

---

# Data Mining en la Empresa y en las Finanzas Utilizando Tecnologías Inteligentes

---

**Richard Weber<sup>1</sup>**

Departamento de Ingeniería Industrial,  
Universidad de Chile

---

## Resumen

---

*El desarrollo de software y hardware computacional ha resultado en la generación de enormes cantidades de datos almacenados en muchas empresas y organizaciones. Estos datos contienen información valiosa, que puede ser usada para mejorar la competitividad de estas instituciones. Data mining (minería de datos) se ha convertido en un área de investigación y desarrollo, en el cual se están proponiendo técnicas que apuntan a encontrar la información oculta en los datos. Este trabajo presenta aplicaciones de técnicas de data mining que han sido realizadas por un departamento de marketing en una institución bancaria.*

---

## 1. Introducción

---

El marketing juega un papel crucial en la mayoría de las empresas, tanto en las áreas de negocios y finanzas, como de comercio, servicios y telecomunicaciones. Entre las metas respectivas se incluyen las de encontrar nuevos clientes y mantener los actuales reconociendo y satisfaciendo sus necesidades. Hoy en día el desarrollo de *software* y *hardware* computacional permite a las empresas recoger y almacenar enormes cantidades de datos descriptivos de sus clientes y su comportamiento. El área de investigación y aplicación en la que se desarrolla estrategias de marketing fundadas en tales bases de datos se llama *database marketing* (Furness 1994; Nash 1993).

El presente trabajo considera aplicaciones de técnicas de *data mining* para el *database marketing*, tratando en mayor detalle del uso de tecnologías inteligentes tales como la lógica difusa (*fuzzy logic*) y las redes neuronales (*neural networks*).

---

1 Dirección: Departamento de Ingeniería Industrial, Universidad de Chile, República 701, Santiago, Chile. E-mail: rweber@dii.uchile.cl

En la Sección 2 se presentan las definiciones básicas junto con un resumen panorámico de *data mining*. La sección 3 presenta las tecnologías inteligentes usadas en esta publicación. En la sección 4 se presentan dos sub-tareas de un proyecto de *database marketing* realizado por una institución bancaria, junto con sus correspondientes resultados. La sección 5 presenta un resumen del trabajo, haciendo alusión además a futuros desarrollos y aplicaciones de los enfoques de *data mining*.

---

## 2. Descubrimiento de Conocimientos en Bases de Datos y Data Mining

---

*Data mining* entrega diversas técnicas para encontrar patrones en grandes conjuntos de datos. Este enfoque multidisciplinario combina los resultados e intuiciones provenientes de varias ramas científicas tales como la estadística, el aprendizaje de máquina (*machine learning*), tecnologías difusas (*fuzzy technologies*), y redes neuronales (véase por ejemplo, Adriaans y Zantinge 1996).

En la literatura existen muchas definiciones para caracterizar expresiones como “descubrimiento de conocimiento en base de datos” (KDD: *Knowledge Discovery in Databases*) y “*data mining*” (véase por ejemplo Frawley *et al.* 1991), donde se hacen las siguientes aseveraciones:

*“El descubrimiento de conocimiento en bases de datos es el proceso no trivial de identificar patrones en datos que sean válidos, novedosos, potencialmente útiles y, por último, comprensibles”* (Fayyad 1996, p. 21)

*“... data mining se refiere al acto de extraer patrones o modelos a partir de los datos ...”* (Fayyad 1996)

Una representación frecuente de un proceso típico de KDD, contempla los siguientes nueve pasos (véase por ejemplo Fayyad 1996, p. 23)

1. Desarrollar una comprensión del dominio de la aplicación
2. Crear un conjunto de datos objetivo
3. Limpieza y pre-procesamiento de los datos
4. Reducción y transformación de los datos
5. Elegir la tarea de *data mining*
6. Elegir los algoritmos de *data mining*
7. *Data mining*
8. Evaluar el resultado del *data mining*
9. Consolidar el conocimiento descubierto

En el paso 2 hay que seleccionar un sub-conjunto de los datos disponibles que normalmente están almacenados en una base de datos o en un *Data Warehouse* (bodega de datos). En una base de datos como sistema tradicional de almacenamiento

se puede captar datos de transacciones de un negocio. En cambio un *Data Warehouse* como sistema analítico apoya la toma de decisiones ofreciendo varias vistas distintas a los datos almacenados (Anahory y Murray 1997). Ambos sistemas sirven como fuente de un conjunto de datos objetivo.

En muchas aplicaciones reales los datos en una base de datos o también en un *Data Warehouse* contienen varias formas de errores, como por ejemplo valores faltantes, valores fuera del rango del atributo considerado o valores no plausibles (Guynes et al. 1996). Esa razón hace necesario el paso 3 de limpieza y pre-procesamiento de los datos (Famili et al. 1997).

En el paso 5 se elige la tarea de *data mining* específica para un análisis más detallado. Aquí se debe decidir si el propósito es, por ejemplo, la agrupación de objetos, la regresión o el modelaje de dependencia. Sobre la base de esta decisión, los más importantes algoritmos de *data mining* deben ser seleccionados (paso 6), los que se usan en la búsqueda real de patrones en los datos (paso 7).

Este trabajo presenta una aplicación del proceso de KDD para un banco donde, en el paso 5, han sido elegidas las tareas de segmentación de clientes y análisis de respuesta. El propósito de la segmentación de clientes es la agrupación (*clustering*) de los clientes, mientras que el análisis de respuesta apunta al modelaje de la dependencia existente entre las medidas de marketing y el comportamiento de respuesta posterior. En la aplicación considerada en el presente trabajo, se ha realizado la segmentación de clientes utilizando la llamada agrupación difusa (*fuzzy clustering*), haciendo uso además de técnicas de redes neuronales y selección automática de atributos, para aumentar las tasas de respuesta.

---

### 3. Fundamentos de Tecnologías Inteligentes

---

En las siguientes sub-secciones se presenta tecnologías difusas y redes neuronales, respectivamente. Ambos métodos se utilizan para el *data mining* en el proyecto descrito en esta publicación.

#### **3.1 Tecnologías Difusas**

La idea básica de la teoría de conjuntos difusos (*fuzzy set theory*) es que a diferencia de la teoría clásica de conjuntos un elemento puede ser miembro de un conjunto con un grado de pertenencia que normalmente es un número entre 0 y 1 (Zadeh 1965). La noción de un conjunto difuso se define de la siguiente manera:

Sea  $X$  un conjunto convencional de elementos. Un conjunto difuso se define como  $\tilde{A}$  sobre  $X$  de tal manera que:

$$\tilde{A} := \{(x, \mu_{\tilde{A}}(x)) ; x \in X\}$$

La expresión  $\mu_{\tilde{A}}(x)$  se denomina grado de pertenencia del elemento  $x$  al conjunto difuso  $\tilde{A}$ , llamándose  $\mu_{\tilde{A}}$  la función de pertenencia, que normalmente asume valores entre 0 y 1.

A base de estas definiciones elementales de un conjunto difuso, en el pasado se han desarrollado muchas teorías y aplicaciones en el área de lógica difusa, un resumen de las cuales se puede encontrar en (Zimmermann 1996). Un método difuso, de especial interés para el presente trabajo, es el algoritmo de agrupación *fuzzy c-means*. Una presentación detallada del *fuzzy c-means* se encuentra en (Bezdek 1981), y las aplicaciones de la agrupación difusa se describen, por ejemplo, en (Meier et al. 1994; Strackeljan, Weber 1999).

El algoritmo *fuzzy c-means* asigna un conjunto de objetos, caracterizados por sus respectivos valores de atributos, a un número determinado de clases. Como resultado del *fuzzy c-means*, cada objeto tiene un grado de pertenencia a cada clase, representada por su centro de clase. Básicamente, el algoritmo *fuzzy c-means* se realiza aplicando los siguientes cuatro pasos (para más detalles véase, por ejemplo, Bezdek et al. 1999):

### Paso 1: Inicialización

Utilizamos la siguiente notación

- número de clases a encontrar:  $c$
- número de objetos a agrupar:  $J$
- vector de atributos del objeto  $j$ :  $x_j, \quad j = 1, \dots, J$
- grado de pertenencia del objeto  $j$  a clase  $i$ :  $\mu_{ij}, \quad i = 1, \dots, c; j = 1, \dots, J.$

Sea  $\Theta^{(0)}$  una matriz ( $c \times J$ ) con el elemento  $\mu_{ij}$  en posición  $(i,j), i = 1, \dots, c; j = 1, \dots, J.$

Esta matriz se inicializa en forma aleatoria con la siguiente restricción:

$$\sum_{i=1}^c \mu_{ij} = 1, \quad \forall j = 1, \dots, J$$

### Paso 2: Cálculo de Centros de Clase

Dados los valores de pertenencia  $\mu_{ij}$ , los centros  $v_i$  de cada clase  $i$  están dados por:

$$v_i = \frac{\sum_{j=1}^J (\mu_{ij})^m x_j}{\sum_{j=1}^J (\mu_{ij})^m}, \quad \forall i = 1, \dots, C$$

El parámetro  $m$ , utilizado en la formula anterior, se llama difusor (*fuzzifier*) y determina el grado de difusión (*fuzziness*) para las clases encontradas ( $1 < m < \infty$ ). Para  $m$  “cercano a 1” se calcula una solución con clases no-difusas (*crisp*); mientras mayor sea  $m$  más difusa se hace la solución.

### Paso 3: Actualización de valores de pertenencia

Dados los centros de clase calculados en el paso 2, los valores de pertenencia  $\mu_{ij}$  son actualizados utilizando la siguiente fórmula:

$$\mu_{ij} = \left( \sum_{k=1}^c \left( \frac{d_{kj}}{d_{ij}} \right)^{\frac{2}{m-1}} \right)^{-1} \quad \forall i = 1, \dots, c \\ \forall j = 1, \dots, J$$

El valor  $d_{ij}$  es la distancia entre el objeto  $j$  y el centro de clase  $i$  ( $v_i$ ). En el cálculo de esta distancia se utilizan los centros de clase  $v_i$  obtenidos en el paso 2.

### Paso 4: El Criterio de Detención (*Stopping*)

Los pasos 2 y 3 se repiten en forma iterativa hasta cumplir con el siguiente criterio de detención:

$$\|\Theta^{(t+1)} - \Theta^{(t)}\| \leq \epsilon$$

donde  $\Theta^{(t)}$  es la matriz de los valores de pertenencia en la iteración  $t$  y  $\epsilon$  es un umbral a ser determinado por el usuario. Las siguientes dos condiciones deben cumplirse para asegurar la convergencia del algoritmo *fuzzy c-means*:

$$\sum_{i=1}^c \mu_{ij} = 1, \quad \forall j = 1, \dots, J$$

$$\mu_{ij} \in [0, 1]; \quad i = 1, \dots, c; \quad j = 1, \dots, J$$

En términos de resultados, el algoritmo *fuzzy c-means* rinde centros de clase  $v_i$  para las  $c$  clases, así como los valores de pertenencia de cada objeto a cada clase  $\mu_{ij}$ .

### Evaluación de las clases encontradas

En la literatura se han propuesto los siguientes parámetros para evaluar la estructura de clase encontrada por la agrupación difusa (Windham 1981):

coeficiente de partición (cp):

$$\text{cp} = \sum_{j=1}^J \sum_{i=1}^C \frac{(\mu_{ij})^2}{J}$$

entropía de partición (ep):

$$\text{ep} = -\frac{1}{J} \sum_{j=1}^J \sum_{i=1}^C \mu_{ij} \log_e \mu_{ij}$$

exponente proporcional (exp):

$$\text{exp} = -\log_e \prod_{j=1}^J \sum_{i=1}^{m_k} (-1)^{i+1} \binom{c}{i} (1 - \mu_{ij})^{c-1}$$

El valor  $m_k$  se define como el entero más grande que sea menor o igual a  $\frac{1}{\mu_k}$ , donde

$$\mu_k := \max \{ \mu_{ik}; i = 1, \dots, c \}$$

Las siguientes relaciones se cumplen para los parámetros definidos anteriormente:

$$\frac{1}{c} \leq \text{cp} \leq 1 \quad 0 \leq \text{ep} \leq \log_e c \quad 0 \leq \text{exp} \leq \infty$$

En la aplicación de la segmentación de clientes que se describe en el capítulo 4.1, los clientes son los objetos considerados, que se describen mediante atributos tales como por ejemplo edad e ingreso. En este capítulo se mostrará también cómo los parámetros de evaluación pueden ser usados para encontrar un número adecuado de clases.

### **3.2 Redes Neuronales**

El término “red neuronal” (NN, por *Neural Network*) se usa para denotar modelos matemáticos de las funciones del cerebro humano, que pretenden expresar las propiedades del procesamiento paralelo masivo y de la representación distribuida existente en el cerebro. Las redes neuronales artificiales aprenden a partir de la experiencia, y es esta característica la que ha suscitado el interés actual en estos métodos. El aprendizaje ofrece una poderosa alternativa a la programación. Se han propuesto diferentes estrategias de aprendizaje para redes neuronales tales como el aprendizaje supervisado y no-supervisado (Rojas 1996).

Un ejemplo de los paradigmas de aprendizaje supervisado es el algoritmo de retropropagación (*backpropagation algorithm*) (Rojas 1996). En el aprendizaje supervisado la red se enfoca en un conjunto de entrenamiento representativo que consiste en pares vectoriales. Uno de los vectores se aplica a la entrada de la red, el otro se usa de objetivo, representativo de la salida deseada. El entrenamiento se logra al ajustar los pesos de las conexiones de la red para minimizar la diferencia entre las salidas deseadas y las salidas computadas.

El aprendizaje no supervisado, muchas veces llamado auto-organización, solamente requiere de vectores de entrada para adiestrar la red. Durante el proceso de entrenamiento, las ponderaciones de la red se ajustan a fin de que las entradas similares produzcan salidas similares. Esto se logra mediante el algoritmo de entrenamiento que extrae regularidades estadísticas desde el conjunto de entrenamiento. Los mapas de atributos auto-organizadores (*self-organizing feature maps*) introducidos por Kohonen pertenecen a la clase de algoritmos de aprendizaje no-supervisado (Kohonen 1995).

En la mayoría de los casos, una estructura bi-dimensional (mapa de atributo) se usa para redes Kohonen, sin embargo, también es posible encontrar estructuras uni-dimensionales (cadenas) o tri-dimensionales (cubos). Cada neurona tiene un vector referencial, que consiste en las ponderaciones  $\mu_{ik}$  entre las neuronas  $i$  y el vector de entrada  $k$  ( $i=1, \dots, N$ ;  $k=1, \dots, M$ ), donde  $N$  es el número de neuronas y  $M$  es el número de entradas. A continuación se presenta el algoritmo de Kohonen, así como sus parámetros de afinamiento.

1. Inicialización de todas las ponderaciones  $\mu_{ik}$  con valores aleatorios
2. Presentación al mapa de un vector de entrada  $e = (e_1, \dots, e_M)$
3. Cálculo del estado de excitación de cada neurona dado el vector de entrada  $e$
4. Búsqueda de la neurona ganadora (neurona con mayor excitación): neurona  $j$
5. Adaptación de las ponderaciones neuronales  $\mu_{ik}$ , de acuerdo a la siguiente formula:

$$\mu_{ik}^{new} = \mu_{ik}^{old} + \epsilon \cdot r_{ij} (e_k - \mu_{ik}^{old}) \quad i, j = 1, \dots, N, k = 1, \dots, M$$

6. Volver al paso 2.

donde,

$\mu_{ik}$ : ponderaciones neuronales;  $i = 1, \dots, N$ ;  $k = 1, \dots, M$

$r_{ij}$ : valores de retroalimentación entre la neurona ganadora (neurona  $j$ ) y la neurona  $i$  ( $i = 1, \dots, N$ )

$\epsilon$ : tasa de aprendizaje

Los valores de retroalimentación indican dependencias geométricas (*neighborhood dependencies*) entre las neuronas. Estas se calculan mediante una función gaussiana:

$$r_{ij} = e^{-\left(\frac{d^2}{\sigma^2}\right)}$$

donde

$\sigma$ : radio de aprendizaje

$d$ : distancia euclidiana entre la neurona ganadora (neurona  $j$ ) y la neurona  $i$ .

La tasa de aprendizaje  $\epsilon$ , junto con el radio de aprendizaje  $s$ , debe ser determinado por el usuario.

En el gráfico siguiente se presenta un mapa de atributo bi-dimensional con  $N$  neuronas. Entre la neurona  $i$  y el valor de atributo  $k$  de un objeto, se considera la ponderación  $\mu_{ik}$  ( $i = 1, \dots, N$ ;  $k = 1, \dots, M$ ).

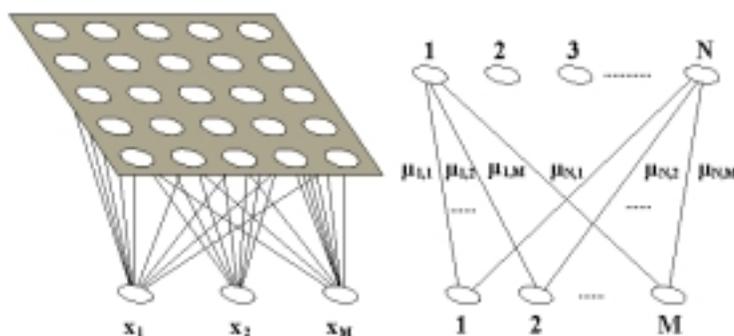


Gráfico 1: Mapa de atributo de Kohonen

En la aplicación descrita en el capítulo 4.2, se han usado mapas de atributos auto-organizadores para analizar las respuestas a *mailings*.

## 4. Aplicaciones de Data Mining en Instituciones Financieras

---

La siguiente sub-sección describe una aplicación de la agrupación difusa para la segmentación de clientes. Basándose en los segmentos encontrados, se aplican redes neuronales para un análisis de respuesta que está descrito en la sub-sección 4.2.

### **4.1 Segmentación de clientes utilizando la agrupación difusa**

---

La segmentación de clientes tuvo su origen en el campo de marketing e investigación de mercados, y entrega una división analítica de todos los potenciales clientes de un mercado de venta de acuerdo a diferentes criterios. Esto resulta en la formación de grupos de clientes, o segmentos de ellos, internamente homogéneos y externamente heterogéneos, planteando así actividades de marketing enfocadas a estos segmentos.

Tanto las instituciones financieras, como muchos departamentos comerciales, se han dado cuenta que un enfoque decisivo para el mejoramiento de la calidad de consulta y servicios reside en la segmentación de clientes en distintos grupos objetivos conformados por clientes similares (Angstenberger et al. 1998). Los productos y los servicios correspondientes a un segmento de clientes específico, deben ser entregados en la forma y al nivel en que el cliente perteneciente a este segmento espera y está dispuesto a pagar. Para aplicar tal enfoque se requiere conciencia de la relevancia de la segmentación orientada al cliente y la identificación de segmentos de clientes distintos mediante el uso de métodos adecuados (Weber 1996).

El análisis de agrupación tradicional no-difusa (*crisp cluster analysis*) es el método más usado tanto para la segmentación de mercado (Hruschka 1986), como para la segmentación de clientes.

Mediante el uso del análisis de agrupación, se descubrirán ciertos segmentos de clientes (del conjunto entero de todos los clientes). Cada cliente será asignado a una determinada agrupación, en forma concreta (no difusa) y completa. Estas agrupaciones se encuentran separadas la una de la otra por umbrales claros y precisos. Si los valores de atributos de cierto cliente sobrepasan estos umbrales, el cliente será asignado a otra agrupación. Si los valores de atributos de un cliente alcanzan exactamente un umbral determinado, este cliente es asignado opcionalmente a una u otra agrupación en forma completa, aún cuando él o ella pertenecería en el mismo grado en ambas agrupaciones.

Lo anterior revela las limitaciones del análisis tradicional de agrupaciones. La aplicación de técnicas de agrupación difusa aquí presentadas fue realizada para un banco alemán, cuyos productos eran, por ejemplo, cuentas corrientes, tarjetas de crédito y esquemas de inversión. La ventaja de la agrupación difusa en esta aplicación es que los clientes son asignados a diferentes segmentos con diferentes grados de pertenencia, lo que entrega un mejor conocimiento del perfil de pertenencia de cada cliente y, por lo tanto, sus necesidades en relación con los productos ofrecidos.

En el proyecto descrito en esta publicación, los datos específicos de clientes proporcionados por un banco alemán han sido analizados para la segmentación. En primer lugar, se seleccionarán los atributos para la segmentación y se analizarán en busca de correlaciones.

La selección de los atributos relevantes para la segmentación es de la mayor importancia para el análisis de agrupación, ya que este paso determina en forma definitiva qué criterios llevan a la agrupación final de los datos de clientes. Existe una enorme variedad de atributos disponibles, clasificables en atributos demográficos, tales como edad, género y situación familiar; y atributos socioeconómicos como educación, profesión, ingreso y propiedad. Los atributos finalmente seleccionados deberían mostrar un potencial discriminatorio suficiente sin correlación. Más aún, deben ser cuantificables (para la aplicación de métodos matemáticos), debiendo tener también el mismo nivel de escala. En el cuadro siguiente se presenta una matriz con los cinco atributos seleccionados en sus columnas. Cada fila representa un cliente. La base de datos completa, que se ha investigado, contiene alrededor de 300.000 clientes.

<b>Código cliente</b>	<b>Edad (años)</b>	<b>Ingreso (\$ por mes)</b>	<b>Propiedades (\$)</b>	<b>Crédito (\$)</b>	<b>Margen de contribución (\$ por año)</b>
101126	35	681,0	0,0	-1177,0	-5,0
607226	60	2218,0	6500,0	0,0	131,0
273974	53	537,0	6577,0	0,0	177,0
827866	56	4397,0	21333,0	0,0	-50,0
626508	38	2813,0	6667,0	0,0	-32,0

**Cuadro 1: Atributos seleccionados para algunos clientes**

El análisis de correlación debería comprobar que los atributos seleccionados son en gran medida independientes y que ningún atributo será sobreponderado en el análisis de agrupación.

	Edad	Ingreso	Propiedades	Crédito	Margen de C.
<i>Edad</i>	1	0,274	0,330	-0,024	0,282
<i>Ingreso</i>	0,274	1	0,437	-0,305	0,194
<i>Propiedades</i>	0,330	0,437	1	0,025	0,336
<i>Crédito</i>	-0,024	-0,305	0,025	1	-0,213
<i>Margen de C.</i>	0,282	0,194	0,336	-0,213	1

Cuadro 2: Resultado del análisis de correlación

Por otra parte, la normalización hace que los datos sean comparables, ya que originalmente los datos pueden tener diferentes dimensiones y diferentes rangos de escala, llevando (sin la normalización) a diferentes ponderaciones de atributos en el proceso de agrupación.

### Resultados y beneficios del análisis de datos difuso para la segmentación de clientes

Tras haber determinado los parámetros para el algoritmo *fuzzy c-means*, se modifica el número de agrupaciones. Para cada número de agrupaciones entre  $c = 2$  y  $c = 10$ , se aplica el algoritmo *fuzzy c-means*. Los valores de pertenencia de todos los clientes a las agrupaciones 1 a  $c$  son calculados y presentados. Adicionalmente, se entrega el centro de todas las agrupaciones (clases)  $c$ .

Clase	Edad	Ingreso	Propiedades	Crédito	Margen de C.
A	32,8	1.946,92	6.315,78	-4.509,91	21,92
B	59,28	1.951,87	9.518,03	-3.667,27	62,94
C	47,58	3.905,84	29.317,29	-13.816,90	171,15
D	10,45	135,03	2.607,43	-467,65	6,18
E	75,49	1.552,54	21.957,89	-1.983,58	203,71
F	41	3.921,11	12.661,52	-8.144,57	68,48

Cuadro 3: Centros de clases para  $c = 6$  agrupaciones

Las medidas de validez de agrupación, descritas anteriormente en la Sección 3.1, se usan para evaluar diferentes soluciones de agrupación. Estas medidas de validez de agrupación dan una pista matemática del “mejor” número de agrupaciones que pueden ser encontradas. Usando por ejemplo la entropía de partición (ep) el número más adecuado se determina como 6 en la aplicación demostrada (véase gráfico 2). Los resultados de la segmentación de clientes, descritos aquí, son verificados mediante entrevistas posteriores con expertos de la industria bancaria involucrados en este proyecto.

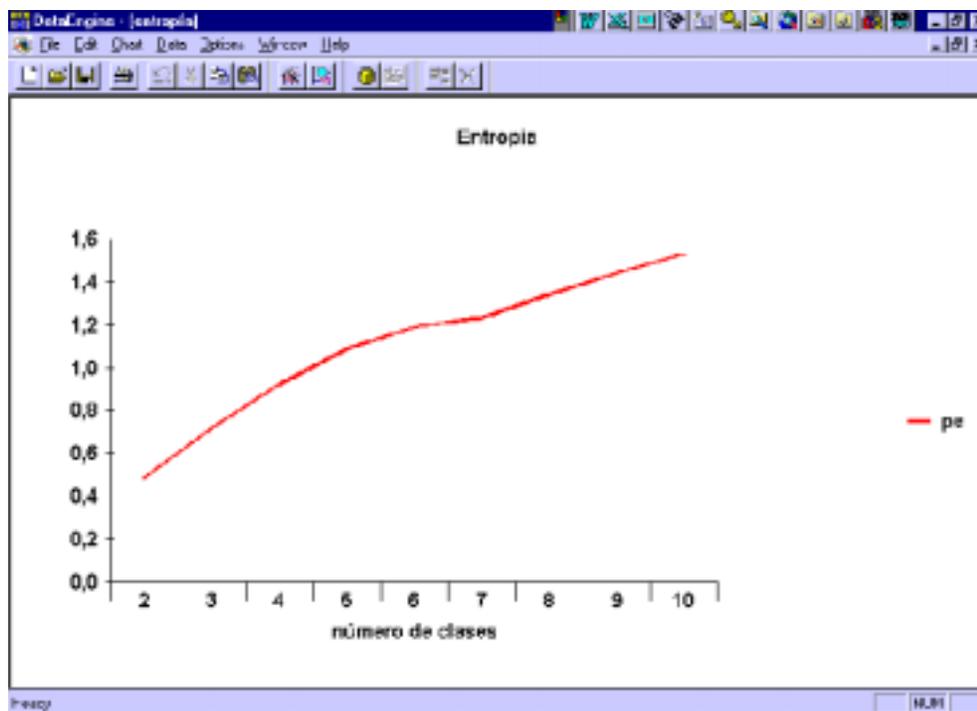


Gráfico 2: Entropía de partición para c=2 a c=10 agrupaciones

A diferencia del enfoque o perspectiva estática resultante de las técnicas tradicionales de agrupación, el análisis de agrupación difusa proporciona grados continuos de pertenencia de los objetos a las agrupaciones. Para cada cliente, se calcularán los valores de pertenencia en todas las clases, llevando a una posición más diferenciada de cada cliente entre las clases. Por ejemplo, los clientes que muestran valores característicos de diferentes segmentos se tratarán con actividades de marketing de acuerdo a sus valores de pertenencia. Como resultado, el experto de marketing del instituto financiero es capaz de ofrecer a sus clientes productos más acordes y los servicios correspondientes.

Además de esto, la agrupación difusa proporciona ventajas en cuanto a los aspectos dinámicos (Joentgen et al. 1999; Weber 1999). La evolución de un cliente a lo largo del tiempo se reconocerá en forma inmediata, a base del cambio de los valores de pertenencia a diferentes clases. Investigaciones periódicas de los datos

de clientes pueden ofrecer pistas precoces del movimiento de un cliente entre las clases y, como consecuencia, para actividades de marketing apropiadas.

## **4.2 Análisis de Respuesta**

Para uno de los segmentos de clientes encontrados, se ha realizado un análisis de respuesta a fin de incrementar la tasa de respuesta de las campañas de marketing realizadas para promocionar ciertos productos. En este caso específico, se consideran los *mailings* que promocionan las tarjetas de crédito.

En el banco correspondiente a este proyecto, la selección de personas a ser incluidas en un *mailing* se ha realizado hasta ahora en forma manual, configurando un perfil del cliente prototipo, y comparando cada registro de la base de datos con este prototipo mediante una consulta a la base de datos (*database query*). Aquellas personas que cumplen con los requerimientos pre-determinados son incluidas en el *mailing*. La composición manual del perfil a ser igualado introduce cierta arbitrariedad al proceso. Tomando en consideración la cantidad de información disponible sobre cada cliente, resulta evidente que es un problema no-trivial establecer un *query* para determinar clientes como, por ejemplo:

```
SELECT "Cliente"      FROM "Base de Clientes"  
  
WHERE "Edad < 25 AND Trabajo = Empleado AND Monto de crédito < 5.000"
```

Sería útil establecer automáticamente el perfil del cliente prototipo. A continuación se muestra cómo una red neuronal ha sido aplicada para realizar esta tarea. Se describe la elaboración de un clasificador neuronal utilizado para extraer de una base de datos aquellas personas que con más probabilidad adquirirán una tarjeta de crédito en respuesta a un *mailing*.

El análisis se basa en aproximadamente 9.000 registros de clientes seleccionados desde un sub-segmento especial denominado "jóvenes adultos". Los registros de datos incluyen datos personales, como edad, género y trabajo, así como información detallada sobre el uso de productos del cliente. Fueron considerados 18 productos distintos, desde cuentas corrientes y de ahorro, hasta planes de inversión y valores. En forma adicional, se dispuso de una clasificación del lugar de domicilio de cada persona incluida en la base de datos. En total, alrededor de 180 campos de datos (atributos) se encontraban disponibles para cada cliente en la base.

A partir de un *mailing* anterior, la respuesta de aproximadamente 2.500 personas ya era conocida (en términos de la compra o no del producto publicitado). La tasa de respuesta a este *mailing* anterior, calculado como la razón entre el número de compradores reales que responden al *mailing* y el volumen del mismo, fue de un 8,2%.

El análisis estadístico preliminar de los datos reveló que las distribuciones de los campos individuales de datos de los compradores eran muy similares a las de

los no-compradores. Por tanto, el perfil de un comprador típico no se pudo establecer a partir de las distribuciones estadísticas. Lo que significa que no era suficiente observar atributos aislados del cliente. Claramente, la configuración del perfil del cliente tenía que considerar la combinación de distintos *ítems* de información sobre el cliente. Sin embargo, la naturaleza de esta combinación de información era desconocida, además de qué información habría que incluir realmente.

Por esta razón, se decidió establecer la relación entre el perfil del cliente por un lado y su tendencia a comprar un producto por otro lado usando una red neuronal, ya que ésta tiene la capacidad de representar el tipo de relación no-lineal y no-homogénea requerida. Sin embargo, a fin de usar sólo los atributos más importantes, el siguiente paso del procedimiento de análisis consistió en reducir el espacio de atributos considerados. Esto también debería entregar una indicación de los atributos que realmente son relevantes.

En el Gráfico 3 se presenta un esbozo del procedimiento de análisis recién descrito. Los atributos más relevantes se seleccionan haciendo uso de una modificación del método descrito en Behr et al. (1997). Este procedimiento tiene la capacidad de considerar inicialmente un espacio de atributos de muchas dimensiones, para luego seleccionar las combinaciones de atributos que dan las tasas de re-clasificación más altas posibles. El procedimiento entrega un listado de combinaciones de atributos junto con una estimación de la tasa de re-clasificación para cada combinación.



Gráfico 3: Esquema del procedimiento de análisis global

La etapa de selección de atributos dio, como resultado, combinaciones de dos a cinco atributos, los que fueron analizados en forma más detallada. Las redes neuronales de tipo Kohonen (1995) (mapas de atributos auto-organizadoras) fueron aplicadas a un sub-conjunto de los datos de respuesta, siendo evaluadas mediante un conjunto distinto para cada una de las combinaciones de atributos seleccionados.

## **Resultados**

Los resultados fueron comparados con los resultados reales del *mailing*, calculándose las tasas de respuesta correspondientes. Estas se presentan en el siguiente cuadro.

<b>Numero de atributos</b>	<b>Tasa de respuesta</b>	<b>Tamaño del grupo objetivo seleccionado</b>
2	20.0%	350
3	18.3%	460
4	17.4%	530
5	17.6%	500

**Cuadro 4: Tasa de respuesta de los clasificadores neuronales**

Al comparar los resultados del análisis presentados en el Cuadro 4 con la tasa de respuesta del 8,2% obtenida anteriormente, resulta obvio que ésta fácilmente podría ser duplicada al realizar el proceso de selección en la forma descrita. Esto implicaría una reducción dramática de los costos del *mailing* y, al mismo tiempo, el volumen del *mailing* seleccionado por la red neuronal sería considerablemente menor que el seleccionado manualmente. Por lo tanto, se lograría un aumento de la eficiencia global de los *mailings*.

En la actualidad, el conocimiento adquirido en este análisis se está transfiriendo para lograr un mejor ajuste de los requisitos de otros productos distintos a la tarjeta de crédito considerada hasta ahora.

### **4.3 Herramientas de Software Utilizadas para Data Mining**

Las investigaciones recién descritas han sido llevadas a cabo utilizando la herramienta de software "DataEngine" un paquete de software que contiene métodos tales como tecnologías difusas y redes neuronales para el análisis inteligente de datos (MIT 2000). La combinación del pre-procesamiento, el análisis estadístico y los sistemas inteligentes para el diseño de un clasificador, junto con el modelaje del sistema lo convierte en un poderoso instrumento de software utilizable en un rango muy amplio de aplicaciones.

El siguiente gráfico da una visión global de la estructura del DataEngine. Su arquitectura abierta permite al usuario extender este programa de software mediante los llamados bloques de función definida por el usuario. Los modelos desarrollados con DataEngine fácilmente pueden ser integrados a otros paquetes de software y programas de aplicación utilizando la biblioteca DataEngine ADL.

Gráfico 4: Estructura de DataEngine

---

## 5. Conclusiones y Perspectivas Futuras

---

El presente trabajo entrega una breve presentación de enfoques de *data mining* y tecnologías inteligentes, así como sus aplicaciones al *database marketing* en un banco alemán. Los resultados son altamente prometedores, aunque basados en un pequeño conjunto de datos de clientes. El siguiente paso consistirá en aplicar los enfoques recién descritos a un conjunto de datos mayor, con el fin de validar su *performance*. Se pretende integrar los respectivos modelos de *data mining* al sistema de software utilizado para apoyar las actividades de marketing.

Las aplicaciones presentadas en la sección 4 pueden ser transferidas a diversas otras áreas en las que los clientes están siendo analizados, a fin de aumentar las ventas mediante el uso de estrategias optimizadas de marketing: por ejemplo, telecomunicaciones, empresas comerciales y la industria de ventas por catálogo (Bitran, Mondschein 1996). Entre otras aplicaciones parecidas donde el procedimiento de análisis aquí presentado puede ser aplicado, se incluye el área de detección de fraude. En vez de distinguir a los compradores de los no-compradores de la manera presentada en este trabajo, podría ser interesante diferenciar el uso fraudulento del uso correcto, por ejemplo, de tarjetas de crédito o teléfonos móviles. Tales aplicaciones quedan para investigaciones futuras.

---

## 6. Referencias bibliográficas

---

- Adriaans, P. y Zantinge, D. (1996): *Data Mining*. Addison-Wesley, Harlow.
- Anahory, S. y Murray, D. (1997): *Data Warehousing in the Real World*. Addison-Wesley, Harlow
- Angstenberger, J., Weber, R. y Poloni, M. (1998): "Data Warehouse Support to Data Mining: A Database Marketing Perspective". *Journal of Data Warehousing* 3, N° 1, 2-11.
- Behr, D., Kocher, T. y Strackeljan, J. (1997): "Fuzzy Pattern Recognition for Automatic Detection of Different Teeth Substances". *Fuzzy Sets and Systems* 85, 275-286.
- Bezdek, J.C. (1981): *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, Nueva York.
- Bezdek, J.C., Keller, J.M., Krishnapuram, R. y Pal, N.R. (1999): *Fuzzy models and algorithms for pattern recognition and image processing*. Kluwer, Boston Londres Dordrecht
- Bitran, G. R. y Mondschein, S. V. (1996): "Mailing Decisions in the Catalog Sales Industry." *Management Science* 42, N° 9, 1364-1381.
- Famili, A., Shen, W.-M., Weber, R., Simoudis, E. (1997): Data Preprocessing and Intelligent Data Analysis. *Intelligent Data Analysis* Vol. 1, N° 1, 3-23.
- Fayyad, U. M. (1996): "Data Mining and Knowledge Discovery: Making Sense out of Data." *IEEE Expert, Intelligent Systems & Their Applications*, Octubre 1996, 20-25.
- Frawley, W. J., Piatetsky-Shapiro, G. y Matheus, C. J. (1991): "Knowledge Discovery in Databases: An Overview". En: Piatetsky-Shapiro, G. y Frawley, W. J. (eds.): *Knowledge Discovery in Databases*. AAAI Press/MIT Press, Menlo Park, 1-27.
- Furness, P. (1994): "New Pattern Analysis Methods for Database Marketing; Part 1". *The Journal of Database Marketing* 1, N° 3, 220-232.
- Guynes, C. S., Prybutok, V. R. y Myers, B. L. (1996): "Evolving Data Quality Considerations for Client/Server Environments". *Data Quality* 2, No.1, 21-27.
- Hruschka, H. (1986): "Market definition and segmentation using fuzzy clustering methods." *International Journal of Research in Marketing* 3, 117-134.
- Joentgen, A., Mikenina, L., Weber, R. y Zimmermann, H.-J. (1999): Dynamic Fuzzy Data Analysis Based on Similarity Between Functions. *Fuzzy Sets and Systems* 105, 1, 81-90
- Kohonen, T. (1995): *Self-Organizing Maps*. Springer-Verlag, Berlin Heidelberg.
- Meier, W., Weber, R. y Zimmermann, H.-J. (1994): "Fuzzy Data Analysis - Methods and Industrial Applications." *Fuzzy Sets and Systems* 61, 19-28.
- MIT (2000): *DataEngine 3.1 - Manual*. Aachen, Alemania.
- Nash, E. L. (1993): *Database Marketing - The Ultimate Marketing Tool*. McGraw-Hill, Nueva York.
- Rojas, R. (1996): *Neural Networks - A Systematic Introduction*. Springer-Verlag, Berlin Heidelberg.
- Strackeljan, J. y Weber, R. (1999): "Quality Control and Maintenance". En: H.-J. Zimmermann (ed.): *Practical Applications of Fuzzy Technologies, Vol. 7 of The Handbooks of Fuzzy Sets Series*, D. Dubois, H. Prade (series editors), Kluwer Academic Publishers, Boston, Londres, Dordrecht, 161-184

- Weber, R. (1996): "Customer Segmentation for Banks and Insurance Groups with Fuzzy Clustering Techniques". En: J. F. Baldwin (ed.): *Fuzzy Logic*, John Wiley and Sons, Chichester, 187 – 196.
- Weber, R. (1999): Dynamic Data Mining with Functional Fuzzy Clustering. *Actas de Resúmenes Extendidos, III Congreso Chileno de Investigación Operativa - OPTIMA '99*, Arica, Chile, 13-15 de octubre de 1999, 217-224
- Windham, M. P. (1981): "Cluster Validity for Fuzzy Clustering Algorithms". *Fuzzy Sets and Systems* 5, 177-185.
- Zadeh, L. A. (1965): "Fuzzy Sets". *Information and Control* 8, 338-353.
- Zimmermann, H.-J. (1996): *Fuzzy Set Theory - and Its Applications*. 3er edición. Kluwer Academic Publishers, Boston Dordrecht, Londres.