

# Minería de datos para descubrir estilos de aprendizaje

ELENA DURÁN  
ROSANNA COSTAGUTA

Universidad Nacional de Santiago del Estero, Argentina

---

## 1. Introducción

Felder y Silverman (1988) afirman que los estudiantes aprenden de muchas maneras: viendo y escuchando, reflexionando y actuando, razonando lógicamente e intuitivamente, memorizando y visualizando, construyendo analogías y modelos matemáticos. También los métodos de enseñanza son variados. Algunos instructores leen, otros demuestran o discuten, algunos se centran en principios y otros en aplicaciones, algunos enfatizan la memorización y otros la comprensión. Cuanto aprenda un estudiante en una clase dependerá de la habilidad innata y de su preparación previa, pero además de la compatibilidad entre su estilo de aprendizaje y el estilo de enseñanza de su instructor.

Como docentes de la carrera Licenciatura en Sistemas de Información (LSI) de la Facultad de Ciencias Exactas y Tecnologías perteneciente a la Universidad Nacional de Santiago del Estero (UNSE) en Argentina, consideramos que nos resultaría útil determinar las características del perfil de aprendizaje de nuestros estudiantes para, en función de ellas, adecuar las estrategias de enseñanza. Con este propósito encuestamos a ochenta y nueve estudiantes, de la carrera antes citada, utilizando el Test propuesto por Felder y Soloman (1984). Con el fin de descubrir el conocimiento implícito en las respuestas, se siguió el Proceso de Descubrimiento de Conocimiento en Bases de Datos, y a través de técnicas de minería de datos se logró detectar el estilo de aprendizaje dominante en nuestro grupo de estudiantes.

El presente trabajo se estructura de la siguiente forma: en la sección 2 se describe brevemente el modelo de estilo de aprendizaje de Felder y Silverman (1988); la sección 3 presenta una pequeña descripción del Proceso de Descubrimiento de Conocimiento en Bases de Datos; la sección 4 documenta la investigación llevada a cabo; la sección 5 contiene algunos lineamientos para adecuar las estrategias de enseñanza; y finalmente, en la sección 6, se sintetizan algunas conclusiones fruto de la investigación desarrollada.

## 2. Modelo de estilos de aprendizaje

Un modelo de estilos de aprendizaje clasifica a los estudiantes de acuerdo con la forma en cómo ellos reciben y procesan la información. El modelo propuesto por Felder y Silverman (1988) es particularmente aplicable a estudiantes de ingeniería. Estos autores proponen, paralelamente, un modelo de estilo de

enseñanza que clasifica los métodos instruccionales de acuerdo a cuan bien direccionan los componentes del estilo de aprendizaje propuesto. En su trabajo original, los autores consideraron cinco dimensiones de análisis: percepción, entrada, organización, procesamiento y comprensión; pero en una posterior versión suprimen la dimensión de organización (inductivo y deductivo). Ambos modelos actualizados se presentan en la Tabla 1.

- APRENDIZAJE VISUAL Y VERBAL: Los estudiantes visuales recuerdan mejor lo que ven: figuras, diagramas, cuadros, líneas de tiempo, películas, demostraciones. Los estudiantes auditivos recuerdan más lo que escuchan y mucho más lo que ellos dicen. Ellos aprenden a partir de la discusión y prefieren las explicaciones verbales a las demostraciones visuales. Aprenden efectivamente cuando ellos pueden explicarle a otros.
- APRENDIZAJE ACTIVO Y REFLEXIVO: El proceso mental complejo por el cual la información percibida es convertida en conocimiento puede ser convenientemente agrupado en dos categorías: experimentación activa y observación reflexiva. La primera implica hacer algo en el mundo externo con la información (discutirla, explicarla o chequearla de alguna manera), y la segunda implica examinar y manipular la información introspectivamente. Según esto, un estudiante activo es el que se siente más cómodo con la experimentación activa que con la observación reflexiva, al revés de un estudiante reflexivo. Los estudiantes activos no aprenden mucho en situaciones en las que ellos deben estar pasivos, tales como lecturas; trabajan bien en grupos y tienden a ser experimentalistas. Los estudiantes reflexivos no aprenden bien en situaciones que no les proporcionan la oportunidad de pensar sobre la información que se les presenta; trabajan mejor solos o a lo sumo con una persona más; tienden a ser teóricos.
- APRENDIZAJE SECUENCIAL Y GLOBAL: El aprendizaje secuencial es el aplicado en la mayoría de la educación formal, e implica la presentación de material en un orden de progresión lógica, con el avance del aprendizaje regido por el tiempo y el calendario. Cuando un cuerpo de material ha sido cubierto los estudiantes son evaluados y recién pueden pasar al nivel o cuerpo siguiente. El aprendizaje global es aquel que no se rige por el tiempo ni el calendario, por el contrario se pueden pasar días o semanas ocupados en resolver un simple problema o demostrando una comprensión rudimentaria hasta que de pronto se le “encienden las luces” y logran una rápida comprensión del todo.

TABLA 1  
Modelo de estilo de aprendizaje-enseñanza

ESTILO DE APRENDIZAJE		CORRESPONDIENTE ESTILO DE ENSEÑANZA	
Sensorial Intuitivo	Percepción	Concreto Abstracto	Competidor
Visual Verbal	Entrada	Visual Verbal	Presentación
Activo Reflexivo	Procesamiento	Activo Pasivo	Participación del estudiante
Secuencial Global	Comprensión	Secuencial Global	Perspectiva

Basándose en lo expuesto, Felder y Silverman (1988) proponen 16 estilos de aprendizaje a partir de las combinaciones posibles de los estilos dados.

### 3. Proceso de descubrimiento de conocimiento

La existencia de voluminosas bases de datos conteniendo grandes cantidades de datos, que exceden en mucho las capacidades humanas de reducción y análisis a fin de obtener información útil, actualmente son una realidad en muchas organizaciones. Debido a esto, frecuentemente, las decisiones importantes se toman en base a la intuición y experiencia del decisor más que considerando la rica información almacenada (Han y Kamber, 2001). Esta situación se intenta solucionar a través del proceso de KDD, el cual implica la realización de tres etapas: preprocesamiento, minería de datos (*datamining*), y postprocesamiento, que se explican brevemente a continuación.

La etapa de preprocesamiento tiene por objetivo preparar los datos para que puedan ser sometidos a la etapa siguiente del proceso. Dentro de las técnicas para realizar el preprocesamiento cabe mencionar a (Han y Kamber, 2001): limpieza de datos, a fin de remover ruido e inconsistencias; integración de datos, para generar un único almacén de datos coherente en aquellos casos donde los datos provienen de diferentes fuentes; transformaciones de datos, para normalizarlos; y reducción de datos, a fin de reducir el tamaño de los datos, por ejemplo, eliminando características redundantes. La importancia del preprocesamiento de los datos se debe a que la calidad de los datos sobre los que se aplican técnicas de KDD impacta de manera directa en la calidad del conocimiento que se descubre a partir de ellos (Han y Kamber, 2001).

La etapa de minería de datos puede definirse sobre la base de un conjunto de primitivas diseñadas especialmente para facilitar un descubrimiento de conocimientos eficiente y fructífero. Tales primitivas incluyen (Han y Kamber, 2001): la especificación de las porciones de la base de datos o del conjunto de datos en los que se quiere trabajar; la clase de conocimiento a ser descubierto; los conocimientos existentes que podrían resultar útiles para guiar el proceso de KDD; las métricas de interés para llevar a cabo la evaluación de patrones en los datos analizados; y finalmente, las formas en que el conocimiento descubierto podría ser visualizado.

La etapa de postprocesamiento implica la realización de algún tipo de reformulación de los resultados obtenidos producto de la minería de datos realizada. Se pretende, así, que los conocimientos encontrados sean más fáciles de entender y utilizar por el usuario a quien finalmente están destinados.

### 4. Desarrollo de la investigación

Como docentes de la LSI, consideramos que nos resultaría útil determinar las características del perfil de aprendizaje de nuestros estudiantes para, en función de ellas, por ejemplo, adecuar las estrategias de enseñanza de los profesores de la carrera.

La principal fuente de datos para llevar a cabo esta investigación la constituyeron los alumnos activos de la mencionada carrera, de los cuales se seleccionaron al azar, ochenta y nueve que constituyeron

la muestra. Tal cantidad representa aproximadamente un 10% de la población. El instrumento utilizado para recolectar los datos fue el Test de Estilos de Aprendizaje propuesto por Felder y Soloman (1984). Este Test consiste en cuarenta y cuatro preguntas con dos opciones de respuesta cada una (a y b), en las que el alumno debe elegir necesariamente sólo una de ellas.

Con la información recabada a través de esta encuesta se generó una base de datos, que fue sometida al proceso de KDD, siguiendo sus tres etapas: pre-procesamiento, minería de datos, y post-procesamiento. La información recabada con el test fue cargada en una planilla de EXCEL (Figura 1), luego convertida a un documento WORD para su posterior conversión a un archivo .arff(Figura 2), requerido como entrada para el software WEKA (Witten y Frank, 1999) (Figura 3).

Entre las actividades de preprocesamiento se aplicó limpieza de los datos mediante el reemplazo de valores faltantes por el signo "?", también se eliminaron algunas inconsistencias a través de correcciones manuales, tales como la existencia de espaciados innecesarios y otros errores de tecleo. Dado que se trabajó con un único archivo no fue necesario realizar tareas de integración ni de transformación. Se llevó a cabo una única tarea de reducción de datos que consistió en eliminar los atributos: nombre del alumno, por resultar irrelevante para los objetivos de este proceso de descubrimiento de conocimiento, y las respuestas a los ítems 8 y 16, por corresponder a la dimensión procesamiento.

FIGURA 1  
 Datos de las encuestas cargados en Excel

	A	B	D	E	F	G	H	J	K	L	M	N	O	P	
		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14
1	Nombre y Apellido														
2	Galvn Fernanc	a	b	a	b	b	a	a	a	a	b	a	a	a	b
3	Pa svecino Alberto Edmundo	b	a	b	b	a	a	b	b	b	b	b	b	b	b
4	Singer Walter Ariel	a	a	a	b	b	a	a	b	a	a	b	c	a	c
5	Colonel Roberto	a	b	a	b	b	a	b	b	a	a	a	a	a	b
6	Diez Cristian Adrian	a	b	a	b	b	a	b	b	a	a	b	a	a	a
7	Quintarc Cristian G.	a	b	a	b	a	a	a	b	b	a	a	b	b	c
8	Farias Roberto Adrian	a	b	a	b	b	a	a	b	b	a	a	a	a	b
9	Qozn Victor Alberto	a	b	a	b	b	a	a	b	a	a	a	a	a	a
10	Correa Luca Carina	a	a	a	b	b	a	a	b	a	a	a	c	b	c
11	Leguizamn Fernanda Magda	a	a	a	h	h	a	a	h	a	a	a	a	a	b
12	Runer Paula de Villagiu	b	a	a	b	b	a	b	a	b	a	a	a	a	b
13	Auran Oscar Alejandro	a	a	a	a	b	a	a	b	a	b	a	a	c	a
14	Torales Villacs Francisco	a	h	a	a	h	a	a	h	a	a	a	b	a	a
15	Gurnez Diego Alejandro	b	a	b	b	a	a	b	a	a	a	b	a	b	b
16	Mora cs Juan Marcelo	a	a	a	b	a	a	a	b	b	a	a	c	a	c
17	Marnez Pablo Luis	a	a	a	h	a	a	h	h	a	a	a	a	a	b
18	Coronel Carlos Stella Mercedes	a	a	b	b	b	a	a	a	a	a	b	b	a	a
19	Leiva Marcela Vernica	a	a	a	a	b	a	b	a	a	a	a	c	a	b
20	Lascano Fernando Daniel	h	a	a	h	a	a	h	a	a	a	h	a	a	b
21	Isaniz Marta	a	a	a	a	b	a	b	a	a	b	a	a	b	a
22	Moro Vernica	b	a	a	a	a	a	b	a	a	a	a	c	a	b

FIGURA 2  
 Conversión en Word para generar archivo .arff

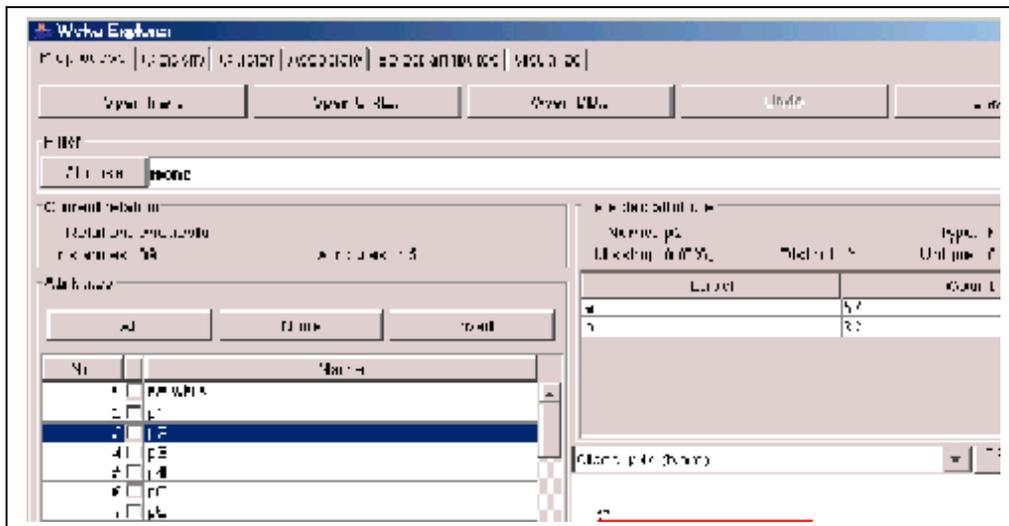
```

@RELATION encuesta

@ATTRIBUTE nombre STRING
@ATTRIBUTE p1 {a,b}
...
@ATTRIBUTE p42 {a,b}
@ATTRIBUTE p43 {a,b}
@ATTRIBUTE p44 {a,b}

@DATA
GalvánFernando,a,b,a,b,b,a,a,a,a,b,a,a,b,b,b,a,a,a,b,
b,b,b,a,a,b,b,a,a,b,a,b,b,b,a,b,b,a,a,a,b,a,b,b
PalavecinoAlbertoEdmundo,b,a,b,b,a,b,b,b,b,b,b,b,b,
a,b,a,b,b,a,?,a,a,a,b,b,?,b,b,a,b,a,a,a,b,?,a,a,a,a,b
,a
SingerWalterAriel,a,a,a,b,b,b,a,b,a,a,b,a,a,a,b,a,b,a,a
,a,b,b,a,a,b,b,b,a
...
    
```

Archivo .arffabierto con Weka



La etapa de minería de datos fue efectuada guiando la investigación a través de las cinco preguntas planteadas por Han y Kamber (2001):

- *¿Cuál es el conjunto de datos que se consideró relevante para realizar el descubrimiento de conocimientos?*

Se decidió trabajar con la base de datos completa resultante del preprocesamiento.

- *¿Qué clase de conocimiento se quiere descubrir?*

En función del tipo de conocimiento a descubrir se decidió utilizar tareas de minería de datos del tipo descriptivo. En particular, se aplicó análisis de cluster (Witten y Frank, 1999) para identificar subgrupos homogéneos dentro de la población de alumnos encuestada. Para ello se utilizó al software *WEKA* y se seleccionó la opción *Cluster*. En particular, se optó por el algoritmo *FarthestFirst* por tratarse de un problema de k centros donde se pretende que la máxima distancia entre una tupla y su centroide sea mínima (Chaudhuri, *et al.*, 1998). Para detectar los estilos de aprendizaje dominantes se decidió mantener en dos la cantidad de clusters a generar.

Como puede apreciarse en los resultados, que se muestran a continuación, no existe necesidad de ampliar este número dado el alto porcentaje de tuplas incluidas en el primer cluster.

```

=== Run information ===
Scheme:   weka.clusterers.FarthestFirst -N 2 -S 1
Relation: encuesta-weka.filters.unsupervised.attribute.Remove-R1
Instances: 89
Attributes: 42
           p1
           p2
           ...
           p44
Test mode: evaluate on training data
=== Clustering model (full training set) ===
FarthestFirst
=====
Cluster centroids:
Cluster 0 a b a b b a a a a a a b a b a a b b b a a b b b b a a a a b a b b a a b a a a b
Cluster 1 b a b a a a b a a b b b a b b a a b b a b b b b a a a a b b b b a a b a a b a b b b
=== Evaluation on training set ===
FarthestFirst
=====
Cluster centroids:
Cluster 0 a b a b b a a b a a a a b a a b a a b b b a a b b b b a a a a b a b b a a b a a a b
Cluster 1 b a b a a a b a a b b b a b b b a a b b a b b b b a a a a b b b b a a b a a b a b b b
Clustered Instances
0    76 ( 85%)
1    13 ( 15%)
    
```

- ¿Qué conocimiento de background resultará de utilidad?

A partir de las características definidas por Felder y Silverman se determinaron las correspondencias entre los estilos de aprendizaje y las posibles respuestas a las cuarenta y cuatro preguntas del test de Felder y Soloman. Los resultados se muestran en la Tabla 2. La información de Tabla 2 puede ser fácilmente expresada como una jerarquía basada en reglas a fin de representar las jerarquías de conceptos obtenidas a partir de nuestro conocimiento del dominio

**TABLA 2**  
**Correspondencia entre los estilos de aprendizaje y las posibles respuestas a las cuarenta y cuatro preguntas del test**

ESTILOS DE APRENDIZAJE	CÓDIGO	RESPUESTAS CONSIDERADAS
SENSORIAL	SN	P2 = a, P6 = a, P10 = a, P14 = a, P18 = a, P22 = a, P26 = a, P29 = a, P30 = a, P34 = a, P38 = a, P42 = a, P43 = a.
INTUITIVO	I	P2 = b, P6 = b, P10 = b, P14 = b, P18 = b, P22 = b, P26 = b, P29 = b, P30 = b, P34 = b, P38 = b, P42 = b, P43 = b.
VISUAL	V	P3 = a, P7 = a, P11 = a, P15 = a, P19 = a, P23 = a, P27 = a, P31 = a, P35 = a, P39 = a.
AUDITIVO	AU	P3 = b, P7 = b, P11 = b, P15 = b, P19 = b, P23 = b, P27 = b, P31 = b, P35 = b, P39 = b.
ACTIVO	AC	P1 = a, P5 = a, P9 = a, P13 = a, P14 = a, P17 = a, P21 = a, P25 = a, P33 = a, P37 = a, P41 = a.
REFLEJO	R	P1 = b, P5 = b, P9 = b, P13 = b, P14 = b, P17 = b, P21 = b, P25 = b, P33 = b, P37 = b, P41 = b.
SECUENCIAL	SC	P4 = a, P12 = a, P20 = a, P24 = a, P28 = a, P32 = a, P36 = a, P40 = a, P44 = a.
GLOBAL	G	P4 = b, P12 = b, P20 = b, P24 = b, P28 = b, P32 = b, P36 = b, P40 = b, P44 = b.

- *¿Qué medidas pueden utilizarse para estimar patrones de interés?*

A partir de este conocimiento se decidió reemplazar cada uno de los atributos incluidos en los dos centroides de los clusters arrojados por Weka por el código del estilo de aprendizaje con el que se vincula (según Tabla 2). Con lo que los centroides quedaron expresados como sigue:

Centroide Cluster 0:

AC, I, V, G, R, SN, V, AC, SN, V, SC, AC, I, V, R, SN, V, G, R, I, V, SC, R, I, AU, G, SN, SN, V, SC, AC, I, V, G, R, SN, V, G, AC, SN, SN, G

Centroide Cluster 1:

R, SN, AU, SC, AC, SN, AU, AC, SN, AU, G, R, SN, AU, R, SN, V, G, R, SN, AU, G, R, I, V, SC, SN, SN, AU, G, R, I, V, SC, R, SN, V, G, AC, I, I, G

A fin de determinar la combinación de estilos de aprendizaje representada por cada centroide, se procedió a contar las cantidad de ocurrencias de cada estilo. Resultando:

Centroide Cluster 0:

SN = 8; I = 5;  
 V = 9; AU = 1;  
 AC = 5; R = 5  
 SC = 3; G = 6

Por lo que este centroide representa el estilo *sensitivo-visual- activo/reflexivo-global (SN, V, AC/R, G)* que resulta ser el dominante en la población de alumnos encuestados.

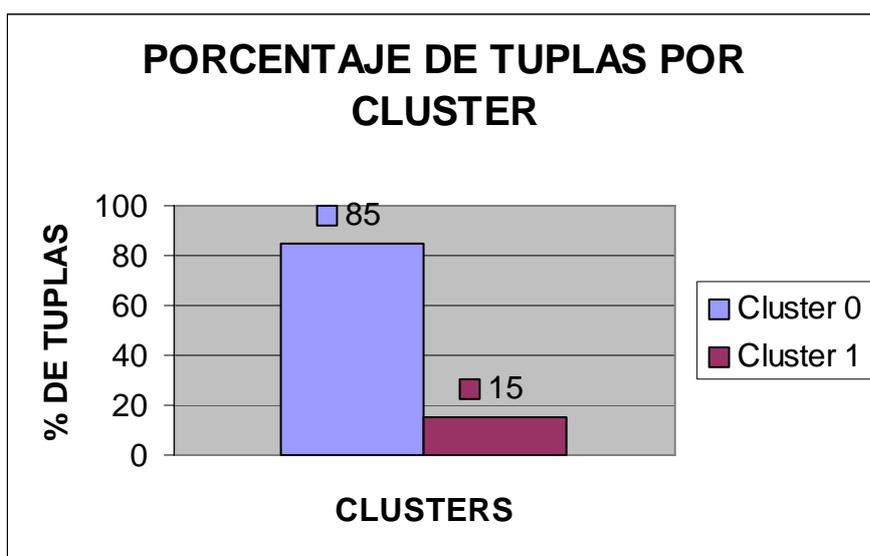
Centroide Cluster 1:  
 SN = 9; I = 4;  
 V = 4; AU = 6;  
 AC = 3; R = 7  
 SC = 3; G= 6

Por lo que este centroide representa el estilo *sensitivo-auditivo-reflexivo-global (SN, AU, R, G)*.

- *¿Cómo se visualizarán los patrones descubiertos?*

Para visualizar los patrones descubiertos se optó por un gráfico de barras que muestre los clusters resultantes en función del porcentaje de tuplas que incluye cada uno (Figura 4). En particular, aproximadamente el 85% de la población encuestada constituyó el primer cluster, y sólo un 15 %, el segundo.

FIGURA 4  
 Visualización de resultados del clustering



## 5. Lineamientos generales para estrategias de enseñanza

Con el objetivo de transferir a la práctica cotidiana de nuestros docentes, y a fin de propiciar la compatibilidad entre el estilo de aprendizaje de los alumnos y el de enseñanza de nuestros profesores, decidimos sugerir algunas estrategias de enseñanza conforme al perfil obtenido de los alumnos encuestados.

Un estilo de enseñanza se define en término de las respuestas a cuatro preguntas planteadas considerando las dimensiones descritas en el apartado 2 de este trabajo Felder y Silverman (1988): ¿qué

tipo de información enfatizar?, ¿en qué modo de presentación se debe hacer hincapié?, ¿qué forma de participación del estudiante debe enfatizarse con la presentación?, ¿qué tipo de perspectiva se proporciona con la información presentada?. De esta manera, y atendiendo al estilo de aprendizaje dominante, se sugiere:

- El tipo de información que se presente a los alumnos debe corresponderse con hechos concretos.
- La información debe ser presentada preferiblemente en forma visual (por ejemplo mediante imágenes, diagramas de flujo, etc.). Por lo tanto, es recomendable acompañar la explicación de procesos y de algoritmos complejos con diagramas de flujos o diagramas lógicos. También las funciones matemáticas complejas deberían ilustrarse, o bien, utilizarse software matemático con capacidades de graficación. Una herramienta importante a tener en cuenta es el uso de software de simulación que permita mostrar y analizar sistemas de información (SI), sistemas de redes, análisis de procesos y flujos de trabajo, etc.
- En las clases sería conveniente incentivar tanto la participación activa del estudiante (opinando, reflexionando, actuando, etc.) como la pasiva (viendo y escuchando). Además, dado el perfil de un egresado de la LSI, es necesario que sean capaces de observar reflexivamente un fenómeno e identificar problemas existentes, para lo cual se requiere una actitud mayormente pasiva. Sin embargo, también es necesaria una participación activa, a la hora de evaluar ideas y soluciones, diseñar y desarrollar sistemas. Teniendo en cuenta el estilo de aprendizaje descubierto, es conveniente que el docente alterne la lectura de material bibliográfico y las clases expositivas, tanto con espacios de reflexión como con pausas para discusión y resolución de problemas.
- La información debe proporcionarse al alumno desde una perspectiva global con visión del contexto y de lo relevante. Es conveniente hacer conocer los objetivos del aprendizaje antes de desarrollar cada tema, y también mostrar la vinculación del tópico con otros de la unidad y de otras asignaturas del currículo. Es aconsejable dejar que los alumnos apliquen sus propios métodos para solucionar problemas, antes que forzarlos a seguir estrategias impuestas por el docente. También sería importante brindarles libertad para crear sus soluciones, en especial en aquellas asignaturas vinculadas con el diseño de SI.

## 6. Conclusiones

La aplicación del proceso de KDD permitió determinar que existe un alto grado de homogeneidad en el estilo de aprendizaje de los alumnos de la LSI, perteneciente a la UNSE. El análisis por cluster realizado permitió identificar al estilo *sensitivo-visual-activo/reflexivo-global* como el dominante.

En este trabajo también se plantean lineamientos generales para adecuar los estilos de enseñanza de los docentes de la carrera mencionada respecto al estilo de aprendizaje dominante en la población de alumnos. Cabe aclarar que estos lineamientos deberían ser particularizados dentro del marco de cada una de las asignaturas. Actualmente, las autoras de este trabajo, se encuentran diseñando un plan de validación

de los resultados obtenidos, que será aplicado durante el año académico 2006, sobre los alumnos de la asignatura Simulación en el ciclo superior de la carrera citada.

## Bibliografía

- CHAUDHURI, S.; GARG, N., y RAVI, R. (1998): "The p-Neighbor k-Center Problem", en *Information Processing Letters*, vol. 65, pp. 131-134. <http://citeseer.ist.psu.edu/chaudhuri98pneighbor.html> [Consulta: Marzo 2006].
- FELDER, R. M., y SOLOMAN, V. (1984): *Index of Learning Styles*. <<http://www.ncsu.edu/felder-public/ILSpage.html>> [Consulta: Marzo 2006].
- FELDER, R. M. y SILVERMAN, L. K. (1988): "Learning and Teaching Styles in Engineering Education Application". *Engr. Education*, vol. 78 (7), pp. 674-681.
- HAN, J., y KAMBER, M. (2001): *Data Mining: Concepts and Techniques*. USA, Academic Press.
- WITTEN, I. y FRANK, E. (1999): *Data Mining: Practical machine, learning tools and techniques with Java implementations*. USA, Morgan Kauffmann Publishers.