraordinario, debido a que la información brindada por las mismas se encontraba fuera del tiempo establecido para aprobar la asignatura. Además, las calificaciones finales no se emplean como tal en el estudio, sino que sirven de base para conformar la variable respuesta: “Promueve en tiempo” (Sí/No). Los estudiantes que repiten el año también fueron excluidos del análisis, pues su carga docente y su experiencia en la educación superior difieren de los que arriban a la educación superior por vez primera. La cantidad de estudiantes que conforman la muestra es de 140, correspondientes a la primera cohorte 2009-2010.

Métodos

Para procesar los datos (la ficha de matrícula, los diagnósticos cognoscitivos y el primer corte evaluativo), se precisa cruzar información que puede provenir de diversas fuentes y generar en consecuencia un modelo que resuma el conocimiento extraído, que se pueda socializar e interpretar de manera mejor a como se hace en la actualidad.

El procedimiento se encuentra compuesto por dos métodos y el objetivo que se pretende alcanzar es el siguiente: identificar los factores que más pueden influir en los resultados finales del semestre. Los dos métodos estadísticos que se emplean utilizan técnicas complementarias para identificar los factores anteriormente mencionados.

Los métodos son los siguientes:

1. Árbol de Decisión (CRT).
2. Regresión Logística Binaria.
 - o Curvas ROC (para 1 y 2) (ver, (García, 2012).

Antes de comenzar a explicar cómo se emplearon cada uno de estos métodos, es importante resaltar que para cada uno de ellos se define la misma medida de eficiencia, o función objetivo para predecir la eficiencia, es la variable dicotómica “Promueve en tiempo” (Sí/No).

El primer método que se emplea es el Árbol de Clasificación y Regresión (CRT) y se utiliza este, debido a que el mismo tiene la capacidad de dividir

a los datos en segmentos que son los más homogéneos posible respecto a la variable dependiente, lo que permite poder encontrar los factores que manifiestan una incidencia directa en la promoción limpia al finalizar el primer semestre de la carrera Ingeniero en Ciencias Informáticas. El análisis multivariado (CRT) permitió detectar los factores de riesgo que influyen en los resultados de promoción al finalizar el primer semestre. Los factores fueron los siguientes:

- La calificación en el 1C1 de Matemática I (M.I.1C1), separada en dos categorías ordinales.
- La calificación en el 1C1 de Matemática Discreta I (MD.I.1C1), separada en dos categorías ordinales.
- La estrategia de aprendizaje por autocontrol motivacional (A.C.M), desglosada en sus dos categorías.
- El nivel de escolaridad del padre (N.E. del padre), separado en dos categorías.
- El centro de procedencia (C. Procedencia), desglosada en dos categorías ordinales.

Las variables restantes fueron probadas, pero no manifiestan asociaciones significativas con la variable respuesta: “Promueve en tiempo” (Sí/No). Es importante señalar que las cinco variables que presentan una asociación significativa con la variable respuesta, destacan como factores de riesgo en los resultados de promoción limpia al finalizar el semestre.

El segundo método que se emplea es la regresión logística binaria, la cual pretende brindar un modelo predictivo de la condición o estado del estudiante al finalizar el semestre, es decir, saber qué posibilidad tiene el estudiante de promover en tiempo (Sí/No). Este instrumento estadístico de análisis multivariado, posibilitó ratificar los resultados alcanzados en el análisis anterior (análisis multivariado (CRT)), pero a pesar de detectar los mismos factores de riesgo, hubo una variación en el desglose de las categorías. De las seis variables que constituyen factores de riesgo, en cinco se modifican sus categorías. Seguidamente, se puede apreciar cómo se modifican las categorías asociadas a cada variable:

- El nivel de escolaridad del padre solo tuvo en cuenta: es universitario (Sí/No).

- La evaluación del corte de matemática I (M.I.1C1) se separa en cuatro categorías:
 - o El estudiante se encuentra evaluado de regular (Sí/No).
 - o El estudiante se encuentra evaluado de bien (Sí/No).
- La evaluación en el corte de matemática discreta I (MD.I.1C1) solo tuvo en cuenta: el estudiante está evaluado de bien (Sí/No).
- El centro de procedencia solo examino: el estudiante pertenece a un Instituto Preuniversitario Vocacional (Sí/No).

Después de haber analizado los resultados que se obtienen tras la aplicación de los métodos anteriores, se evidencia que los factores obtenidos constituyen factores de riesgo de los resultados de promoción limpia al finalizar el semestre.

ANÁLISIS DE RESULTADOS

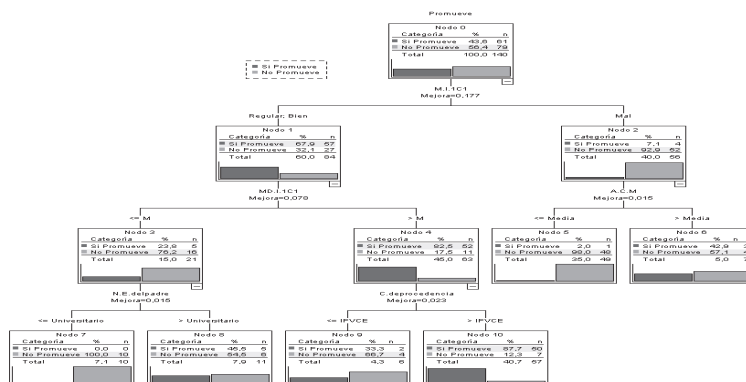
Fase 1: Identificación de los factores que más influyen en los resultados finales.

Los resultados obtenidos por cada una de las técnicas utilizadas para identificar a los factores que más influyen en los resultados finales se detallan en los próximos sub acápite.

Árbol de Decisión CRT

En la Figura 1 se muestra el árbol de decisión obtenido para la muestra en general. El árbol de decisión CRT, brinda un conjunto de reglas las cuales son útiles para el trazado de “Estrategias Pedagógicas” y es una técnica robusta ante la ausencia de valores perdidos. Las interacciones que se detectan son interesantes, pues en primer lugar se conjugan todas las variables utilizadas hasta el momento en los anteriores análisis, es decir: el Centro de Procedencia del estudiante (C.Procedencia), la calificación en el primer corte en las asignaturas de matemáticas (MD.I y M.I); también, se detectan la Estrategia de Aprendizaje por Autocontrol Motivacional (A.C.M) y el Nivel de Escolaridad del Padre (N.E. del padre). El total de nodos en este árbol es igual a 11, el número total de nodos terminales del árbol es igual a 6, que a su vez tienen asociados 6 caminos o trayectorias para llegar a cada uno de ellos. Para predecir la medida de eficiencia desde el nodo raíz hasta el nodo terminal, se define en cada caso una regla.

Figura 1: Árbol de decisión obtenido por la técnica CRT.



La matriz de confusión del clasificador multivariado empleado anteriormente, evidencia los excelentes resultados que se obtienen al utilizar este clasificador multivariado. El porcentaje de estudiantes con tendencia a no promover en tiempo es de un 91% aproximadamente y el porcentaje de estudiantes con tendencia a promover en tiempo es de un 82%, como se puede apreciar el porcentaje en ambos casos es bueno, demostrando la potencia del clasificador. Igualmente, la exactitud general alcanza un buen nivel del 87% aproximadamente, lo cual es muy apropiado para investigaciones en el campo de las Ciencias Sociales.

Regresión Logística Binaria

La regresión logística binaria pretende a partir de los datos de la ficha del estudiante, de los diagnósticos cognoscitivos y del primer corte evaluativo que se emite poder construir un modelo predictivo de la condición o estado del estudiante al finalizar el semestre, es decir, saber qué posibilidad tiene el estudiante de Promover en tiempo (Si/No).

Las seis variables que constituyen factores de riesgo pasan a formar parte de la ecuación logística, pues sus coeficientes son significativos estadísticamente según el valor del estadístico de Wald (evalúa la hipótesis nula). Las variables detectadas fueron:

- El nivel de escolaridad del padre solo tuvo en cuenta: es universitario (Si/No).
- La evaluación del corte de matemática I (M.I.1C1) se separa en cuatro categorías:

- o El estudiante se encuentra evaluado de regular (Sí/No).
- o El estudiante se encuentra evaluado de bien (Sí/No).
- La evaluación en el corte de matemática discreta I (MD.I.1C1) solo tuvo en cuenta: el estudiante está evaluado de bien (Sí/No).
- El centro de procedencia solo examino: el estudiante pertenece a un Instituto Preuniversitario Vocacional (Sí/No).

Es importante destacar que todas las variables que se incluirán en la ecuación de regresión constituyen factores de riesgo, excepto la estrategia de autocontrol motivacional (A.C.M) que evidencia no ser un factor de riesgo, sino de protección.

A partir de las variables, cuyos coeficientes son significativos estadísticamente según el valor del estadístico de Wald, se puede construir la ecuación de regresión logística, que en el presente estudio sería:

$$P(\text{Estado} = \text{No Promueve}) = \frac{1}{1 + \exp(3.879 - 1.560xN.Edelpadre_U + 2.920x A.C.M - 1.425x M.I.1C1_R - 2.978x M.I.1C1_B - 2.361x MDI.1C1_B - 1.548x C.deprocedencia_IPV)}$$

La anterior ecuación puede ser utilizada para predecir la probabilidad de tener el resultado (Estado) de “No Promueve” de un estudiante que presenta los factores de riesgo anteriormente expuestos. Así, un estudiante que su padre no tenga un nivel de escolaridad universitario, que su evaluación en matemática sea de mal, que su evaluación en matemática discreta sea de mal o de regular y que el centro de procedencia no sea un instituto preuniversitario vocacional, presenta una probabilidad de No Promover en tiempo igual a:

$$P(\text{Estado} = \text{No Promueve}) = \frac{1}{1 + \exp(3.879 - 1.560x1 + 2.920x0.5 - 1.425x1 - 2.978x1 - 2.361x1 - 1.548x1)} = 0.98$$

Por lo tanto, con esta probabilidad predicha, como es mayor que 0.50 se clasificaría como “Estado=No Promueve en tiempo”.

El modelo permite calcular una predicción del resultado en escala de probabilidad. También, se puede clasificar a un estudiante en cualquiera de los grupos de sucesos del presente estudio, es decir: (Grupo de estudiantes que No Promueven) y (Grupo de estudiantes que Promueven), ello está dado por el hecho que su probabilidad supere (Sí/No) un valor de corte dado para la clasificación, en este caso 0.50 es adecuado.

La Tabla 3 es un extracto de los pronósticos hechos por el modelo de regresión logística, se conoce que las seis variables que conforman al modelo presentan dos categorías (1 ó 0). Por lo tanto, se tendrían 2^6 combinaciones posibles a partir de los dos niveles en los que se miden cada variable predictora del modelo, lo que equivale a 64 evaluaciones del modelo de regresión logística. Se puede apreciar que el modelo es capaz de diferenciar de forma correcta los estudiantes que promueven de los que no promueven, lo cual es muy útil para la toma de decisiones de los directivos docentes.

Tabla 3: Resultados de la Regresión Logística.

N.E.P	A.C.M	M.I.IC1_R	M.I.IC1_B	MD.I.IC1_B	C.P_IPV	Pronóstico	Promueve
0	0	1	1	1	0	0,95	No Promueve
0	0	1	1	1	1	0,99	No Promueve
0	1	1	1	1	0	0,22	Promueve
0	1	0	1	1	0	0,14	Promueve

Fuente: elaboración propia.

CONCLUSIONES

Los resultados del estudio realizado evidencian que el procedimiento implementado para el procesamiento de los datos educacionales correspondientes a la etapa inicial del primer año de la carrera de Ingeniería en Ciencias Informáticas es de gran importancia. En primer lugar, permite determinar los posibles factores de riesgo que pueden influir de forma directa en los resultados finales del semestre. También, se muestra el modo de empleo de métodos multivariados, a través de los cuales se logró describir la interrelación que existe entre los factores de riesgo detectados y la medida de eficiencia utilizada: “Promueve en tiempo (Sí/No)”.

El procedimiento brindado permite sintetizar el posible desempeño de cada estudiante a partir de los factores de riesgo. La interpretación de los resultados obtenidos, constituyen el punto de partida para el trazado de estrategias por parte de los directivos docentes; las cuales enriquecerán el cúmulo de experiencias que se acopian cada día del proceso docente educativo en nuestras universidades.

REFERENCIAS BIBLIOGRÁFICAS

- Álvarez, J. H. Q. (2012). Procesamiento del Diagnóstico Pedagógico mediante Algoritmos de Minería de Datos. Tesis de maestría, Departamento de Ciencias Básicas, Facultad 1, Universidad de las Ciencias Informáticas (UCI), Cuba.
- Brito, R. (2008). Minería de Datos aplicada a la Gestión Docente del Instituto Superior Politécnico José Antonio Echeverría. Tesis de maestría, Instituto Superior Politécnico “José Antonio Echeverría” (ISPJAE), Cuba.
- Delavari, N. a. M. R. B. (2008). Data Mining Application in Higher Learning Institutions. *Informatics in Education* 7(1): 31-54.
- Espinosa, I. a. S. P. (2007). Obtención de Reglas y Patrones en el Proceso Académico de la Universidad de Ciencias Informáticas. Tesis de ingeniería, Instituto Superior Politécnico “José Antonio Echeverría” (ISPJAE), Cuba.
- Gallander, M. y Dilouya, B. (2011). Academic achievement in first-year university: who maintains their High School average? *Higher Education*, V. 62, p. 467-481.
- García, L., Grau, I. y Grau, R. (2012). Predicting the Efficiency with Knowledge Discovery of a Budgeted Company: a Cuban University. Validation through three semesters. *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval, KDIR 2012. # pages 4. SCITEPRESS.*
- Garnica, E. (1997). El rendimiento estudiantil: una metodología para su medición. *Revista de Economía* 13: 5-26.
- King, O. A. (2008). Sistema basado en instancias para pronosticar el éxito de un estudiante a su ingreso en una carrera universitaria. Tesis de licenciatura, Universidad Central de las Villas, Cuba.
- Moral de la Rubia, J. (2006). Predicción del rendimiento académico universitario. *Perfiles Educativos* 28(113): 43-61.
- Remón, N. (2009). Análisis para la predicción del éxito o fracaso académico de estudiantes de la Universidad de las Ciencias Informáticas mediante la teoría de conjuntos aproximados. Tesis de ingeniería. Universidad de las Ciencias Informáticas (UCI), Cuba.

- Reyes, J. y Escobar, C. (2007). Una aplicación del modelo de regresión logística en la predicción del rendimiento estudiantil. *Estudios pedagógicos* 33(2): 101-120.
- Rico, J. J. H. (2011). Análisis de datos en apoyo a la productividad en el proceso de formación de ingenieros Facultad de Ingeniería Industrial. Tesis de maestría. Instituto Superior Politécnico “José Antonio Echeverría” (ISPJAE), Cuba.
- Rodríguez G., y. C. (2009). Análisis Multivariado. La Habana, Instituto Superior Politécnico “José Antonio Echeverría” (ISPJAE). Monografía.
- Zúnica, L. R., R. M. Alcover (2005). Relación entre el rendimiento de dos asignaturas de segundo curso y las asignaturas de primer curso en Ingenierías Técnicas de Informática de la UPV. XI Jornadas de Enseñanza Universitaria de la Informática. Madrid, ATI.