MODELOS ESTADÍSTICOS Y CONEXIONISTAS PARA PREDECIR EL RENDIMIENTO ACADÉMICO DE ALUMNOS UNIVERSITARIOS

MARÍA V. LÓPEZ , MARÍA G. LONGONI, EDUARDO A. PORCEL
Facultad de Ciencias Exactas y Naturales y Agrimensura
Universidad Nacional del Nordeste - ARGENTINA
mvlopez@exa.unne.edu.ar - eporcel@exa.unne.edu.ar - magalo82@hotmail.com

Fecha Recepción: Marzo 2011 - Fecha Aceptación: Agosto 2012

RESUMEN

En este trabajo se analiza la relación del rendimiento académico de los alumnos ingresantes a las carreras de perfil profesional la FACENA – UNNE en Corrientes, Argentina, durante el primer año, con sus características socioeducativas. El rendimiento fue medido por la aprobación de los exámenes parciales de las asignaturas del primer cuatrimestre del primer año. Se ajustaron un modelo de Regresión Logística Multinomial (RLM) y dos modelos de redes neuronales de tipo Perceptrón Multicapa (PM) y de Función de Base Radial (FBR) a dos conjuntos de datos: a) alumnos ingresantes a Bioquímica, cuyos plan de estudios incluye dos asignaturas en el primer cuatrimestre del primer año; b) alumnos ingresantes a carreras cuyos planes de estudios incluyen tres asignaturas en el primer cuatrimestre del primer año.

En ambos casos el modelo PM produjo el mejor ajuste, observándose que en el caso b) las tres técnicas utilizadas registraron altos porcentajes de clasificación correcta. Los resultados obtenidos contribuyen a orientar las políticas y estrategias institucionales para mejorar los preocupantes índices de desgranamiento, abandono y bajo rendimiento de los estudiantes en el primer año de universidad.

PALABRAS CLAVE: Rendimiento académico - Ingresantes universitarios - Regresión Logística multinomial - Redes neuronales - Perceptrón multicapa - Función de Base Radial.

ABSTRACT

This paper analyzes the relationship between the academic performance of students entering professional profile's careers in the FACENA - UNNE in Corrientes, Argentina, during the first year, and their social-educational characteristics.

Performance was measured by the approval of the partial evaluation of the subjects in the first semester of the first year. A model of Multinomial Logistic Regression (MLR) and two models of neural networks of type Multilayer Perceptron (MP) and Radial Basis Function (RBF) were fitted to two data sets: a) students entering in Biochemistry, whose curriculum includes two subjects in the first semester of the first year, b) students entering careers whose curriculum includes three subjects in the first semester of the first year.

In both cases, the PM model produced the best fit, and besides it was observed that in the case b) the three techniques showed high percentages of correct classification. The obtained results contribute to guide policies and strategies to improve the worrying levels of dropout and low performance of students in the first year of college.

KEY WORDS: Academic Performance - University freshmen - Multinomial Logistic Regression - Neural Networks - Multilayer perceptron - Radial Basis Function.

1. INTRODUCCIÓN

A partir de la década del ‘80 surge en las universidades de todo el mundo la preocupación por la calidad del servicio educativo que prestan. Esto dio lugar a procesos de evaluación a fin de detectar las debilidades y fortalezas institucionales y generar acciones correctivas de las deficiencias encontradas.

En Argentina, en la década del ’90, el Estado Nacional incluye en su agenda de política educativa la evaluación de la calidad del accionar universitario, y la mayoría de las universidades nacionales inician procesos de evaluación institucional. En 1996, se conocieron los primeros resultados referidos al rendimiento académico de los estudiantes de las trece carreras que constituyen la oferta académica de la Facultad de Ciencias Exactas y Naturales y Agrimensura de la Universidad Nacional del Nordeste (FACENA – UNNE) en Corrientes (Argentina).

Dicha información hace referencia a elevados índices de desgranamiento en todos los años de estudios pero, fundamentalmente, al término del primer cuatrimestre del primer año. Asimismo, da cuentas de que el retraso promedio en el egreso de todas las carreras alcanza al 50% de la duración teórica de las mismas, llegando en algunas a superarlo (Comisión de Autoevaluación FACENA-UNNE, 1996).

Según diversos estudios, el rendimiento académico de los estudiantes se ve influenciado por la interacción de diversos factores, que están ligados a características socioeducativas y culturales, los cuales afectan de manera importante el desempeño de los mismos, ya que son determinantes en la preparación del alumno desde antes de su entrada al sistema educativo y durante toda su trayectoria académica.

Por lo tanto, identificar estos factores y analizar conjuntamente su influencia en el rendimiento académico de los alumnos resulta una estrategia interesante de llevar a cabo, para lograr la identificación temprana de elementos de riesgo, y permitir así la realización oportuna de acciones correctivas en el proceso educativo.

Muchos autores están estudiando las relaciones entre las técnicas estadísticas convencionales y los modelos conexionistas, como herramientas de clasificación (Cherkassky et al, 1994; Flexer, 1995; Michie et al, 1994; Ripley, 1996; Sarle, 1994).

Este trabajo tiene los siguientes objetivos:

a) Desarrollar e implementar modelos que permitan predecir el rendimiento académico de los alumnos de primer año de las carreras de perfil profesional (Licenciatura en Sistemas de Información, Ingeniería en Electrónica, Ingeniería Eléctrica, Bioquímica y Agrimensura) de la Facultad de Ciencias Exactas y Naturales y Agrimensura de la UNNE, en base a los datos socioeducativos disponibles de los mismos.

b) Contrastar el rendimiento de las redes neuronales (redes de tipo perceptrón muticapa y de base radial) con modelos estadísticos convencionales (regresión logística multinomial) en problemas de clasificación de una variable cualitativa de tres categorías.

2. TÉCNICAS ESTADÍSTICAS Y CONEXIONISTAS. ANTECEDENTES

2.1. Regresión Logística Multinomial (RLM)

La distribución multinomial surge cuando una variable de respuesta es categórica por naturaleza, es decir, consiste en datos que describen la pertenencia de los casos respectivos a una categoría en particular (Agresti, 1996). La distribución multinomial es una generalización de la distribución binomial a más de dos categorías.

La **Regresión Logística Multinomial (RLM)** es un método estadístico clásico, que permite efectuar el análisis de los posibles efectos de variables independientes sobre una variable categórica dependiente con tres o más categorías.

El modelo de regresión logit multinomial es una extensión del modelo de regresión logit estándar para los casos donde la variable dependiente tiene más de dos categorías, es decir, cuando la variable dependiente o de respuesta de interés sigue una distribución multinomial en vez de una distribución binomial.

Hosmer et al. (2000), expresan que el modelo de regresión logística usa variables binarias, y es parametrizado en términos del logit de y = 1 versus y = 0. Esta idea puede extenderse el modelo a tres categorías de resultados, para lo cual se necesitan dos funciones logit. Si se usa y = 2 como referencia, y para formar los logit se comparan y = 0 e y = 1 con respecto a éste, las dos funciones logit se expresan como sigue, asumiendo que se tienen p variables:

$$g_1(x) = \ln\left[\frac{P(y=0|x)}{P(y=2|x)}\right] = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \ldots + \beta_{1p}x_p$$

$$g_2(x) = \ln\left[\frac{P(y=1|x)}{P(y=2|x)}\right] = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \ldots + \beta_{2p}x_p$$

Las probabilidades condicionales o de pertenencia a cada categoría de resultado, dado el vector de variables x, son las siguientes:

$$p_0 = P(y=0 \mid x) = \frac{e^{g_1(x)}}{1 + e^{g_1(x)} + e^{g_2(x)}}$$

$$p_1 = P(y=1 \mid x) = \frac{e^{g_2(x)}}{1 + e^{g_1(x)} + e^{g_2(x)}}$$

$$p_2 = P(y=2 \mid x) = 1 - p_0 - p_1 = \frac{1}{1 + e^{g_1(x)} + e^{g_2(x)}}$$

2.2. Redes Neuronales Artificiales (RNA)

Las **redes neuronales artificiales (RNA)** cuentan con el potencial para lograr los objetivos que se proponen en este trabajo, gracias a su excelente comportamiento en problemas de predicción y clasificación. De esta forma, se podrá concluir sobre el rendimiento futuro de los estudiantes teniendo información de entrada de diversos factores.

Una red neuronal es un sistema informático reticular (de inspiración neuronal) que aprende de la experiencia mediante la auto-modificación de sus conexiones (Hectht-Nielsen, 1990; Hertz et al, 1991; Wasserman, 1989; Hilera y Martínez, 1995; Martín y Sanz, 1997).

Las RNA son herramientas que permiten analizar los datos con el objeto de descubrir y modelar las relaciones funcionales existentes entre las variables. Permiten explorar relaciones o modelos que no podrían ser descubiertos usando procedimientos estadísticos más tradicionales (Rzempoluck, 1997). Se encuentran dentro de los métodos inteligentes, y tienen ventajas sobre los métodos estadísticos cuando se las aplica a situaciones en donde los datos de entrada son incompletos o ambiguos por naturaleza. Tampoco dependen de relaciones funcionales particulares, y no requieren una comprensión a priori de las relaciones entre variables. Asimismo, se caracterizan por su buen rendimiento ante problemas no lineales o datos con mucho "ruido", y presentan la ventaja de poder utilizarse independientemente del cumplimiento de los supuestos teóricos relativos a las técnicas estadísticas, debido a que para su uso no es necesario formular previamente una hipótesis, ya que desentrañan la información implícita en los datos. Hoy en día, las RNA son aplicadas a problemas de índole estadística como lo son la predicción y clasificación (Pitarque et al, 2000).

A modo de antecedentes de aplicación de la técnica de redes neuronales en el ámbito de educación pueden mencionarse los trabajos de González (1999), Salgueiro et al (2006), Borracci y Arribalzaga (2005), Zamarripa Topete et al., Santín González (1999), entre otros.

Las RNA de tipo **Perceptrón Multicapa (PM)** se encuentran entre las arquitecturas de red más poderosas y populares. Están formadas por una capa de entrada, un número arbitrario de capas ocultas, y una capa de salida. Cada una de las neuronas ocultas o de salida recibe una entrada de las neuronas de la capa previa (conexiones hacia atrás), pero no existen conexiones laterales entre las neuronas dentro de cada capa (Castillo et al, 1999).

La capa de entrada contiene tantas neuronas como categorías correspondan a las variables independientes que se desean representar. La capa de salida corresponde a la variable respuesta, que en este caso es una variable categórica.

Las RNA de **Función de Base Radial (FBR)** son aquellas cuyas funciones de activación en los nodos ocultos son radialmente simétricas. Se dice que una función es radialmente simétrica (o es una Función de Base Radial, FBR) si su salida depende de la distancia entre un vector que almacena los datos de entrada y un vector de pesos sinápticos, que recibe el nombre de centro o centroide (Vélez-Langs et al, 2007).

Fueron usadas por primera vez por (Broomhead et al, 1988), y pueden encontrarse contribuciones a su teoría, diseño y aplicaciones en los trabajos de (Moody et al, 1989) y (Poggio et al, 1990).

Las redes FBR presentan tres capas de conexión hacia adelante: la capa de entrada, la capa oculta o intermedia y la capa de salida. Las neuronas de la capa de entrada simplemente envían la información a la capa intermedia. Las neuronas de la capa oculta se activan en función de la distancia que separa cada patrón de entrada con respecto al centroide que cada neurona oculta almacena, a la que se le aplica una función radial con forma gaussiana. Las neuronas de la capa de salida son lineales, y simplemente calculan la suma ponderada de las salidas que proporciona la capa oculta.

3. METODOLOGÍA

3.1. Datos

La población analizada consiste en 1614 alumnos ingresantes a las carreras profesionales de la FACENA-UNNE en los años 2004 y 2005. Los datos de sus características socioeducativas se obtuvieron del formulario de ingreso a la universidad, mientras que los correspondientes a su desempeño académico, se obtuvieron del sistema informático de gestión de alumnos de la facultad.

Esta información se incorpora periódicamente en un único almacén de datos con un diseño orientado a las decisiones. Este proceso incluye la integración, depuración y formateo de los datos, siguiendo las técnicas usuales de preprocesado, constituyentes de las etapas previas al modelado y análisis de los datos (Dapozo et al, 2005) (Dapozo et al, 2007).

Dentro de las carreras profesionales en la oferta académica de la FACENA-UNNE, existen:

a) Una carrera cuyo plan de estudios incluye dos asignaturas en el primer cuatrimestre del primer año (Bioquímica), a la cual ingresaron 490 alumnos.

b) Cuatro carreras cuyos planes de estudios incluyen tres asignaturas en dicho cuatrimestre (Agrimensura, Ingeniería Eléctrica, Ingeniería en Electrónica y Licenciatura en Sistemas de Información), a las cuales ingresaron 1124 alumnos.

Todas estas carreras tienen en el primer cuatrimestre del primer año una materia con contenidos matemáticos (principalmente Álgebra).

Para poder avanzar en las materias del segundo cuatrimestre del primer año, el esquema de correlatividades de los planes de estudios requiere, como mínimo, tener aprobados los exámenes parciales de esta asignatura (situación que se denomina "regularizar" la asignatura), hecho que configura un fuerte condicionamiento de dicho avance.

Por tal motivo, a los fines de los objetivos del presente estudio, se han diseñado dos modelos diferentes para los casos a) y b) mencionados anteriormente:

a) En el caso a, el rendimiento académico se midió mediante una variable ***y*** que toma los siguientes valores:

   – 0 (cero), si el alumno no regularizó ninguna asignatura en el primer cuatrimestre (rendimiento malo).
   – 1 (uno), si el alumno regularizó una asignatura diferente de Álgebra en el primer cuatrimestre (rendimiento regular).
   – 2 (dos), si el alumno regularizó Álgebra en el primer cuatrimestre ó si regularizó las dos asignaturas del primer cuatrimestre (rendimiento bueno).

b) En el caso b, el rendimiento académico se midió mediante una variable ***y*** que toma los siguientes valores:

   – 0 (cero), si el alumno no regularizó ninguna asignatura en el primer cuatrimestre (rendimiento malo).
   – 1 (uno), si el alumno regularizó una asignatura en el primer cuatrimestre ó si regularizó dos asignaturas diferentes de Álgebra (rendimiento regular).
   – 2 (dos), si el alumno regularizó dos asignaturas siendo Álgebra una de ellas ó si regularizó las tres asignaturas del primer cuatrimestre (rendimiento bueno).

A continuación se enuncian las variables socieducativas (independientes) que se incluyeron en ambos modelos, y las categorías que asumen, indicándose en negrita (para el modelo de regresión logística), la categoría de referencia.

a) AÑO DE INGRESO: 2004, **2005.**
b) CARRERA: En el modelo diseñado para el caso a) la carrera es **Bioquímica**. En el modelo descripto para el caso b) las carreras son: Licenciatura en Sistemas de Información; Agrimensura; Ingeniería Eléctrica; **Ingeniería en Electrónica**.
c) SEXO: Varón; **Mujer**.
d) TIENE MAIL: No; **Sí**.

e) TITULO SECUNDARIO: Economía y Gestión de las Organizaciones; Humanidades y Ciencias Sociales; Comunicación, Arte y Diseño; Producción de bienes y servicios; Bachiller común; Peritos Mercantiles; Técnicos; Otros títulos; **Ciencias Naturales**.
f) DEPENDENCIA DEL ESTABLECIMIENTO: Nacional; Provincial; Dependiente de la Universidad; Privados religiosos; Privados particulares; **Institutos militares.**;
g) COBERTURA OBRA SOCIAL: De los padres; Del cónyuge; Propia; **Ninguna**.
h) ESTUDIO DE LOS PADRES: Se consideró el mayor nivel educativo alcanzado por el padre o la madre. Las categorías son: No hizo estudios/Escuela Primaria Incompleta; Escuela Primaria Completa/ Escuela Secundaria Incompleta; Escuela Secundaria Completa/Estudio Superior No Universitario Incompleto; Estudio Superior No Universitario Completo/Estudio Universitario Incompleto; **Estudio Universitario Completo/Estudios de Posgrado**.

Las variables relacionadas a la actividad laboral del alumno y de los padres no pudieron ser incluidas en el modelo debido a la notable falta de respuesta registrada en los formularios.

3.2. Modelo de regresión logística multinomial para predecir el rendimiento académico

Se formularon dos modelos de regresión logística multinomial de efectos principales, y se ajustaron con los conjuntos de datos descriptos en la sección 3.1.

Dado un vector ***x*** de variables socioeducativas de los alumnos, siendo ***y*** la variable dependiente (rendimiento académico) que asume las tres categorías descriptas en 3.1, el modelo de regresión logística multinomial intenta modelar las probabilidades de que dicha variable *y* sea igual a 0, 1 ó 2, dados los valores del vector de variables ***x***.

3.3. Modelos de redes neuronales de tipo perceptrón multicapa y base radial para predecir el rendimiento académico

Se entrenaron modelos de redes neuronales de tipo PM y FBR con los conjuntos de datos descriptos en la sección 3.1.

En el entrenamiento de las redes, se presentó un conjunto de patrones de entrada, constituido por las variables que definen el perfil socioeducativo enumeradas precedentemente, y su correspondiente valor de salida (rendimiento académico) esperado.

Para el entrenamiento de las redes PM, se emplearon los algoritmos de aprendizaje supervisado de Retropropagación (BackPropagation - BP) (Patterson, 1996) (Fausett, 1994) (Haykin, 1994) y Gradiente descendente (Conjugate Gradient Descent - CG) (Bishop, 1995) (Shepherd, 1997).

El entrenamiento de las redes FBR comprende dos fases: una no supervisada y otra supervisada. En la fase no supervisada se empleó el algoritmo K-medias para la determinación de centros, y el método del vecino más cercano para el cálculo de las amplitudes de las neuronas de la capa oculta (Vélez-Langs y Staffetti, 2007) (Palacios Burgos, 2003) (Lanzarini, 2003). La fase supervisada consiste en la determinación de los pesos y umbrales en la capa de salida. Para minimizar la diferencia entre las salidas de la red y las salidas deseadas, se utilizó el método de los mínimos cuadrados (Vélez-Langs y Staffetti, 2007).

Para efectuar la validación, se utilizó en cada caso una lista de mediciones independientes de los datos para todas las variables a fin de determinar el grado de predicción de cada modelo. Este conjunto constituyó el 25 % del total de los datos y fue seleccionado al azar.

Debido a que se trata de un problema de clasificación, el objetivo de la red es el de asignar a cada caso, una de tres clases (Bueno, Regular o Malo), estimando la probabilidad de pertenencia del caso a cada clase.

Con respecto a la determinación del nivel de corte o umbral de clasificación, en este trabajo se indicó que no se utilizarían umbrales, por lo que la red utilizó el algoritmo “winner takes all”. En este algoritmo, la neurona de mayor activación determina la clase, y no hay “opción de duda".

Finalmente, se determinó la importancia de las variables de entrada, mediante un análisis de sensibilidad, el cual cuantifica el porcentaje de contribución de cada variable de entrada a la variable respuesta en el modelo, permitiendo diferenciar las variables socioeducativas que tienen una influencia estadísticamente significativa de aquellas que no difieren significativamente del azar. El análisis se llevó a cabo tratando a cada una de las variables de entrada por vez, como si estuviese "no disponible" (Hunter et al, 2000). Una vez que se calcularon las sensibilidades para todas las variables, éstas fueron clasificadas en orden. El análisis de sensibilidad permitió obtener pistas importantes sobre la utilidad de las variables individuales, identificando las variables que podrían ser ignoradas en los análisis posteriores, y aquellas variables clave que siempre deberían mantenerse. Este análisis valora cada variable de acuerdo con el deterioro en el rendimiento del modelo que se produce si esa variable ya no está disponible para el mismo, asignando un valor de calificación o ranking único a cada variable.

4. RESULTADOS

Para el modelo de regresión logística multinomial ajustado para el caso a) descripto en 3.1, las estimaciones de los coeficientes β de los $g_i(x)$ se muestran en la Tabla 1, siendo su expresión la siguiente:

$g_1(x)$ = -1.298 + 0.846 AÑO + 0.228 SEXO + 0.231 MAIL + 0.916 TECON + 1.802 THUM + … + 0.270 MSCUI

$g_2(x)$ = -17.790 + 0.995AÑO + 0.245 SEXO + 0.234 MAIL + 18.104 TECON + 18.261 THUM + … -0.756 MSCUI

Para el modelo de regresión logística multinomial ajustado para el caso b) descripto en 3.1., las estimaciones de los coeficientes β de los $g_i(x)$ se muestran en la Tabla 2, siendo su expresión la siguiente:

$g_1(x)$ = 33.864 + 1.208 AÑO - 0.109 CLSI - 0.437 CAG + 0.221 CIE + 0.389 SEXO + 0.315 MAIL + … + 0.656 MSCUI

$g_2(x)$ = 1.490 - 1.148 AÑO + 1.583 CLSI + 0.525 CAG + 0.206 CIE + 0070 SEXO – 0.025 MAIL + … - 0.209 MSCUI

En cuanto a los modelos de redes neuronales, la arquitectura que presentan los modelos PM y FBR entrenados tienen la forma I:N-N-N:O, donde I es el número de variables de entrada, O es el número de variables de salida, y N es el número de unidades en cada capa (Figuras 1 a 4).

A partir de las matrices de clasificación se puede concluir acerca de la eficacia predictiva de todos los modelos ajustados.

Para los modelos correspondientes a las carreras que poseen dos asignaturas en el primer cuatrimestre de primer año, se obtuvieron, en el caso de la RLM un porcentaje de clasificación correcta total de 58,8 %, para el modelo PM de 72,4 %, y para el modelo FBR de 58,9% (Tablas 3 a 5).

Para los modelos correspondientes a las carreras que poseen tres asignaturas en el primer cuatrimestre de primer año, se obtuvieron, en el caso de la RLM un porcentaje de clasificación correcta total de 71,4 %, para el modelo PM de 80,5 %, y para el modelo FBR de 74,7 % (Tablas 6 a 8).

Pudo apreciarse que el modelo de red PM predijo adecuadamente tanto el rendimiento malo (0) como el rendimiento regular (1) y el bueno (2), situación que no ocurre con el modelo de red FBR ni con el modelo RLM. En el caso del modelo de red FBR, se observó que éste clasifica mejor a los alumnos de rendimiento malo, que a los de rendimiento bueno y regular.

En el caso del modelo RLM, éste clasifica mejor a los alumnos de rendimiento bueno, que a los malos y regulares.

Se observó que las tres técnicas (RLM y redes FBR y PM) tuvieron un comportamiento similar, tanto para los modelos correspondientes a las carreras que poseen tres asignaturas en el primer cuatrimestre de primer año, como para los modelos correspondientes a las carreras que poseen dos asignaturas en el primer cuatrimestre de primer año.

En la carrera de Bioquímica (carrera con dos asignaturas en el primer cuatrimestre de primer año), el análisis de sensibilidad mostró que la variable más importante para ambos modelos de redes fue el título secundario, la cual también resultó estadísticamente significativa en el modelo de regresión logística multinomial. Los órdenes obtenidos para las restantes variables fueron poco concordantes en ambos tipos de redes. El método RLM registró también como significativas al estudio de los padres y el año de ingreso.

En las carreras que tienen tres asignaturas en el primer cuatrimestre de primer año, ambos modelos de redes registraron como las variables más importantes al año de ingreso, el estudio de los padres y el título secundario obtenido. Estas tres variables también resultaron estadísticamente significativas en el modelo de regresión logística multinomial, concordando además con lo obtenido en la carrera de Bioquímica por el modelo RLM. Asimismo, se observó que las restantes variables obtuvieron órdenes de importancia muy similares en ambos modelos de redes, a diferencia de lo ocurrido para Bioquímica. Esta situación podría quizás deberse al distinto número de registros de datos usados en cada caso.

La variable tenencia de mail resultó poco relevante en todos los modelos ajustados, y la variable sexo sólo resultó significativa en la red de tipo perceptrón multicapa, para la carrera de Bioquímica (Tabla 9).

5. CONCLUSIONES

Se concluye que la técnica de RNA ha mostrado en general una buena capacidad clasificatoria, mediante los modelos PM y FBR orientados a la predicción del rendimiento académico de los alumnos ingresantes en función de sus características socioeducativas.

En particular, los ajustes ofrecidos por las redes PM, son superiores a los obtenidos con las redes FBR y el modelo de RLM, aplicados a los mismos conjuntos de datos.

Con respecto a los diferentes modelos entrenados, se ha observado que el modelo diseñado para las carreras con tres asignaturas en el primer cuatrimestre de primer año ha permitido obtener mejores porcentajes de clasificación que el modelo diseñado para la carrera con dos asignaturas en el primer cuatrimestre del primer año, lo que podría atribuirse al número de registros de datos utilizados en cada caso.

A futuro se profundizará este estudio con otras metodologías de análisis de datos, tales como árboles de clasificación, y se realizarán comparaciones entre las diferentes técnicas.

De este modo, estos modelos orientados al rendimiento académico han permitido obtener información precisa sobre la controversia planteada en relación a qué técnicas (estadísticas o neuronales) son más eficientes en la solución de problemas de clasificación.

A nivel de la gestión de la educación superior, los modelos desarrollados en este trabajo, mediante la aplicación de técnicas estadísticas y conexionistas, permiten obtener información acerca del rendimiento académico de los alumnos ingresantes a las carreras de perfil profesional de la FACENA-UNNE al finalizar el primer cuatrimestre del primer año de estudios, contribuyendo de este modo a orientar las políticas y estrategias institucionales para mejorar los preocupantes índices de desgranamiento, abandono y bajo rendimiento.

Tabla 1: Modelo de Regresión Logística Multinomial para la carrera con dos asignaturas en el primer cuatrimestre de primer año. Estimaciones de los parámetros

Rendimiento (y) (*)	Variables (x)	β
0	Constante	-1.298
	AÑO DE INGRESO (AÑO)	0.846
	SEXO (SEXO)	0.228
	TIENE MAIL (MAIL)	0.231
	TITULO SECUNDARIO	
	Econ. y Gestión de las Org. (TECON)	0.916
	Humanidades y Cs. Sociales (THUM)	1.802
	Comunicación, Arte y Diseño (TCOM)	1.024
	Producción de Bienes y Servicios (T_PROD)	19.553
	Bachiller Común (TBACH)	1.580
	Perito Mercantil (TPER)	1.106
	Técnicos (TTEC)	3.578
	Otros Títulos (TOTROS)	2.001
	DEPENDENCIA DEL ESTABLECIMIENTO SECUNDARIO	
	Nacional (DNAC)	0.179
	Provincial (DPROV)	0.190
	Dependiente de la Universidad (DUNIV)	20.199
	Privados religiosos (DPRIVR)	0.091
	COBERTURA OBRA SOCIAL	

Tabla 1: Modelo de Regresión Logística Multinomial para la carrera con dos asignaturas en el primer cuatrimestre de primer año. Estimaciones de los parámetros

	De los padres (OSPAD)	-0.324
	Del cónyuge (OSCONY)	-0.324
	Propia (OSPROP)	18.964
	ESTUDIO DE LOS PADRES	
	No hizo estudios/Escuela Primaria Incompleta (MNEPI)	0.712
	Primaria Completa / Secundaria Incompleta (MPCSI)	1.388
	Secundaria Completa / Superior No Univ. Incompleto (MSCSI)	0.762
	Superior No Univ. Completo / Universitario Incompleto (MSCUI)	0.270
1	Constante	-17.790
	AÑO DE INGRESO (AÑO)	0.995
	SEXO (SEXO)	0.245
	TIENE MAIL (MAIL)	0.234
	TITULO SECUNDARIO	
	Econ. y Gestión de las Org. (TECON)	18.104
	Humanidades y Cs. Sociales (THUM)	18.261
	Comunicación, Arte y Diseño (TCOM)	17.953
	Producción de Bienes y Servicios (T_PROD)	38.181
	Bachiller Común (TBACH)	18.679
	Perito Mercantil (TPER)	18.646
	Técnicos (TTEC)	20.367
	Otros Títulos (TOTROS)	18.715
	DEPENDENCIA DEL ESTABLECIMIENTO SECUNDARIO	
	Nacional (DNAC)	-0.229
	Provincial (DPROV)	0.229
	Dependiente de la Universidad (DUNIV)	1.132
	Privados religiosos (DPRIVR)	0.136
	COBERTURA OBRA SOCIAL	
	De los padres (OSPAD)	-1.330
	Del cónyuge (OSCONY)	-1.135
	Propia (OSPROP)	-1.159
	ESTUDIO DE LOS PADRES	
	No hizo estudios/Escuela Primaria Incompleta (MNEPI)	-19.042
	Primaria Completa / Secundaria Incompleta (MPCSI)	0.833
	Secundaria Completa / Superior No Univ. Incompleto (MSCSI)	-0.062
	Superior No Univ. Completo / Universitario Incompleto (MSCUI)	-0.756

(*) La categoría de referencia es la 2.

Tabla 2: Modelo de Regresión Logística Multinomial para las carreras con tres asignaturas en el primer cuatrimestre de primer año. Estimaciones de los parámetros

Rendimiento (y) (*)	Variables (x)	B
0	Constante	33.864
	AÑO DE INGRESO (AÑO)	1.208
	CARRERA	
	Licenciatura en Sistemas de Información (CLSI)	-0.109
	Agrimensura (CAG)	-0.437
	Ingeniería Eléctrica (CIE)	0.221
	SEXO (SEXO)	0.389
	TIENE MAIL (MAIL)	0.315
	TITULO SECUNDARIO	
	Econ. y Gestión de las Org. (TECON)	-17.133
	Humanidades y Cs. Sociales (THUM)	-16.272
	Comunicación, Arte y Diseño (TCOM)	-15.908
	Producción de Bienes y Servicios (T_PROD)	2.008
	Bachiller Común (TBACH)	-16.795
	Perito Mercantil (TPER)	-16.138
	Técnicos (TTEC)	-16.312
	Otros Títulos (TOTROS)	-16.963
	DEPENDENCIA DEL ESTABLECIMIENTO SECUNDARIO	
	Nacional (DNAC)	-16.404
	Provincial (DPROV)	-16.243
	Dependiente de la Universidad (DUNIV)	-36.722
	Privados religiosos (DPRIVR)	-16.280
	Privados particulares (DPRIVP)	-15.757
	COBERTURA OBRA SOCIAL	
	De los padres (OSPAD)	-0.972
	Del cónyuge (OSCONY)	-1.217
	Propia (OSPROP)	-1.633
	ESTUDIO DE LOS PADRES	
	No hizo estudios/Escuela Primaria Incompleta (MNEPI)	1.661
	Primaria Completa / Secundaria Incompleta (MPCSI)	1.118
	Secundaria Completa / Superior No Univ. Incompleto (MSCSI)	0.733
	Superior No Univ. Completo / Universitario Incompleto (MSCUI)	0.656
1	Constante	1.490
	AÑO DE INGRESO (AÑO)	-1.148
	CARRERA	
	Licenciatura en Sistemas de Información (CLSI)	1.583
	Agrimensura (CAG)	0.525
	Ingeniería Eléctrica (CIE)	0.206
	SEXO (SEXO)	0.070
	TIENE MAIL (MAIL)	-0.025
	TITULO SECUNDARIO	

Tabla 2: Modelo de Regresión Logística Multinomial para las carreras con tres asignaturas en el primer cuatrimestre de primer año. Estimaciones de los parámetros

	Econ. y Gestión de las Org. (TECON)	-0.599
	Humanidades y Cs. Sociales (THUM)	0.141
	Comunicación, Arte y Diseño (TCOM)	0.870
	Producción de Bienes y Servicios (T_PROD)	19.059
	Bachiller Común (TBACH)	0.194
	Perito Mercantil (TPER)	0.103
	Técnicos (TTEC)	0.239
	Otros Títulos (TOTROS)	-0.070
	DEPENDENCIA DEL ESTABLECIMIENTO SECUNDARIO	
	Nacional (DNAC)	-1.443
	Provincial (DPROV)	-1.207
	Dependiente de la Universidad (DUNIV)	-20.703
	Privados religiosos (DPRIVR)	-0.965
	Privados particulares (DPRIVP)	-1.037
	COBERTURA OBRA SOCIAL	
	De los padres (OSPAD)	-1.089
	Del cónyuge (OSCONY)	-1.265
	Propia (OSPROP)	-20.962
	ESTUDIO DE LOS PADRES	
	No hizo estudios/Escuela Primaria Incompleta (MNEPI)	0.593
	Primaria Completa / Secundaria Incompleta (MPCSI)	0.253
	Secundaria Completa / Superior No Univ. Incompleto (MSCSI)	0.045
	Superior No Univ. Completo / Universitario Incompleto (MSCUI)	-0.209

(*) La categoría de referencia es la 2.

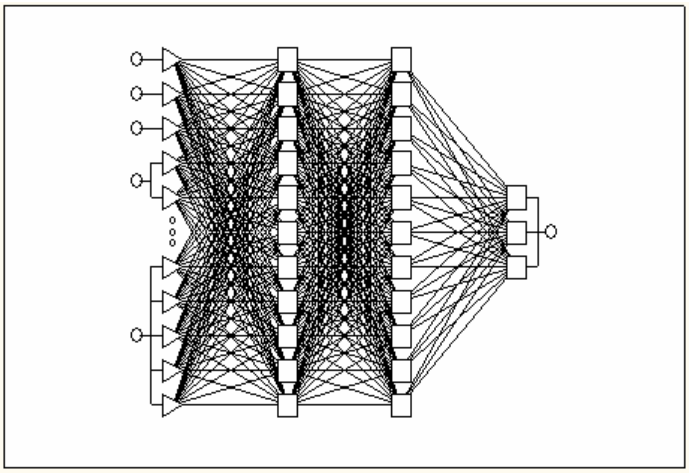

Figura 1. Arquitectura del modelo de red PM para la carrera con dos asignaturas en el primer cuatrimestre de primer año (7:26-11-11-3:1)

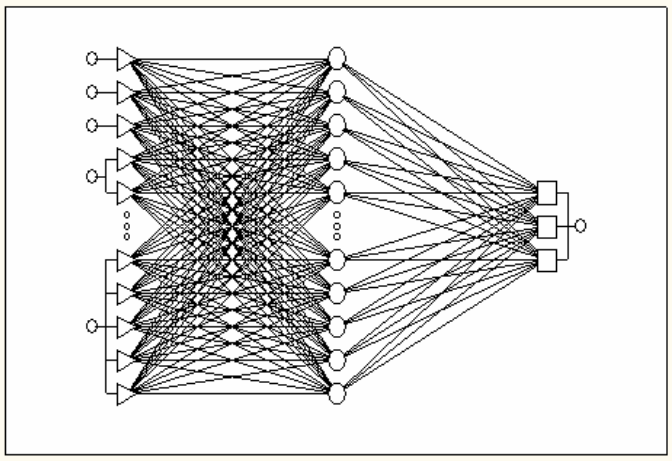

Figura 2. Arquitectura del modelo de red FBR para la carrera con dos asignaturas en el primer cuatrimestre de primer año (7:26-34-3:1)

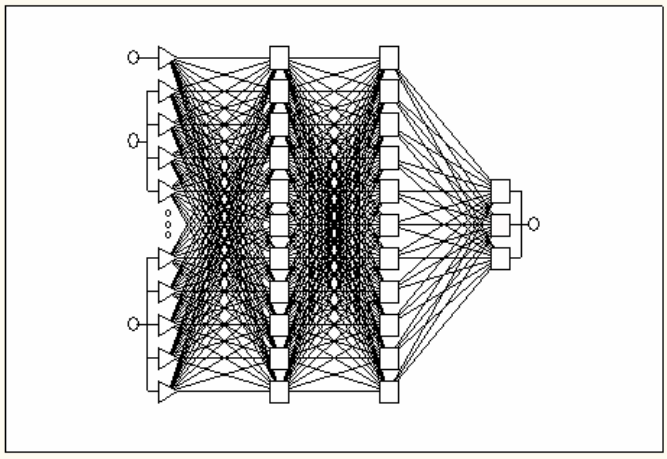

Figura 3. Arquitectura del modelo de red PM para las carreras con tres asignaturas en el primer cuatrimestre de primer año (8:31-11-11-3:1)

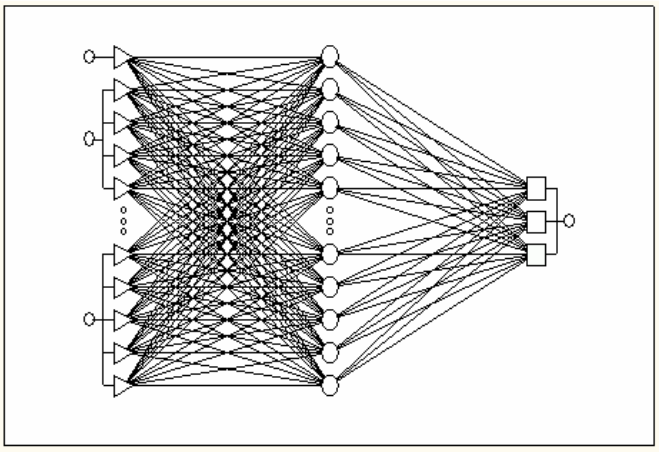

Figura 4. Arquitectura del modelo de red FBR para las carreras con tres asignaturas en el primer cuatrimestre de primer año (8:30-155-3:1)

Tabla 3: Modelo de regresión logística multinomial (RLM) para la carrera con dos asignaturas en el primer cuatrimestre de primer año. Matriz de clasificación

Clasificación	Rendimiento académico			
	Bueno (2)	Regular (1)	Malo (0)	Total
Bueno (2)	234	78	70	
Regular(1)	14	24	8	
Malo (0)	15	17	30	
Total	263	119	108	490
% de clasif. Correcto	88,97	20,17	27,78	58,78

Tabla 4: Modelo de red PM para la carrera con dos asignaturas en el primer cuatrimestre de primer año. Matriz de clasificación

Clasificación	Rendimiento académico			
	Bueno (2)	Regular (1)	Malo (0)	Total
Bueno (2)	78	14	34	
Regular(1)	11	75	27	
Malo (0)	19	30	202	
Total	108	119	263	490
% de clasif. Correcto	72,2	63	76,8	72,4

Tabla 5: Modelo de red FBR para la carrera con dos asignaturas en el primer cuatrimestre de primer año. Matriz de clasificación

Clasificación	Rendimiento académico			
	Bueno (2)	Regular (1)	Malo (0)	Total
Bueno (2)	27	7	15	
Regular(1)	13	34	20	
Malo (0)	68	78	228	
Total	108	119	263	490
% de clasif. Correcto	25	28,5	86,6	58,9

Tabla 6: Modelo de regresión logística multinomial (RLM) para las carreras con tres asignaturas en el primer cuatrimestre de primer año. Matriz de clasificación

Clasificación	Rendimiento académico			
	Bueno (2)	Regular (1)	Malo (0)	Total
Bueno (2)	730	144	110	
Regular(1)	43	64	20	
Malo (0)	0	4	9	
Total	773	212	139	1124
% de clasif. Correcto	94,4	30,1	6,4	71,4

Tabla 7: Modelo de red PM para las carreras con tres asignaturas en el primer cuatrimestre de primer año. Matriz de clasificación

Clasificación	Rendimiento académico			
	Bueno (2)	Regular (1)	Malo (0)	Total
Bueno (2)	87	22	39	
Regular(1)	9	124	40	
Malo (0)	43	66	694	
Total	139	212	773	1124
% de clasif. Correcto	62,5	58,4	89,7	80,5

Tabla 8: Modelo de red FBR para las carreras con tres asignaturas en el primer cuatrimestre de primer año. Matriz de clasificación

Clasificación	Rendimiento académico			
	Bueno (2)	Regular (1)	Malo (0)	Total
Bueno (2)	44	13	17	
Regular(1)	18	76	36	
Malo (0)	77	123	720	
Total	139	212	773	1124
% de clasif. Correcto	31,6	35,8	93,2	74,7

Tabla 9: Análisis de sensibilidad de los modelos PM, FBR y RLM para las carreras analizadas

Variables	Carreras con tres asignaturas en el primer cuatrimestre de primer año					Bioquímica (carrera con dos asignaturas en el primer cuatrimestre de primer año)				
	PM		PM		FBR	PM		FBR		RLM
	Relac.	Ord.	Relac.	Ord.	p	Relac.	Ord.	Relac.	Ord.	P
Titulo secundario	1,390	3	1,184	1	0,008	1,380	1	1,039	1	0,013
Sexo	1,102	8	1,021	8	0,143	1,350	2	1,006	7	0,646
Año de ingreso	1,520	1	1,093	3	0,000	1,281	3	1,009	5	0,001
Dependencia del establecimiento	1,260	5	1,078	6	0,453	1,269	4	1,010	4	0,936
Estudio de los padres	1,453	2	1,137	2	0,008	1,265	5	1,032	2	0,0003
Tenencia de mail	1,162	7	1,026	7	0,089	1,207	6	1,007	6	0,637
Cobertura Obra social	1,231	6	1,089	4	0,130	1,180	7	1,012	3	0,409
Carrera	1,339	4	1,084	5	0,000	-	-	-	-	-

6. REFERENCIAS

- AGRESTI, A. (1996): "AN INTRODUCTION TO CATEGORICAL DATA ANALYSIS". New York: Wilely.

- BISHOP, C. (1995): "NEURAL NETWORKS FOR PATTERN RECOGNITION". Oxford: University Press. En: LÉVY MANGIN, J.; VARELA MALLOU, J. (2003): "ANÁLISIS MULTIVARIABLE PARA LAS CIENCIAS SOCIALES". Pearson Educación S. A.

- BORRACCI, R. A. y ARRIBALZAGA, E. B. (2005): "APLICACIÓN DE ANÁLISIS DE CONGLOMERADOS Y REDES NEURONALES ARTIFICIALES PARA LA CLASIFICACIÓN Y SELECCIÓN DE CANDIDATOS A RESIDENCIAS MÉDICAS". Educación Médica Vol 8 Nº 1. ISSN 1575-1813. Barcelona.

- BROOMHEAD, D.S.; LOWE, D. (1988): "MULTIVARIABLE FUNCTIONAL INTERPOLATION AND ADAPTIVE NETWORK". Complex Systems, 2, 321–355.

- CASTILLO, E.; COBO, A.; GUTIÉRREZ, J.M.; PRUNEDA, R.E. (1999). Introducción a las Redes Funcionales con Aplicaciones. Un Nuevo Paradigma Neuronal". Editorial Paraninfo S.A. Madrid. España. pp.5-8; 8-16; 21-24, 30-34, 53-100.

- CHERKASSKY, V.; FRIEDMAN, J.H. Y WECHSLER, H. (1994): "FROM STATISTICS TO NEURAL NETWORKS". Springer- Verlag. Berlin.

- COMISIÓN DE AUTOEVALUACIÓN FACENA-UNNE. (1996): "INFORME DE AVANCE PRE-DIAGNÓSTICO 1996". Facultad de Ciencias Exactas y Naturales y Agrimensura. Universidad Nacional del Nordeste.

- DAPOZO, G.; PORCEL, E. (2005): "METODOLOGÍA DE INTEGRACIÓN DE DATOS PARA APOYAR EL SEGUIMIENTO Y ANÁLISIS DEL RENDIMIENTO ACADÉMICO DE LOS ALUMNOS DE LA FACENA". Comunicaciones Científicas y Tecnológicas de la Universidad Nacional del Nordeste 2005. Corrientes. Argentina. Disponible en: http://www.unne.edu.ar/Web/cyt/com2005/8-Exactas/E-032.pdf.

- DAPOZO, G.; PORCEL, E.; LÓPEZ, M. V.; BOGADO, V. (2007): "TÉCNICAS DE PREPROCESAMIENTO PARA MEJORAR LA CALIDAD DE LOS DATOS EN UN ESTUDIO DE CARACTERIZACIÓN DE INGRESANTES UNIVERSITARIOS". IX Workshop de Investigadores en Ciencias de la Computación (WICC 2007). Trelew. Chubut. Argentina.

- FAUSETT, L. (1994): "FUNDAMENTALS OF NEURAL NETWORKS". New York: Prentice Hall. En: LÉVY MANGIN, J.; VARELA MALLOU, J. (2003): "ANÁLISIS MULTIVARIABLE PARA LAS CIENCIAS SOCIALES". Pearson Educación S. A.

- FLEXER, A. (1995): "CONNECTIONIST AND STATISTICIANS, FRIENDS OR FOES?". The Austrian Research Institute for Artificial Intelligence. Acceso FTP. Servidor: ai.univie.ac.at.

- GONZÁLEZ, D.S. (1999): "DETECCIÓN DE ALUMNOS DE RIESGO Y MEDICIÓN DE LA EFICIENCIA DE CENTROS ESCOLARES MEDIANTE REDES NEURONALES". Biblioteca de Económicas y Empresariales. Servicios de Internet. Universidad Complutense de Madrid.

- HAYKIN, S. (1994): "NEURAL NETWORKS: A COMPREHENSIVE FOUNDATION". New York: Macmillan Publishing. En: Lévy Mangin, J.; Varela Mallou, J. (2003). Análisis multivariable para las Ciencias Sociales. Pearson Educación S. A.

- HECTHT-NIELSEN, R. (1990): "NEUROCOMPUTING". Addison-Wesley. Cal.

- HERTZ, J. KROGH, A. y PALMER, R. (1991): "INTRODUCTION TO THE THEORY OF NEURAL COMPUTATION". Addison-Wesley. Cal.

- HILERA, J.R. y MARTÍNEZ, V.J. (1995): "REDES NEURONALES ARTIFICIALES: FUNDAMENTOS, MODELOS Y APLICACIONES". Ra-ma. Madrid.

- HOSMER, D.; LEMESHOW, S. (2000): "APPLIED LOGISTIC REGRESSION. 2ND EDITION JOHN WILEY & SONS INC". En: GARCÍA, T.; MONTERO, C.; RUÍZ, V.; VÁSQUEZ, M.; ÁLVAREZ, W. (2008): "APLICACIÓN DE LA REGRESIÓN LOGÍSTICA MULTINOMIAL EN LA DETECCIÓN DE FACTORES ECONÓMICOS QUE INFLUYEN LA PRODUCTIVIDAD DE LOS SECTORES INDUSTRIALES". Ingeniería UC, Vol. 15, Núm. 3, pp. 19-24. ISSN 1316-6832. Universidad de Carabobo. Venezuela.

- HUNTER, A.; KENNEDY, L.; HENRY, J; FERGUSON, R.I. (2000): "APPLICATION OF NEURAL NETWORKS AND SENSITIVITY ANALYSIS TO IMPROVED PREDICTION OF TRAUMA SURVIVAL". Computer Methods and Algorithms in Biomedicine 62, 11-19.

- LANZARINI, L. (2003): "REDES NEURONALES DE BASE RADIAL (EJEMPLO K-MEDIAS)". Material didáctico. Cátedra "Redes Neuronales y algoritmos evolutivos". Instituto de Investigación en Informática LIDI. Facultad de Informática. Universidad Nacional de La Plata. Buenos Aires. Argentina. Fecha de consulta: Abril de 2010. Disponible en: http://weblidi.info.unlp.edu.ar/catedras/neuronales/05_RBF.pdf

- MARTÍN, B. y SANZ, A. (1997): "REDES NEURONALES Y SISTEMAS BORROSOS". Ra-ma. Madrid.

- MICHIE, D., SPIEGELHARTER, D.J. y TAYLOR, C.C. (1994): "MACHINE LEARNING, NEURAL AND STATISTICAL CLASSIFICATION". Londres: Ellis Horwood.

- MOODY, J.; DARKEN, C. (1989): "FAST LEARNING IN NETWORKS OF LOCALLY TUNED PROCESSING UNITS". Neural Computation, 1 (2), 281–294.

- PALACIOS BURGOS, F. J. (2003). "REDES NEURONALES CON GNU/LINUX COPYRIGHT (c)"- Herramientas en GNU/Linux para estudiantes universitarios- Capítulo 3. Tipos de Redes Neuronales. Fecha de consulta: Abril de 2010. Disponible en: http://softwarelibre.unsa.edu.ar/docs/descarga/2003/curso/htmls/redes_neuronales/x185.html

- PATTERSON, D. (1996): "ARTIFICIAL NEURAL NETWORKS". Singapore: Prentice Hall. En: Lévy Mangin, J.; Varela Mallou, J. (2003). Análisis multivariable para las Ciencias Sociales. Pearson Educación S. A.

- PITARQUE, A.; RUIZ, J. C.; ROY, J. F. (2000): "LAS REDES NEURONALES COMO HERRAMIENTAS ESTADÍSTICAS NO PARAMÉTRICAS DE CLASIFICACIÓN". Psicothema ISSN 0214 - 9915 CODEN PSOTEG. Vol. 12, Supl. nº 2, pp. 459-463. Disponible en: http://www.psicothema.com/psicothema.asp?id=604. 2000.

- POGGIO, T.; GIROSI, F. (1990): "NETWORK FOR APPROXIMATION AND LEARNING". Proceedings of IEEE, 78 (9), 1491–1497.

- RIPLEY, B.D. (1996). "PATTERN RECOGNITION AND NEURAL NETWORKS". Cambridge Univ. Press. Cambridge, G.B.

- RZEMPOLUCK, E. J. (1997): "NEURAL NETWORK DATA ANALYSIS USING SIMULNET". Simon Fraser University. Burnaby. B.C. Canadá. ISBN: 0-387-98255-8. pp. 1-3, 13-75.

- SALGUEIRO, F.; COSTA, G.; CÁNEPA, S.; LAGE, F.; KRAUS, G.; FIGUEROA, N.; CATALDI, Z. (2006): "REDES NEURONALES PARA PREDECIR LA APTITUD DEL ALUMNO Y SUGERIR ACCIONES". Workshop de Investigadores en Ciencias de la Computación 2006.

- SANTÍN GONZÁLEZ, D. (1999): "DETECCIÓN DE ALUMNOS DE RIESGO Y MEDICIÓN DE LA EFICIENCIA DE CENTROS ESCOLARES MEDIANTE REDES NEURONALES". Disponible en: http://eprints.ucm.es/6674/1/9902.pdf.

- SARLE, W.S. (1994): "NEURAL NETWORKS AND STATISTICAL MODELS". Proceedings of the 19th Annual SAS Group conference, Cary, NC. pps. 1538-1550.

- SHEPHERD, A. J. (1997): "SECOND-ORDER METHODS FOR NEURAL NETWORKS". New York: Springer. En: Lévy Mangin, J.; Varela Mallou, J. (2003). Análisis multivariable para las Ciencias Sociales. Pearson Educación S. A.

- VÉLEZ-LANGS, O.; STAFFETTI, E. (2007): "COMPUTACIÓN NEURONAL Y EVOLUTIVA. REDES DE FUNCIONES DE BASE RADIAL". Material didáctico. Asignatura "Computación Neuronal y Evolutiva". Escuela Superior de Ingeniería Informática. Universidad Rey Juan Carlos. Fecha de consulta: Abril de 2010. Disponible en: http://www.ia.urjc.es/~ovelez/docencia/cne/Redes%20de%20Funciones%20de%20Base%20Radial.pdf

- WASSERMAN, P.D. (1989): "NEURAL COMPUTING: THEORY AND PRACTICE". Van Nostrand Reinhold. N.Y.

- ZAMARRIPA TOPETE, J.; SÁNCHEZ RODRÍGUEZ, J. (2007): "PERFILES DE CALIDAD EN EVALUACIÓN INSTITUCIONAL Y PROGRAMA ACADÉMICO, APLICANDO REDES NEURONALES". Anales del VII Congreso internacional "Retos y expectativas de la Universidad". Junio de 2007. Universidad Autónoma de Nuevo León. México. Disponible en: http://www.congresoretosyexpectativas.udg.mx/Congreso%201/Mesa%20E/mesa-e_6.pdf. Fecha de consulta: Febrero de 2010.