

**Análisis del perfil de aspirantes a carreras profesionalistas
en una institución de educación superior aplicando
técnicas de aprendizaje no supervisado.**

Trabajo presentado y publicado en los Anales de las:

**Decimoterceras Jornadas en Estadística e Informática
Llevadas a cabo en Guayaquil, Ecuador, durante octubre de 2006.**

Análisis del perfil de aspirantes a carreras profesionalistas en una institución de educación superior aplicando técnicas de aprendizaje no supervisado.

Dr. PAULINO E. MALLO; *Universidad Nacional de Mar del Plata*;
Tel.-Fax: +542234747906, paulinomallo@uolsinectis.com.ar
CPN MARÍA A. ARTOLA; *Universidad Nacional de Mar del Plata*;
Tel.: +54 223 4735059, Fax: +54 223 4737631, martola@infovia.com.ar
Mg. Lic. ALICIA I. ZANFRILLO; *Universidad Nacional de Mar del Plata*;
Tel.: +54 223 4752240, Fax: +54 223 4749696 int. 309, alicia@mdp.edu.ar
CPN Lic. MARIANO MORETTINI; *Universidad Nacional de Mar del Plata*;
Tel.: +54 223 4736470, Fax: +54 223 4750377, mmoretti@mdp.edu.ar
CPN Lic. MARCELO J. GALANTE; *Universidad Nacional de Mar del Plata*;
Tel.-Fax: +54 223 4925006, mjgalante@uolsinectis.com.ar
CPN Lic. ADRIÁN R. Busetto; *Universidad Nacional de Mar del Plata*;
Tel.: +54 223 4896320, adrianbusetto@hotmail.com
CPN Lic. MARIANO E. PASCUAL; *Universidad Nacional de Mar del Plata*;
Tel.: +54 223 4796568, mpascual@copetel.com.ar

Resumen

Uno de los propósitos del quehacer de las Instituciones de Educación Superior es la promoción del nivel académico del alumnado, dotándolo de las herramientas necesarias para su eficaz desenvolvimiento en el campo profesional y científico, para lo cual es condición necesaria que el aspirante reúna determinados requisitos mínimos de conocimientos y aptitudes que le permitan desenvolverse exitosamente a lo largo de la carrera elegida y egresar con un nivel académico adecuado a las exigencias del medio y la evolución de la sociedad en el campo científico – tecnológico.

El debate producido con el nivel secundario o polimodal de nuestro país, trajo como corolario la necesidad de implementar nuevas metodologías pedagógicas, ya sea semipresencial y/o a distancia, que permitan la evaluación de conocimientos previamente, que resulten más eficaces, dinámicas y superadoras con relación a los resultados de admisión.

El presente trabajo tiene como propósito identificar grupos de aspirantes con características comunes en cuanto a su desempeño en los exámenes de admisión a las carreras profesionalistas de la Facultad de Ciencias Económicas y Sociales de la Universidad Nacional de Mar del Plata de la República Argentina. Para proveer de este tipo de información a la gestión académica se

utiliza una técnica multivariante de análisis exploratorio, el análisis cluster, a fin de obtener una clasificación de los datos que resulte relevante para la adopción de políticas conducentes al mejoramiento de la calidad educativa.

Palabras clave: ingreso universitario – instituciones de educación superior — perfil aspirantes – análisis cluster

1. Introducción

El ingreso universitario constituye hoy en día un tema importante en la agenda universitaria. Ya sea desde las áreas de ingreso de las unidades académicas, de la unidad central de la Universidad o desde la Secretaría de Políticas Universitarias dependiente del Ministerio de Educación, se realizan acciones concretas para la articulación con el nivel de Educación Superior en cuanto a financiamiento de proyectos de articulación.

Numerosos estudios se han llevado a cabo a fin de establecer las causas del fracaso de los aspirantes en las pruebas de admisión. Así, la deficiencia en el nivel secundario o polimodal, el nivel socio-económico y educativo de sus padres son factores determinantes en su paso a la Universidad. El análisis de las causas de la lentificación - atraso en el avance de la carrera - y el abandono presenta dos tipos de factores: exógenos y endógenos. Para el caso de análisis, el ingreso universitario, se

mencionan los factores exógenos a la institución universitaria que resultan más relevantes: género, edad, residencia, nivel socio-económico, nivel educativo de los padres, condición de actividad económica del estudiantes, formación académica previa y las aspiraciones y motivaciones individuales (García de Fanelli, 2005: 72-75).

La evaluación del desempeño de los aspirantes de la Institución de Educación Superior - IES - se lleva a cabo por lo general, utilizando técnicas descriptivas univariadas o bivariadas. La distribución de aspirantes según sus calificaciones en las pruebas de admisión por carrera, por grupo étnico, por sexo, por tipo de establecimiento y por titulación secundaria o polimodal son las presentaciones tradicionales para determinar la distribución del rendimiento según las variables clasificatorias disponibles.

La búsqueda de correlación o asociación entre las variables del análisis constituye la práctica usual en este tipo de problemática. En la práctica, dichas variables pueden no presentar la correspondencia deseada, es en este tipo de situaciones en que las técnicas de minería de datos resultan útiles para la obtención de información a partir de un conjunto de variables independientes. Su utilización se justifica además en la respuesta a interrogantes planteados por (Luan, 2001: 6).

Así, la utilización de otras técnicas en el análisis del proceso de admisión a los estudios de nivel superior, enriquecen la información obtenida anteriormente, con la adición de la selección de las variables más determinantes según la configuración presentada y su importancia o jerarquía en el proceso.

Con la incorporación de la minería de datos al bagaje de herramientas de análisis para las problemáticas de la Educación Superior, se busca favorecer la búsqueda de respuestas a interrogantes del tipo: por qué fracasan los alumnos en su examen de admisión? O bien: cuál es el perfil del alumno que ingresa efectivamente a la unidad académica? (Luan, 2001: 3).

2. Metodología utilizada

El análisis cluster o de conglomerados que se emplea para clasificar objetos o individuos

donde no se conoce a priori los grupos que se formarán. La finalidad de la técnica consiste en la formación de grupos, de modo tal que cada agrupamiento sea lo más homogéneo posible y lo más distintos entre los diferentes grupos conformados.

En nuestro caso, los individuos corresponden a los aspirantes. La aplicación de esta técnica de carácter descriptivo tiene como propósito determinar el número de grupos y las características que los definen.

Los datos utilizados en la matriz de datos – 1374 casos válidos – para la construcción del modelo corresponden a los aspirantes inscriptos para el Ingreso 2006 a la FCEyS de la UNMdP, en las carreras de Contador Público – CP –, Licenciatura en Administración – LA –, Licenciatura en Economía – LE – y Licenciatura en Turismo – LT – del plan de estudios 2005.

Se seleccionaron las variables consideradas como más significativas para el análisis y con la confiabilidad requerida para el estudio. Al conjunto de variables seleccionadas se le aplicaron los análisis estadísticos tradicionales de asociación y correlación. Sobre este conjunto de variables se aplica la técnica de análisis cluster y se realiza la validación del modelo a través de una matriz de confusión.

El análisis de los datos utilizando análisis bietápico se realiza a través del software estadístico SPSS, versión de evaluación 13.0 (SPSS, 2001).

2.1 Variables utilizadas

Las variables seleccionadas para el modelo son:

- *Carrera*: CP, LA, LE y LT.
- *Sexo*: Masculino o Femenino.
- *Tipo secundario / polimodal*: tipo de institución donde el aspirante cursó sus estudios secundarios o de polimodal. Valores: Nacional, Provincial, Municipal, Privado no religioso, Privado religioso, Instituto militar, Seminario, Extranjero y Otros.
- *Años egreso secundario / polimodal*: cantidad de años transcurridos desde la fecha de egreso de sus estudios secundarios o de polimodal. Valores: 0 a 30 años.
- *Título secundario*: título otorgado por la entidad en que el aspirante cursó sus estudios secundarios o de

polimodal. Valores: Bachiller, Perito Mercantil, Técnico, Maestro y Otros.

Las variables que corresponden a país, provincia, partido y ciudad de procedencia no se utilizan en el análisis por las inconsistencias detectadas en los valores de las distintas categorías de las mismas.

3. Análisis descriptivo

3.1 Análisis estadístico

El porcentaje de ingresantes es del 25,1%, el 7,3% corresponde a los exceptuados del examen de admisión – y el 67,6% a los no ingresantes – Figura 1 -.

La excepción al examen de admisión se efectúa cuando el aspirante acredita su titulación de nivel secundario o polimodal del Colegio Nacional Arturo Illia - dependiente de la UNMDP - o es uno de los tres mejores promedios correspondientes a establecimientos secundarios o polimodales públicos nacionales o bien, ha aprobado una asignatura de otra carrera de una institución universitaria.

	Condición			Total
	Ingresante	No ingresante	Exceptuado	
Frequency	345	929	100	1374
Percent	25,1	67,6	7,3	100,0
Valid Percent	25,1	67,6	7,3	100,0
Cumulative Percent	25,1	92,7	100,0	

Figura 1

Para nuestro estudio se ha realizado una selección de aquellos aspirantes cuya condición frente al examen de admisión es de desaprobado, descriptos anteriormente como 'no ingresantes'.

3.2 Análisis bi-etápico

Con el propósito de obtener las variables que resultan más determinantes en la caracterización del perfil de los aspirantes que no han concluido exitosamente su prueba de admisión, se propone la utilización de un modelo de agrupamiento que permita el análisis de variables cuantitativas y

cualitativas para determinar su contribución en la caracterización de los distintos agrupamientos.

El análisis bi-etápico revela cuatro agrupamientos. En la tabla que se presenta en la Figura 2 se muestra la proporción de observaciones y su relación con respecto al total en cada uno de los grupos identificados.

Cluster Distribution

	N	% of Combined	% of Total
Cluster 1	159	17,1%	17,1%
2	277	29,8%	29,8%
3	214	23,0%	23,0%
4	279	30,0%	30,0%
Combined	929	100,0%	100,0%
Total	929		100,0%

Figura 2

A continuación se presenta el porcentaje de contribución de cada variable en la definición de cada conglomerado. En la Figura 3 se observa el aporte de la variable carrera.

En la Figura 4 se observa el aporte de la variable sexo.

En la Figura 5 se observa el aporte de la variable tipo secundario/polimodal.

En la Figura 6 se observa el aporte de la variable título secundario/polimodal.

En la Figura 7 se observa el aporte de la variable años transcurridos desde el egreso del secundario/polimodal.

Para validar la solución obtenida se recurrió al análisis discriminante, utilizando como variable dependiente una variable nueva, relativa a la pertenencia de cada caso al cluster creado por el análisis bi-etápico y como variables independientes las mismas que las empleadas para el análisis cluster.

La matriz de confusión - tabla de resumen - generada entre los conglomerados predichos y la nueva variable del análisis discriminante muestra un porcentaje de asignaciones correctas superiores al 80% como se muestra en la Figura 8, lo que indica una bondad del análisis aceptable (Penelas, 1999; Santesmases Mestre, 1997).

Classification Results^a

TwoStep Cluster Number		Predicted Group Membership				Total
		1	2	3	4	
Original	Count	1	2	3	4	
		156	0	0	3	159
		21	219	16	21	277
		15	7	189	3	214
		53	0	39	187	279
%		1	2	3	4	
		98,1	,0	,0	1,9	100,0
		7,6	79,1	5,8	7,6	100,0
		7,0	3,3	88,3	1,4	100,0
		19,0	,0	14,0	67,0	100,0

a. 80,8% of original grouped cases correctly classified.

Figura 8

En la definición de los clusters se observan las siguientes características:

- En el cluster 1 se observan aspirantes de la carrera de Contador Público, de sexo femenino, con titulación de bachiller mayoritariamente provenientes de escuelas públicas.
- En el cluster 2 se observan aspirantes de la mayoría de las carreras – en menor medida Lic. en Economía – provenientes de escuelas públicas y privadas con titulación de perito mercantil y técnico en su gran mayoría con la mayor cantidad de años transcurridos desde el egreso.
- En el cluster 3 se observan aspirantes de la mayoría de las carreras – en mínima proporción Lic. en Turismo – mayoritariamente masculinos y provenientes principalmente de escuelas públicas, de titulación bachiller y una moderada cantidad de años transcurridos desde el egreso.
- En el cluster 4 se observan aspirantes mayoritariamente de las carreras de Lic. en Economía y en Turismo y en menor medida Lic. en Administración, con una fuerte presencia de las escuelas privadas y de titulación bachiller con la más baja cantidad de años transcurridos desde el egreso.

Conclusiones

Los procesos decisivos de alto nivel de las IES requieren de una etapa de diagnóstico que aporte información sustantiva para la formulación de políticas o el planeamiento estratégico. Esta información sustantiva comprende la identificación de características o comportamientos tácitos o no evidentes que enriquezcan y faciliten la comprensión del escenario donde se desarrollan dichas instituciones.

La interpretación del análisis realizado permite identificar agrupamientos de aspirantes con dificultades en la superación de las pruebas de admisión. En estos agrupamientos se evidencian algunas titulaciones como más determinantes que otras según el sexo y la carrera elegida. En otros agrupamientos la procedencia unida a una titulación técnica o comercial y con uno o más años transcurridos desde el egreso no resulta favorable para la aprobación de las pruebas.

En el último agrupamiento – con una amplia cantidad de observaciones – si bien han terminado recientemente sus estudios secundarios o polimodales de escuelas mayoritariamente privadas y se vuelcan hacia las licenciaturas de titulación bachiller no logra superar los exámenes previstos. Este grupo resulta de interés para un análisis posterior, pues presenta características que tradicionalmente han sido contributivas en la admisión de aspirantes.

Surge de la interpretación de los resultados del análisis que las políticas de la unidad académica deberán estar orientadas hacia los procesos de articulación con las

instituciones de Enseñanza Media, tal como se encuentran en este momento, propiciando el interés de los alumnos por la continuidad de sus estudios en un futuro próximo y facilitando los entornos de aprendizaje necesarios para superar las dificultades detectadas en la aprobación de las pruebas de admisión enfocados en algunas procedencias y titulación específicas.

La utilización de la herramienta propuesta, de amplia difusión en el ámbito empresarial, posibilita un avance en la identificación de características comunes y de variables relevantes del ingreso universitario, en este contexto se presenta a esta herramienta como un instrumento de importancia para el bagaje de herramientas de los procesos decisivos en las instituciones de educación superior.

Bibliografía

KASS, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29.

GARCÍA de FANELLI, Ana M. (2005). Indicadores y estrategias en relación con la graduación y el abandono universitario. In: La agenda universitaria: propuestas de políticas públicas para la Argentina. Carlos Marquís, comp. Escuela de Educación Superior. Universidad de

Palermo.

KAUFMAN, Leonard; ROUSSEEUW, Peter J. (1990). *Finding Groups in Data: an Introduction to Cluster Analysis*. New York: John Wiley & Sons.

LUAN, Jing. (2001). "Data Mining as Driven by Knowledge Management in Higher Education. Persistence Clustering And Prediction". [en línea]. In: *Keynote for SPSS Public Conference, UCSF*. <http://www.cabrillo.edu/services/pro/oir_reports/UCSFpaper.pdf> [Consulta: 21 sep. 2006].

PENELAS, Azucena (1999). "Aplicación del análisis cluster". En Hair, Joseph; Anderson, Rolph; Tatham, Ronald; Black, William. *Análisis multivariante*. pp. 533-536. 5a. ed. Madrid: Prentice Hall.

SANTESMASES MESTRE, Miguel. (1997). *DYANE. Diseño y análisis de encuestas en investigación social y de mercados*. Madrid: Pirámide.

SPSS (2001). "The SPSS TwoStep Cluster Component: A Scalable Component Enabling More Efficient Customer Segmentation" [en línea]. In: *White Paper: Technical Report*. <<http://www.spss.com/downloads/Papers.cfm?List=all&Name=all>>. [Consulta: 01 jun. 2006], disponible bajo registro.

Figuras

Carrera

	CP		LA		LE		LT	
	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
Cluster 1	159	33,7%	0	,0%	0	,0%	0	,0%
2	138	29,2%	63	30,6%	8	16,7%	68	33,5%
3	131	27,8%	71	34,5%	8	16,7%	4	2,0%
4	44	9,3%	72	35,0%	32	66,7%	131	64,5%
Combined	472	100,0%	206	100,0%	48	100,0%	203	100,0%

Figura 3

Sexo

	Masculino		Femenino	
	Frequency	Percent	Frequency	Percent
Cluster 1	0	,0%	159	30,0%
2	115	28,8%	162	30,6%
3	196	49,1%	18	3,4%
4	88	22,1%	191	36,0%
Combined	399	100,0%	530	100,0%

Figura 4

Tipo Sec./Pol.

		Cluster				
		1	2	3	4	Combined
Nacional	Frequency	25	29	30	30	114
	Percent	21,9%	25,4%	26,3%	26,3%	100,0%
Provincial	Frequency	95	134	103	76	408
	Percent	23,3%	32,8%	25,2%	18,6%	100,0%
Municipal	Frequency	0	6	36	0	42
	Percent	,0%	14,3%	85,7%	,0%	100,0%
Privado no relig.	Frequency	36	55	45	59	195
	Percent	18,5%	28,2%	23,1%	30,3%	100,0%
Privado relig.	Frequency	0	50	0	114	164
	Percent	,0%	30,5%	,0%	69,5%	100,0%
Instituto militar	Frequency	0	1	0	0	1
	Percent	,0%	100,0%	,0%	,0%	100,0%
Extranjero	Frequency	2	0	0	0	2
	Percent	100,0%	,0%	,0%	,0%	100,0%
Otros	Frequency	1	2	0	0	3
	Percent	33,3%	66,7%	,0%	,0%	100,0%

Figura 5

Título Sec./Pol.

	Bachiller		Perito mercantil		Técnico		Otros	
	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
Cluster 1	159	24,6%	0	,0%	0	,0%	0	,0%
2	4	,6%	61	98,4%	42	87,5%	170	98,3%
3	204	31,6%	1	1,6%	6	12,5%	3	1,7%
4	279	43,2%	0	,0%	0	,0%	0	,0%
Combined	646	100,0%	62	100,0%	48	100,0%	173	100,0%

Figura 6

Centroids

	Años egreso	
	Mean	Std. Deviation
Cluster 1	2,06	3,249
2	2,63	4,575
3	1,85	2,579
4	1,25	2,054
Combined	1,94	3,332

Figura 7