

# Uso de Tecnologías de la Información para Detectar Posibles Deserciones Universitarias.

Magister Especialista Marisa Fabiana Haderne  
Universidad Nacional de Cuyo  
Universidad del Aconcagua

Mendoza, Argentina

## Resumen

El tema de la deserción es central al momento de definir políticas universitarias que promuevan principios de equidad social y oportunidades educativas. El objetivo es alcanzar condiciones igualitarias para todos los alumnos, contando con la oportunidad de acceso a las instituciones educativas y de permanencia en ellas.

La deserción puede ser vista desde el punto de la la calidad, como un indicador de falta de eficacia al medir la incapacidad del sistema para conservar a los estudiantes y permitirles cursar sin retrasos ni salidas del sistema educativo.

En la actualidad se utilizan sistemas de información para realizar seguimiento o registros de deserción, pero es importante detectar las variables que inciden en el proceso de aprendizaje y poder comprender su incidencia en el fracaso antes de producirse la deseción.

Este trabajo destaca la importancia del conocimiento al momento de realizar seguimientos académicos y presenta un conjunto de técnicas y herramientas de tecnologías de la información para brindar un complemento a las actividades de seguimiento ya utilizadas. El objetivo es encontrar tendencias y patrones en la información que permitan conocer al alumno y su rendimiento académico y detectar con anticipación posibles deserciones. Es fundamental contar con información útil y oportuna que permita tomar medidas correctivas a tiempo.

Conocer la situación de cada asignatura o detectar grupos de riesgo entre los alumnos es fundamental para cumplir un papel correcto en el desempeño de toda unidad académica.

Hoy en día la información dejó de ser un recurso escaso. Se recibe mucha más información de la que se puede asimilar, siendo el verdadero recurso escaso el tiempo disponible para analizarla, por lo tanto se debe contar con herramientas que permitan acceder al conocimiento.

Para lograr este objetivo, el trabajo se centró en aplicar minería de datos a la información de una unidad académica en la Universidad Nacional de Cuyo. Este proceso consistió en extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados.

Se analizó el rendimiento académico de 3444 alumnos y se aplicó minería de datos para buscar relaciones entre los resultados académicos y diferentes variables que podrían afectar su rendimiento o posible deserción. El trabajo contribuyó mediante el uso de TI a encontrar patrones y tendencias entre estos datos.

Los resultados fueron utilizados por la Secretaría Académica, y sirvieron de complemento al sistema de información ya existente.

## Introducción

La deserción es un problema que se puede ver reflejado en los diferentes niveles educativos. El informe sobre la educación superior en América Latina y el Caribe 2000-2005,

[UNE05]<sup>1</sup> “La metamorfosis de la educación superior” de la Unesco, indica: “La deserción en las universidades está provocando afecciones en la salud física y mental de los jóvenes, y un elevado costo, ya que se estima que al año en América Latina y el Caribe se pierden entre 2 y 415 millones de dólares por país, producto del abandono de los estudios universitarios.”

Según Fernández Lamarra, [LAM02]<sup>2</sup>: “El grave problema en cuanto a equidad e inclusión social de la Argentina, no es el del acceso sino el de las altas tasas de deserción de los estudiantes de la educación superior, en especial de la universitaria”.

Un estudio del CIN muestra que Argentina tiene 4,3 egresados por cada 100 alumnos universitarios, Alemania: 14,8; Canadá: 15,6; Estados Unidos: 14,6; España: 10,6; Japón: 23,4 y Francia: 23,9.

Una unidad académica debe contar con herramientas que le permitan diagnosticar, conocer y analizar las diferentes situaciones conflictivas de aprendizaje que puedan derivar en una deserción, y así implementar estrategias de enseñanza adaptadas a cada necesidad. Es fundamental contar con información útil y oportuna que permita la toma de medidas correctivas a tiempo.

Conocer la situación de cada asignatura o detectar grupos de riesgo en los alumnos es fundamental para cumplir un papel correcto en el desempeño de todo director de carrera, docente o psicopedagogo.

---

<sup>1</sup> [UNE05] UNESCO *La educación superior en el siglo XXI: Visión y acción*. 2005

<sup>2</sup> [LAM02] Fernández Lamarra N. *La educación Superior en la Argentina. Informe para la UNESCO*. 2002

## 1. Objetivos

La Facultad de Ciencias Económicas implementó el proyecto “Mejora en el egreso”, cuyo objetivo es crear las condiciones que permitan a los alumnos próximos a recibirse, con escasa o nula actividad académica en los últimos 3 años, finalizar sus estudios. Contar con una herramienta que permita descubrir patrones de relación entre la actividad académica y el resultado de desempeño, deserción o graduación, permitirá contribuir con este proyecto. Transformar datos en conocimiento es fundamental para encontrar las diferentes problemáticas de estos alumnos.

Se buscó descubrir la relación entre los resultados académicos y la deserción de los alumnos de carreras relacionadas con las Ciencias Económicas. Para ello se aplicaron técnicas de minería de datos (clasificación, estimación, cluster y reglas de asociación), para descubrir relaciones y patrones ocultos en la información académica.

### 1.1 La importancia del conocimiento

Existe una realidad indiscutible en nuestras organizaciones, las educativas entre ellas, es el aumento del volumen de información y sus diferentes formatos (texto, voz, imágenes, videos, entre otros) que se encuentran en varias fuentes internas o externas a la institución.

La mayor parte de esa información representa transacciones históricas, útil para explicar el pasado, conocer el presente y proyectarse hacia el futuro. La mayoría de las decisiones que se toman se basan en experiencias pasadas plasmadas en diferentes fuentes. Surge la necesidad de transformar esta gran cantidad de datos en conocimiento útil.

El valor de la información está dado por el uso que las organizaciones y las personas hacen de ella. Con frecuencia se centra la atención en las tecnologías de la información, las cuales son vistas como la solución a todos

los problemas de una organización y se pierde de vista la información en sí misma. Esto explica en parte por qué algunas implementaciones de TICs (tecnologías de la información y las comunicaciones) fallan.

La información debe ser considerada como el combustible en un auto, por más que se cuente con el mejor vehículo del mercado, con el peor combustible, no llegará muy lejos. El problema radica en cómo distinguir la información útil entre las cantidades inmensas de la misma con que se cuenta, en diferentes formatos y soportes que fluyen por toda la organización y en la mayoría de los casos sin un responsable capaz de evaluar su impacto, valor o costo.

Hoy en día la información dejó de ser un recurso escaso. Se recibe mucha más información de la que se puede asimilar, siendo el verdadero recurso escaso el tiempo disponible para analizar la información. El premio Nobel de Economía Herbert Simon dijo “¿Qué consume la información? Consume la atención de sus destinatarios. En consecuencia, una riqueza de información crea pobreza de atención, y la necesidad de asignar de manera eficiente la atención que entre la sobreabundancia de fuentes de información que se pueda consumir es primordial.”

Por lo tanto, si consideramos a la información como un recurso en exceso, debemos encontrar una herramienta que permita acercarnos al conocimiento, recurso escaso.

Dentro del campo de la inteligencia artificial se encuentran herramientas que ayudan en el proceso de toma de decisiones, como la minería de datos, proceso que permite extraer conocimiento útil y comprensible de grandes cantidades de datos almacenados en distintos formatos. Se utiliza para la clasificación, agrupamiento y reconocimiento de patrones, con grandes volúmenes de datos.

## 1.2 La Minería de Datos

La minería de datos integra numerosas técnicas de análisis de datos y extracción de modelos. Es capaz de describir tendencias, extraer patrones y predecir comportamientos. [OQR04]<sup>3</sup>

Hoy se reconoce la importancia de la extracción de conocimiento a partir de los datos, donde la minería de datos es solo una etapa. Otras de las técnicas que incorpora este importante proceso de obtención de conocimiento se encuentran en el campo del aprendizaje automático, la estadística, las bases de datos, los sistemas de toma de decisiones, la inteligencia artificial.

El aumento de volumen y variedad de información que se encuentra informatizada en bases de datos digitales y otras fuentes ha crecido en las últimas décadas. Surge la necesidad de analizar los datos para la obtención de conocimiento útil para la organización.

La enorme abundancia de datos desborda la capacidad humana de comprenderlos sin utilizar una herramienta poderosa. Muchas veces las decisiones importantes se basan en la intuición, ante la imposibilidad de contar con estas herramientas. Esta es la principal función de la minería de datos: resolver los problemas analizando los datos presentes en las bases de datos.

En el 2000, Witten [OQR04]<sup>3</sup> definió la minería de datos como el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos. Es decir, la tarea fundamental de la minería de datos es encontrar modelos inteligibles a partir de los datos. Para que este proceso sea efectivo debería ser automático o semiautomático (asistido) y el uso de los patrones descubiertos debería ayudar a la toma de decisiones.

---

<sup>3</sup> [OQR04] Orallo J., Ramirez Quintana, Ferri Ramirez C. *Introducción a la minería de datos*. 2004

Por lo tanto dos son los retos de la minería de datos:

- ✓ Trabajar con grandes volúmenes de datos, procedentes en su mayoría de sistemas de información, con sus defectos (ruido, datos ausentes, volatilidad de los datos, etc.)
- ✓ Utilizar técnicas adecuadas para analizar los datos y extraer conocimiento útil y novedoso.

En muchos casos la utilidad del conocimiento obtenido está íntimamente relacionado con la comprensibilidad del modelo inferido. El usuario final suele no ser un experto en el uso de las técnicas de minería de datos, ni tampoco puede perder tiempo interpretando los resultados. Por eso, muchas aplicaciones presentan los resultados en forma de gráficos, convirtiendo los patrones a lenguaje natural.

## 2. Metodología

### 2.1 El proceso de extracción de conocimiento

La minería de datos es una fase en el proceso de extracción del conocimiento, el cual es iterativo e interactivo. Es iterativo ya que la salida de una de las fases puede hacer volver a pasos anteriores y porque a menudo son necesarias varias iteraciones para obtener conocimiento de calidad e interactiva porque el usuario o especialista en el dominio del problema debe ayudar en la preparación de los datos, validación del conocimiento extraído, etc.

En la fase de integración y recopilación de datos se determinan las fuentes de información que pueden ser útiles y dónde conseguirlas. El departamento de procesamiento de datos aportó los datos históricos, se contó con información de 19.017 alumnos inscriptos en la carrera de Contador desde en año 1940.

A continuación se transforman todos los datos a un formato común. Este almacén de datos facilitó enormemente la navegación y

visualización previa de los datos, para discernir que aspectos iban a ser estudiados.

Dado que los datos provenían de diferentes fuentes, podían contener valores erróneos o faltantes. Estos errores se trataron en la fase de selección, limpieza y transformación, en la que se eliminaron o corrigieron los datos incorrectos y se decidió la estrategia a seguir con los datos incompletos. Además se proyectaron los datos para considerar únicamente aquellas variables o atributos que fueran relevantes, con el objetivo de hacer más fácil la tarea de minería y para que los resultados de la misma fueran más útiles. Se trabajó con 3204 patrones de entrada que correspondieron a alumnos que ingresaron entre 1999 hasta 2004, donde 734 se recibieron y 2470 aun no. Se realizó una clasificación entre alumnos recibidos y no recibidos, hombres y mujeres.

En la fase de minería de datos se decidió cual sería la tarea a realizar (clasificar, agrupar, etc.) y se eligió el método a utilizar.

En la fase de evaluación e interpretación se evaluaron los patrones y se analizaron los resultados por la Secretaría Académica, para revisar la necesidad de repetir las fases anteriores y mejorar los resultados. Finalmente en la fase de difusión se hizo uso del nuevo conocimiento.

### 2.2 Proceso de extracción realizado con el software:

*Integración y recopilación:* se realizó mediante consultas a los datos históricos de los alumnos de la base de datos académica de la facultad. Se debió migrar al formato reconocido por excel y depurar los campos no necesarios para la minería.

*Selección, limpieza y transformación:* se utilizaron las herramientas brindadas por la aplicación

*Minería de datos:* se probó con los diferentes modelos para obtener el que mejor se

adaptaba a los requerimientos de información resultante.

*Evaluación e interpretación:* se utilizaron las herramientas brindadas por la aplicación

### 2.3 Datos de trabajo:

3204 alumnos ingresados entre 1999 y 2004

	A	B	C	D	E	F	G	H	I
1	rreg	recibido	titulo	sexo	años al Ingresar	años en carrera	Trabaja	Veces que rindió mal	promedio
2	21209	NO	PERITO MERCANTIL - PRIV.	F	19,00	11	1	13	4,35
3	21210	NO	PERITO MERCANTIL - PRIV.	F	20,00	11	0	9	4,33
4	21211	NO	PERITO MERCANTIL - U.N.C.	F	20,00	11	1	11	4,79
5	21212	NO	PERITO MERCANTIL - PROV.	M	19,00	11	1	20	4,1

Figura 1. Conversión de datos

Se realizó una exploración de datos, para conocer de manera rápida el tipo y distribución de datos con que se contaba. Por la columna de estado de carrera: recibido o no recibido, la figura 2 muestra información de la cantidad de alumnos recibidos respecto de los no recibidos.

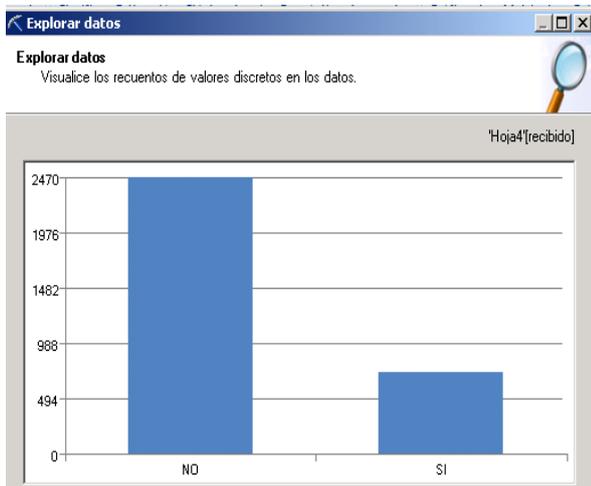


Figura 2. Exploración de datos

### 3. Análisis e interpretación de los resultados

Se seleccionaron diferentes modelos para el análisis de la información.

**3.1. Clasificación:** Generó un modelo de clasificación que predijo los valores de una columna en función de los valores de otras columnas del modelo. El modelo mostró una red de dependencias y un árbol de decisión. La red de dependencia mostró las dependencias entre los atributos de entrada y los atributos de predicción del modelo. Se puede observar en la figura 3 que los atributos: sexo, años del alumno al ingresar, si trabaja no fueron tenidos en cuenta por el algoritmo, y determinó que los atributos relacionados con la variable “si se recibió” fueron: promedio, años de duración de la carrera, título secundario y veces que rindió mal un final.

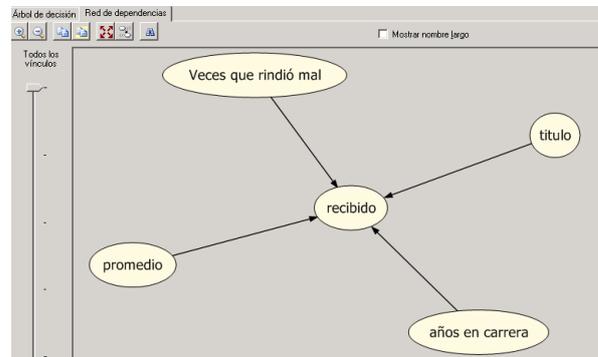


Figura 3 Gráfico de dependencias para clasificación

En el árbol de decisión, el atributo de predicción fue el estado de recibido, el cual es discreto, por lo tanto el visor muestra la condición que provocó la división y un histograma que representa la distribución de los estados del atributo de predicción, ordenados por popularidad. El azul corresponde a no recibido y el rojo al estado recibido.

El árbol completo de decisión para este caso es:



Figura 4 Árbol de decisión

### 3.2. Gráfico de precisión

El gráfico de precisión permite aplicar un modelo a un nuevo conjunto de datos y, a continuación, evaluar su rendimiento. Es un gráfico de elevación, que se suele utilizar para medir la precisión de un modelo de minería de datos. Muestra una representación gráfica de la mejora obtenida al usar el modelo de minería de datos especificado en comparación con predicciones aleatorias y en el caso ideal de que el cien por cien de esas predicciones fuera correcto. La figura 5 muestra el gráfico de precisión para la clasificación anteriormente realizada.

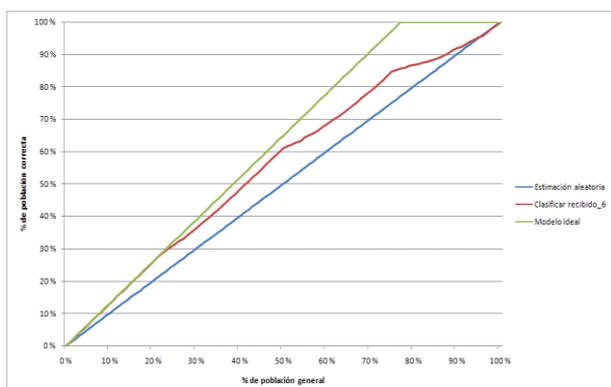


Figura 5 Gráfico de precisión

### 3.3. Cluster

El modelo de clúster detecta grupos de filas que comparten características similares. Este asistente resulta útil para examinar patrones en todo tipo de datos. En la figura 6 se puede observar cómo se relacionan los clúster encontrados por el modelo.

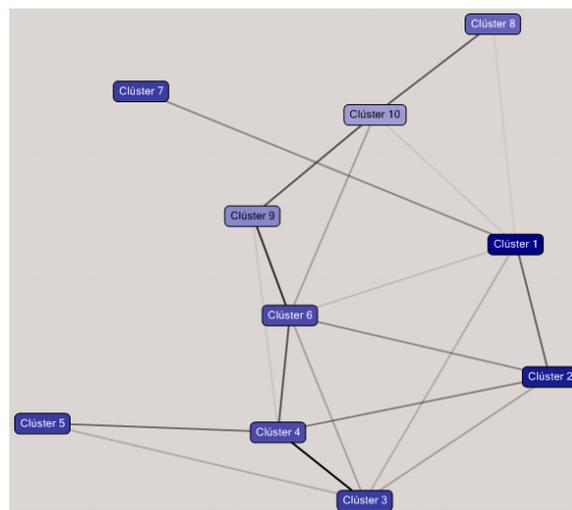


Figura 6 Relación entre clúster

En el gráfico de perfiles del clúster, figuras 7 se puede observar el perfil de cada uno de los clúster.

Variables	Estado	Perfiles del clúster										
		Población Tamaño: 3	Clúster 3 Tamaño: 4	Clúster 2 Tamaño: 1	Clúster 1 Tamaño: 1	Clúster 7 Tamaño: 3	Clúster 5 Tamaño: 1	Clúster 6 Tamaño: 1	Clúster 8 Tamaño: 1	Clúster 4 Tamaño: 1	Clúster 9 Tamaño: 1	Clúster 10 Tamaño: 2
años al ingreso	25,50 19,30 17,00											
años en cama	6 7 8 9 Otros											
promedio	9,30 5,57 1,75											
recibido	NO SI ausente											
sexo	F M ausente											

Figura 7 Pantalla de perfiles de cluster

Se puede observar que el cluster número 3, 2 y 1 corresponden a alumnos no recibidos, mientras que el cluster número. 6 es el que posee más alumnos recibidos, por lo tanto se comparó el cluster 1 con el 6, en la figura 8:

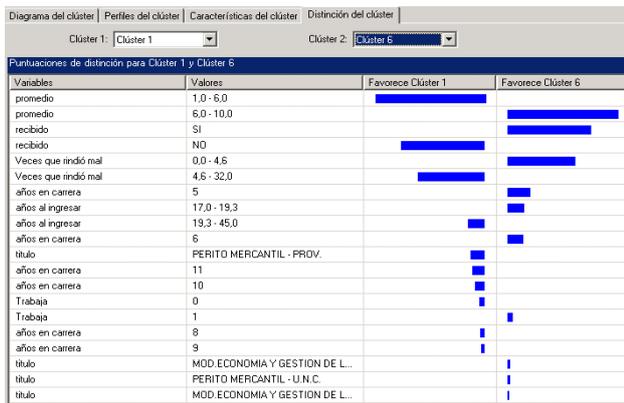


Figura 8 Comparación de cluster 1 y 6

### 3.4. Estimación

Un modelo de estimación extrae patrones de los datos y utiliza la información para predecir los factores que afectan a los valores continuos. Se consideraron los datos de alumnos recibidos en las tres carreras de la Facultad y se estimó la cantidad de posibles recibidos en los próximos 4 años.

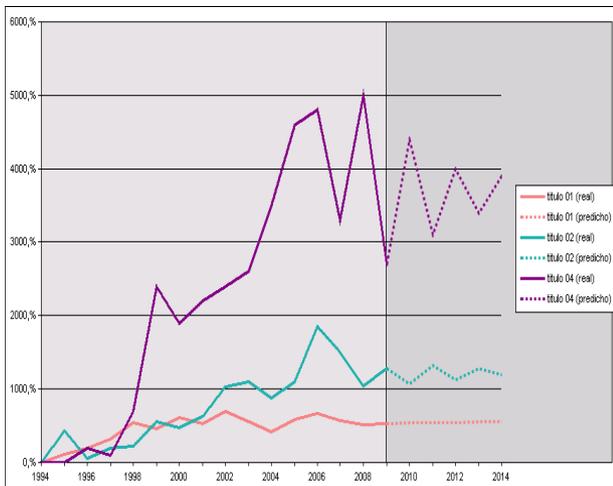


Figura 9 Estimación de egresados por título

### 3.5. Selección de algoritmo

**Reglas de asociación:** El algoritmo de reglas de asociación genera reglas que describen los elementos con mayores probabilidades de aparecer juntos en una transacción. Las reglas se pueden utilizar para predecir la presencia

de un elemento en función de la presencia de otros elementos en una transacción.

En la figura 10 se puede observar el gráfico de dependencias con promedio de notas de la carrera como atributo de predicción:

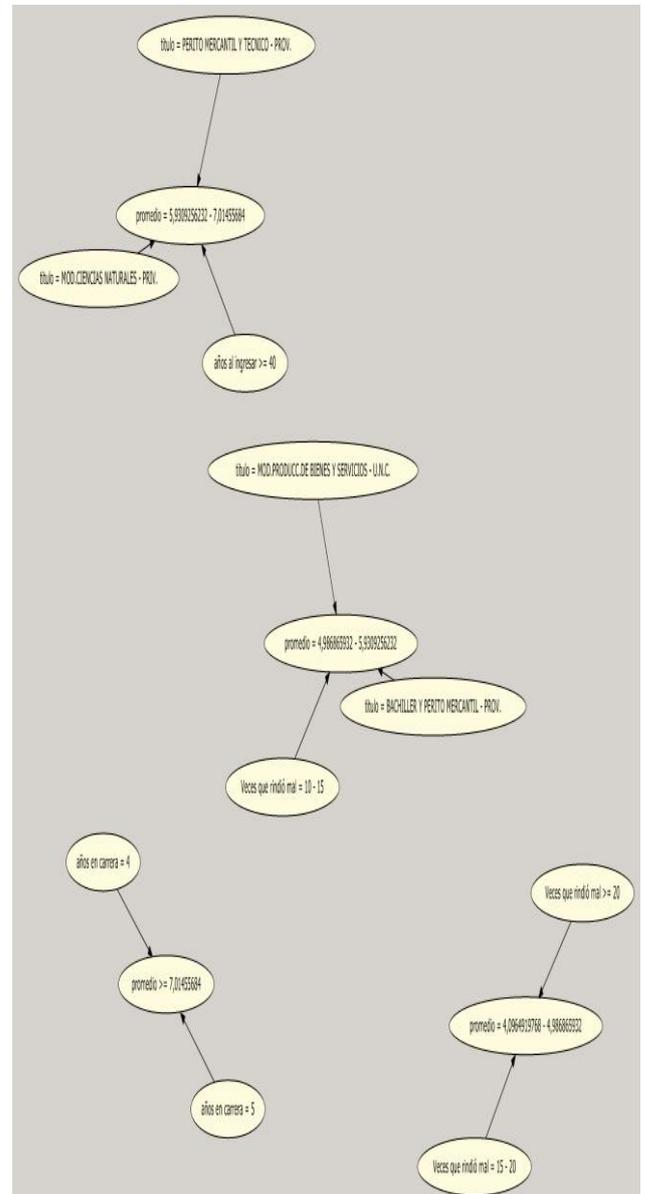


Figura 10 Gráfico de dependencias

La figura 11 corresponde al gráfico de dependencias con la duración de carrera como atributo de predicción:

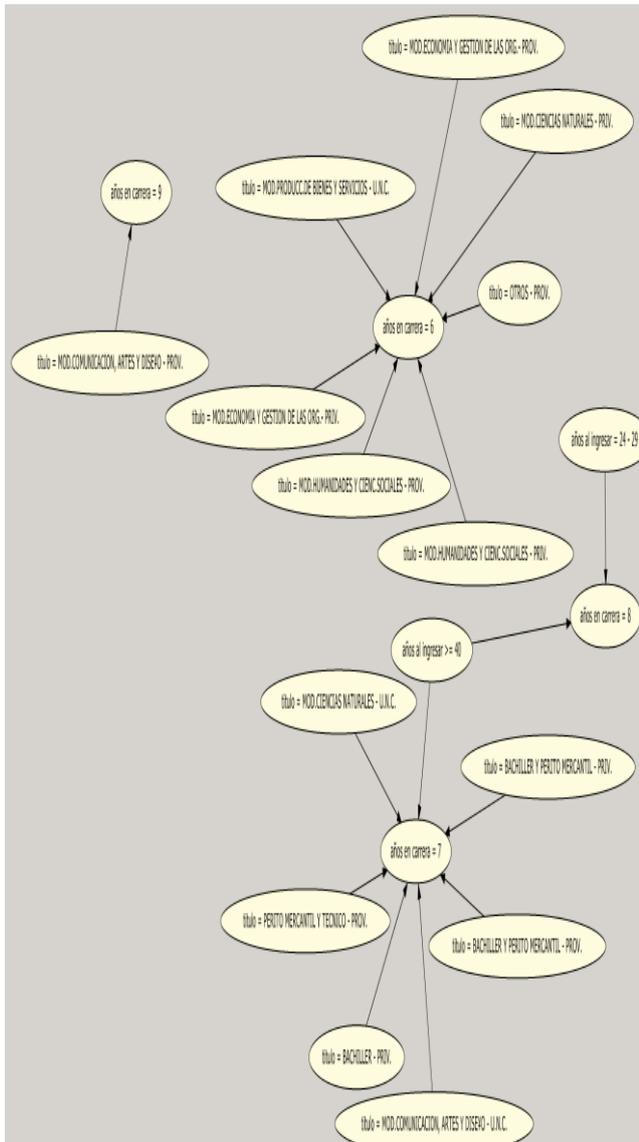


Figura 11 Gráfico de dependencias

#### 4. Transferencia y/o impacto

Se analizó el rendimiento académico de 3444 alumnos de la Facultad de Ciencias Económicas y se aplicó minería de datos para buscar relaciones entre los resultados académicos y diferentes variables que podrían afectar su rendimiento o posible deserción. El trabajo contribuyó mediante el uso de TI a encontrar patrones y tendencias entre estos datos. Los resultados fueron utilizados por la Secretaría Académica, y sirvieron de

complemento al sistema de información ya existente.

Se colaboró con proyecto “Mejora en el egreso”, cuyo objetivo es crear las condiciones que permitan a los alumnos próximos a recibirse, con escasa o nula actividad académica en los últimos 3 años, finalizar sus estudios. Contar con esta herramienta que permite descubrir patrones de relación entre la actividad académica y el resultado de desempeño, deserción o graduación, contribuye con este proyecto. Transformar datos en conocimiento es fundamental para descubrir las diferentes problemáticas de estos alumnos.

Se descubrieron relaciones entre los resultados académicos y la deserción de los alumnos de carreras relacionadas con las Ciencias Económicas. Con esta información se pueden realizar desde la Secretaría Académica diversas actividades que promuevan en los alumnos las competencias necesarias

#### 5. Conclusiones

Se presentaron durante este trabajo herramientas muy potentes que servirán al trabajo de la Secretaría Académica, al conocer mejor las características de los alumnos que se reciben y de aquellos que no lo logran o demoran la obtención de su título.

El cuidado es una característica esencial al momento de aplicar los modelos y algoritmos, pues son muchas las relaciones y patrones que se pueden descubrir, y no siempre sirven al propósito que se busca. Como complemento de este trabajo se dejó copia de todos los resultados de los diferentes modelos, con alumnos recibidos y no recibidos, clasificados por sexo, por edad o colegio secundario. También se incluyeron los análisis de asignaturas por alumno, relacionando el resultado en los exámenes finales con el resultado final de carrera.

Sería beneficioso si se pudieran incorporar al sistema académico los datos correspondientes

al cursado del alumno, notas de parciales, trabajos prácticos entre otras evaluaciones de proceso, porque esa información de seguimiento permitiría detectar problemas de manera temprana, aun cuando el alumno no rindió ningún final.

Se pudo observar de manera gráfica que existe una relación entre el historial académico y el resultado de los alumnos de carreras relacionadas con las Ciencias Económicas, y la misma quedó en evidencia al usar redes neuronales y otros algoritmos de minería de datos.

## 6. Referencias bibliográficas

- [ANT03] Antoni, E.J., *Alumnos universitarios: el por qué de sus éxitos y fracasos*, Buenos Aires. Ed. Miño y Dávila. 2003.
- [GIU02] Giuvagnoli P. *Determinantes de la deserción y graduación universitaria: una aplicación utilizando modelos de duración*, Universidad Nacional de La Plata, Facultad de Ciencias Económicas, disponible en dirección web <http://depeco.econo.unlp.edu.ar> 2002
- [GFM01]Gallarreta S.Felipe A., Merino G.: *Capacidades intelectuales en el nivel universitario: su diagnóstico mediante pruebas de lápiz y papel*, Revista Contexto educativo, año IV, número 24. 2001.
- [GR01] Giarratano, Riley. *Sistemas expertos. Principios y programación*. México. Ed.Thomson. 2001
- [HUI94] Huit W. *A transactional model of teaching/learning process*, disponible en la dirección web [www.chiron.valdosta.edu/whitt/materials/tchlrmd.html](http://www.chiron.valdosta.edu/whitt/materials/tchlrmd.html). 1994
- [HPM99] Hernández López J.M. y Pozo Muñoz C. *El fracaso académico en la universidad: diseño de un sistema de evaluación y detección temprana*, Revista de Psicología educativa, vol.5 nro.1- Pág.27-40. 1999
- [JAH00] Jewsbury A. y Haefeli I. *Análisis de la deserción en Universidades Públicas Argentinas*, Trabajo presentado en el V Congreso Internacional del CLAD de Santo Domingo, disponible en la dirección web <http://www.tribunaldecuentas.santafe.gov.ar/congreso>. 2000.
- [KAP03] Kaplan C.V. Buenos y malos alumnos. Buenos Aires. Ed. Aique. 2003.
- [KLL02] K. Laudon y J. Laudon: *Sistemas de Información Gerencial*. Prentice Hall. 2002
- [LAM02] Fernández Lamarra N. *La educación Superior en la Argentina*. Informe para la UNESCO. 2002.
- [NIL01] Nilsson N.J. *Inteligencia Artificial. Una nueva síntesis*. Madrid. Ed. Mc Graw Hill.. 2001.
- [OQR04] Orallo J., Ramirez Quintana, Ferri Ramirez C. *Introducción a la minería de datos*. Madrid Pearson. . 2004.
- [PRE06] Pressman, Roger. *Ingeniería del software. Un enfoque práctico* Mc Graw Hill. Madrid. 2006
- [RNP96] Russell S. y Norvig P. *Inteligencia Artificial*. México. Pearson Educación.. 1996
- [SEH00] Sposetti A. y Echevarría H. “*El factor educacional como causa potencial de la deserción en primer año de la universidad*”, proyecto aprobado y subsidiado por la SeCyT. UNRC, disponible en la dirección web [www.unr.edu.ar/publicar/cde/h21.htm](http://www.unr.edu.ar/publicar/cde/h21.htm) 1999
- [VGL04] Viñuela Pedro, Galván León Inés. *Redes neuronales artificiales. Un enfoque práctico*. Madrid. España. Pearson Educación. 2004.