

Estudio del rendimiento académico y determinación temprana de perfiles de alumnos en la Facultad Politécnica de la Universidad Nacional del Este de Paraguay.

Gabriela Bobadilla de Almada¹ y David L. la Red Martínez².

Facultad Politécnica - Universidad Nacional del Este¹, Facultad Ciencias Exactas y Naturales y Agrimensura - Universidad Nacional del Nordeste ²
Ciudad del Este - Paraguay¹, Corrientes - Argentina²
gaby@fpune.edu.py¹, laredmartinez@gigared.com²

Resumen

La motivación para realizar este trabajo proviene de la acumulación de alumnos en los primeros semestres de las carreras tecnológicas de la Facultad Politécnica de la Universidad Nacional del Este (FPUNE), Paraguay; y del bajo porcentaje de egresados en su cohorte. Por esta razón se buscó identificar las variables que influyen en el rendimiento académico de los alumnos de las carreras de Análisis de Sistemas, Ingeniería de Sistemas e Ingeniería Eléctrica, utilizando tecnologías de almacén de datos (DW) y de minería de datos (DM). Con el DW se han preparado los datos sobre los cuales se aplicaron las técnicas de clústeres, asociación y clasificación de la DM, que han evidenciado las características representativas de alumnos con rendimiento académico muy bueno, regular y reprobado. Se observó que el grado educacional de los padres, la actitud general hacia el estudio y la utilización de las TICs inciden en el rendimiento académico de los alumnos y que los promedios generales del segundo semestre correlacionan significativamente con los valores de la situación académica global de los alumnos de los cinco primeros semestres. Con esta investigación se obtuvo una herramienta útil para identificar en los primeros semestres del cursado, a los alumnos que más tarde pudieran tener dificultades e incluso llegar a desertar de sus carreras.

Descriptores: Minería de Datos, Almacenes de Datos, Rendimiento Académico.

Abstract

The motivation for this work comes from the accumulation of students in the first semesters of technological degree programs from Polytechnic Faculty of the Eastern National University (FPUNE), Paraguay; and the low percentage of graduates in their cohort. For this reason it was sought to identify variables that affect the academic performance of students from degree programs: Systems Analysis, Systems Engineering, and Electrical Engineering; employing techniques of data warehouse (DW) and data mining (DM). DW was used in preparing data on which DM techniques of clustering, association and classification were applied; representative characteristics have been highlighted of students with very good, fair and failed academic performance. It was observed that educational level of parents, general attitude towards study and use of ICTs affect the academic performance of students and the overall averages of the second semester correlate significantly with the values of the global academic situation of students of the first five semesters. With this research a useful tool was acquired to identify in the first semesters, students who later might have difficulty even deserting their careers.

Keywords: Data Mining, Data Warehouse, Academic Performance.

1. Introducción.

La motivación para realizar este trabajo proviene de la acumulación de alumnos en los primeros semestres de las carreras tecnológicas de la Facultad Politécnica de la Universidad Nacional del Este (FPUNE), Paraguay, y del bajo porcentaje de egresados en su cohorte. Esta realidad puede corroborarse en los gráficos de la figura 1, donde se analiza el porcentaje de egresados sobre los ingresantes en el periodo 1996 al 2010 y en la figura 2, donde se observa un desfase de egreso respecto del año de cohorte.

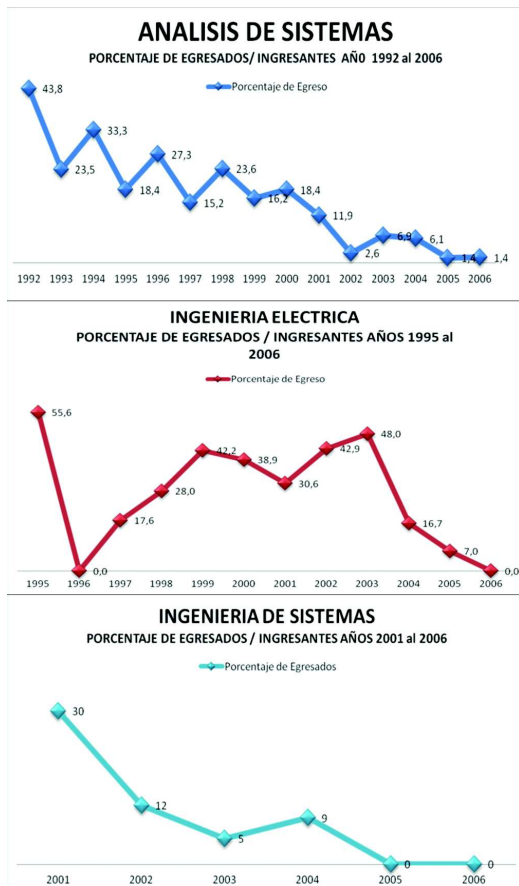


Figura 1. Porcentaje de egreso - (Dirección de Informática FPUNE, 2013).

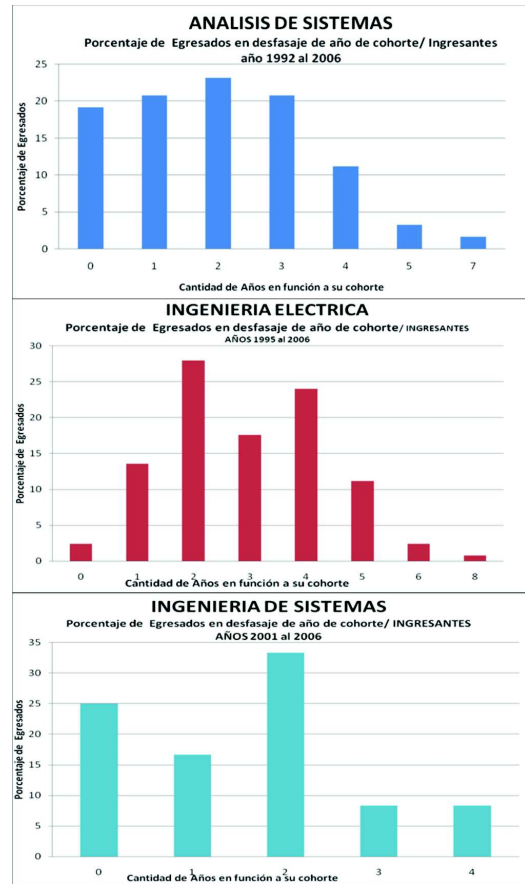


Figura 2. Porcentaje de egresados en desfase de cohorte - (Dirección de Informática FPUNE, 2013).

En este contexto se ha considerado oportuno realizar una investigación que permita determinar las variables que inciden en el rendimiento académico de los alumnos [20], [2], [3], [17], [6], [13],[10], [11], [8], [4], [21], [16],[18], [19], [24], identificar perfiles de alumnos exitosos (los que promocionan en su cohorte), como así también los perfiles de alumnos que no lo logran (los que quedan desfasados a su cohorte o desertan). Una vez determinados los perfiles de alumnos, esto podrá servir de información referencial para plantear acciones tendientes a evitar potenciales fracasos académicos. Se utilizó como herramientas para el estudio de la problemática presentada, las técnicas de Almacenes de Datos (Data Warehouse: DW) y Minería de Datos (*Data Mining*: DM) [7], [14], [15], [22], [23], [5], [1].

2. Modelo utilizado.

Se aplicó la metodología CRISP-DM [3] para el desarrollo de las etapas del proceso de descubrimiento del conocimiento (KDD), figura 3 se muestran las fases del proceso de modelado.

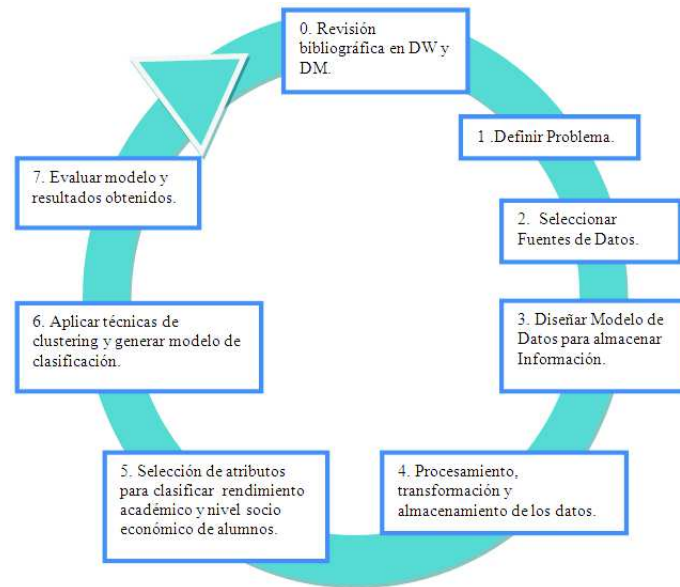


Figura 3. Fases del proceso de modelado de la metodología CRISP-DM aplicado (fuente propia).

Los datos sobre los cuales se realizó el estudio se obtuvieron a través de las encuestas socio económicas aplicadas a los alumnos de las carreras de Licenciatura en Análisis de Sistemas, Ingeniería Eléctrica e Ingeniería de Sistemas y de la base de datos del sistema de gestión académica de la FPU-NE, de donde se tomaron los promedios generales obtenidos por los alumnos desde el primer semes-

tre al quinto semestre. El universo estuvo constituido por los alumnos que ingresaron en los años 2007 al 2013 en las carreras mencionadas.

El DW es construido según el modelo User Driven [9] y la metodología Big Bang [12], de estructura multidimensional con el esquema en estrella que se muestra en la figura 4.

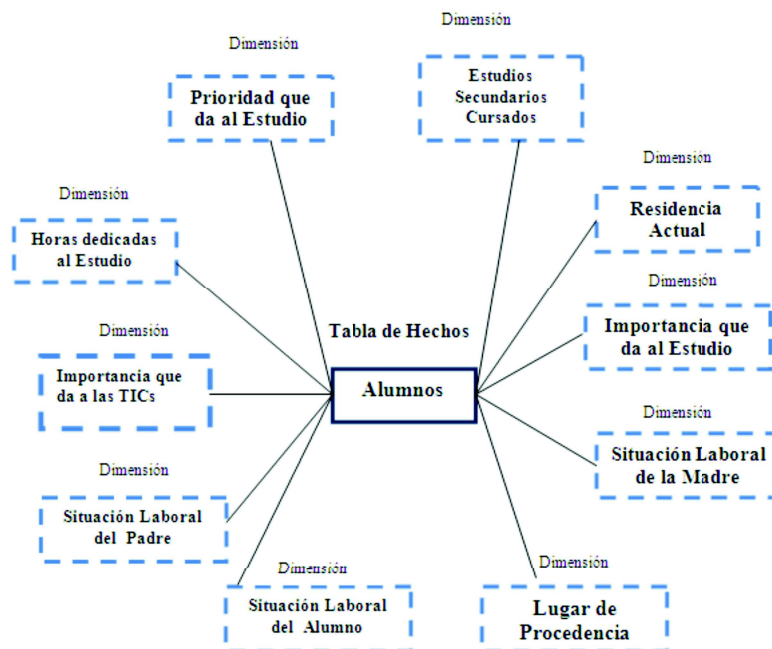


Figura 4. Esquema básico del Almacén de Datos (fuente propia).

Para modelar la técnica de DM se utilizó el software de IBM Business Intelligence, particularmente DB2 Warehouse Design Studio.

Se ha aplicado las técnicas de DM de clústeres, asociación y clasificación, sobre el DW construido, que han evidenciado las características de las

clases representativas de alumnos con rendimiento académico regular muy bueno, regular y reprobado.

3. Resultados obtenidos.

Con el modelo de clústeres se identificó en las agrupaciones formadas las características de los alumnos de acuerdo a su situación final, figura 9.

Clúster[1] de tamaño 28,90%	Clúster[4] de tamaño 17,73%	Clúster [9] de tamaño 11,75%
Situación final: Regular	Situación final: Regular Muy Bueno	Situación final: Reprobado
Género: Masculino	Género: Femenino	Género: Masculino
Opinión sobre Utilización de las TICs: facilitan el proceso de enseñanza	Opinión sobre Utilización de las TICs: será imprescindible su dominio para el ejercicio profesional	Opinión sobre Utilización de las TICs: facilitan el proceso de enseñanza
Último estudio del Padre: secundaria completa	Último estudio del Padre: universitario completo	Último estudio del Padre: primaria completa
Situación Laboral del Padre: ocupado	Situación Laboral del Padre: ocupado	Situación Laboral del Padre: ocupado
Categoría Ocupacional del Padre: otra ocupación no específica	Categoría Ocupacional del Padre: propietario de mediano y pequeño negocio	Categoría Ocupacional del Padre: no específica
Último estudio de la Madre: secundaria completa	Último estudio de la Madre: universitario completo	Último estudio de la Madre: primaria incompleta
Situación Laboral de la Madre: no trabaja	Situación Laboral de la Madre: ocupada	Situación Laboral de la Madre: no trabaja
Categoría Ocupacional de la Madre: no registra	Categoría Ocupacional de la Madre: profesional independiente	Categoría Ocupacional de la Madre: no registra

Figura 5. Cuadro resumen de resultados de Cluster (fuente propia).

Los resultados obtenidos con Árboles de Decisión se han resumido en las tablas 1, 2, 3, correspondientes a las tres categorías en que se ha clasificado la situación final según el rendimiento académico: regular, regular muy bueno y reprobado.

Tabla 1. Resumen de resultados de la clase Regular (fuente propia).

Clase Regular	Atributo	Valor Predominante
44,75 % de la población	Situación final	Regular SIT_FINAL ≥ 2 y $< 3,5$ (100 %).
	Género del alumno	Masculino (74 %)
	Tipo de Bachillerato	Bachiller Técnico en Informática (31 %)
	Tipo de institución educativa en que culminó sus estudios secundarios	Nacional (72 %)
	Opinión sobre la utilización de las TICs	Facilitan el proceso de enseñanza (45 %)
	Importancia que le da al estudio	Más que a la diversión (66 %)
	Tipo de residencia	Con familiares (82 %)
	Situación laboral del alumno	No trabaja (60 %)
	Situación laboral del padre	Ocupado (81 %)
	Último estudio realizado por el padre	Escuela Secundaria Completa (32 %)

Tabla 2. Resumen de resultados de la clase Regular Muy bueno (fuente propia).

Clase Regular Muy Bueno	Atributo	Valor Predominante
33,25 % de la población	Situación final	Regular Muy Bueno SIT_FINAL $\geq 3,5$ y ≤ 5 (100 %)
	Género del alumno	Femenino (64 %)
	Tipo de Bachillerato	Bachiller Científico Humanístico (38 %)
	Tipo de institución educativa en que culminó sus estudios secundarios	Nacional (83 %)
	Opinión sobre la utilización de las TICs	Será imprescindible su dominio para el ejercicio profesional (68 %)
	Importancia que le da al estudio	Más que a la diversión (83 %)
	Tipo de residencia	Con familiares (95 %)
	Situación laboral del alumno	No trabaja (54 %)
	Situación laboral del padre	Ocupado (73 %)
	Último estudio realizado por el padre	Escuela Primaria Incompleta (25 %)
Situación laboral de la madre	Ocupado (59 %)	
Último estudio realizado por la madre	Escuela Primaria Completa (29 %)	

Tabla 3. Resumen de resultados de la clase Reprobado (fuente propia).

Clase Reprobado	Atributo	Valor Predominante
22 % de la población	Situación final	Regular SIT_FINAL ;2 (100 %)
	Género del alumno	Masculino (74 %)
	Tipo de Bachillerato	Bachiller Científico Humanístico (50 %)
	Tipo de institución educativa en que culminó sus estudios secundarios	Nacional (75 %)
	Opinión sobre la utilización de las TICs	Facilitan el proceso de enseñanza (56 %)
	Importancia que le da al estudio	Más que a la diversión (50 %)
	Tipo de residencia	Con familiares (88 %)
	Situación laboral del alumno	Trabaja (57 %)
	Situación laboral del padre	Ocupado (76 %)
	Último estudio realizado por el padre	Escuela Primaria Completa (30 %)
	Situación laboral de la madre	No trabaja (52 %)
	Último estudio realizado por la madre	Escuela primaria Incompleta (26 %)

Con la minería de predicción con regresión a los datos de la tabla de notas, se obtiene como resultado que los promedios del segundo semestre influyen en mayor porcentaje en las del quinto semestre de la carrera que cursa, figura 6.

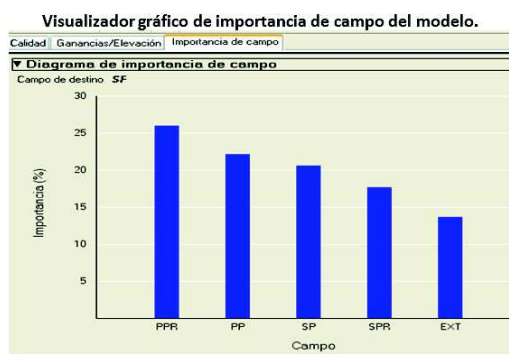


Tabla de importancia de campo.

Campo	Descripción	Importancia
PPR	segundo semestre	25,97%
PP	primer semestre	22,13%
SP	tercer semestre	20,59%
SPR	cuarto semestre	17,66%
EXT	quinto semestre	13,65%

Figura 6. Resultados obtenidos con predicción con regresión (fuente propia).

4. Conclusiones.

Los resultados obtenidos con la aplicación de las técnicas de minería de datos de clústeres, asociación y clasificación han evidenciado las características de las clases representativas de alumnos con rendimiento académico regular muy bueno, regular y reprobado.

Con el modelo de clústeres se identificó en las agrupaciones formadas las características de los alumnos de acuerdo a su situación final. Con la aplicación de la asociación se logró fijar las variables que consistentemente se asocian en función de las características de la situación final del alumno. Con el modelo de clasificación a través de árboles de decisión, se predijo las características de las clases formadas de acuerdo a la situación final del alumno. Con el modelo de clasificación con regresión se predijo cuál de los promedios de las notas de los cinco primeros semestres influyen en la situación final del alumno.

Con esta metodología se podría identificar en los primeros semestres del cursado de la carrera, a los posibles alumnos que podrían llegar a desertar de sus carreras; con la aplicación a tiempo del programa de tutoría y otras medidas a los alumnos identificados se podría evitar su mal desempeño académico y que llegaran a la deserción.

Referencias bibliográficas

- [1] 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.
- [2] Carrasco Pradas, A.; Gracia Expósito, E.; de la Iglesia Villasol, 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.
- [3] Di Gresia, L. (2007). *Rendimiento Académico Universitario*. Tesis Doctoral. Universidad Nacional de La Plata. Argentina.
- [4] Fazio, M. V. (2004). *Incidencia de las horas trabajadas en el rendimiento académico de estudiantes universitarios argentinos*. Documentos de Trabajo UNLP, 52. Argentina.
- [5] Fayyad, U. M.; Grinstein, G. & Wierse, A. (2001). *Information Visualization in Data Mining and Knowledge Discovery*. Morgan Kaufmann. Harcourt Intl.
- [6] García, M. M.; San Segundo, M. J. (2001). *El Rendimiento Académico en el Primer Cur-*

- so Universitario*. X Jornadas de la Asociación de Economía de la Educación. Libro de Actas, pp. 435-445. España.
- [7] García Jiménez, M. V.; Alvarado Izquierdo, J. M.; Jiménez Blanco, A. (2000). La predicción del rendimiento académico: regresión lineal versus regresión logística. *Psicothema* Vol. 12, Supl. N° 2, pp. 248-252. España.
- [8] Giovagnoli, P. I. (2002). *Determinantes de la deserción y graduación universitaria: Una aplicación utilizando modelos de duración*. Documentos de Trabajo UNLP, 37. Argentina.
- [9] Gutting, R. (1994). An Introduction to spatial database systems. *VLDB Journal*, 3, 357-399.
- [10] Hanushek, E. (1979). Conceptual and empirical issues in the estimation of educational productions. *Journal of Human Resources*, (14), pp. 351 - 388.
- [11] Hanushek, E. (1986). The economics of schooling production and efficiency in public schools. *Journal of Economic Literature*, (24), pp. 1141-1177.
- [12] Harinarayan V.; Rajaraman, A.; Ullman, J. (1996). Implementation data cubes efficiently. *ACM SIGMOD Record*, 25 (2), 205 - 216.
- [13] Herrera Clavero, F. et al. (2004). Cómo Interactúan el Autoconcepto y el Rendimiento Académico en un Contexto Educativo Pluricultural. *Revista Iberoamericana de Educación*. España.
- [14] Inmon, W. H. (1992). *Data Warehouse Performance*. John Wiley & Sons. USA.
- [15] Inmon, W. H. (1996). *Building the Data Warehouse*. John Wiley & Sons. USA.
- [16] Johnson, D.; Johnson, R. (1985). *Motivational processes in cooperative, competitive, and individualistic learning situations*. New York. C. Ames & R. Ames Eds. Research on motivation in education. Vol. 2: The classroom milieu (pp. 249-286). Academic Press. USA.
- [17] 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.
- [18] La Red Martínez, D. L.; Podestá Gómez, C. E. (2014). Contributions from Data Mining to Study Academic Performance of Students of a Tertiary Institute; Volume 02 - N° 9; *American Journal of Educational Research*; pp. 713-726; ISSN N° 2327-6126; U.S.A.
- [19] La Red Martínez, D. L.; Karanik, M.; Giovannini, M.; Pinto, N. (2015). Academic Performance Profiles: A Descriptive Model Based on Data Mining; Volume 11 - N° 9; *European Scientific Journal (ESJ)*; pp. 17-38; ISSN N° 1857-7881; University of the Azores, Portugal.
- [20] Montero Rojas, E.; Villalobos Palma, J.; Valverde Bermúdez, A. (2007). Factores Institucionales, Pedagógicos, Psicosociales y Sociodemográficos Asociados al Rendimiento Académico en la Universidad de Costa Rica: Un Análisis Multinivel. *Revista Electrónica de Investigación y Evaluación Educativa (RELIEVE)*. Universidad de Costa Rica, 13 (2), 215-234.
- [21] Navarro, R. E. (2003). Factores Asociados al Rendimiento Académico. *Revista Iberoamericana de Educación*. España.
- [22] Simon, A. (1997). *Data Warehouse, Data Mining and OLAP*. John Wiley & Sons. USA.
- [23] Trujillo, J. C.; Palomar M.; Gómez, J. (2000). *Applying Object-Oriented Conceptual Modeling Techniques to the Design of Multidimensional Databases and OLAP Applications*. First International Conference On Web-Age Information Management (WAIM'00). Lecture Notes in Computer Science 1846:83-94.
- [24] 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.
- [25] Chapman, P. et al. (2000). CRISP-DM 1.0: Step-by-step. Data Mining Guide. <http://www.crisp-dm.org/CRISPWP-0800.pdf>