



Asignatura: **CI5128 – Ingeniería del conocimiento**
Cuatrimestre: **2º** Examen: **Parcial** Convocatoria: **Ordinaria**
Grupo: **5IT1** Curso: **2008/2009** Fecha: **12-3-2009**

ALUMNO: _____

NOTA: Se puede usar todo tipo de libros y apuntes de manera personal e intransferible.

Se dispone de cierta información sobre varios mercados empresariales, y sobre las empresas que compiten en dichos mercados. En concreto, la información de la que se dispone está contenida en una base de datos que contiene las siguientes tablas:

1. Sobre las empresas:

- Identificador (ID-E): tipo entero
- Capital (C): tipo entero, en millones de euros
- Nivel tecnológico (NT): puede ser bajo, medio, o alto
- Gastos publicidad (GP): puede ser bajo, medio, o alto

2. Sobre los productos:

- Identificador (ID-P): tipo entero
- Nombre de venta del producto (NP): una cadena de caracteres
- Identificador de Mercado (ID-M): entero que identifica el mercado en el que se encuadra este producto
- Identificador de Empresa (ID-E): entero que identifica la empresa que fabrica este producto

3. Sobre los mercados (todos los parámetros de los mercados se consideran constantes en el tiempo):

- Identificador (ID-M): entero
- Nombre del mercado (NM): sólo hay dos mercados: tomates y lechugas
- IPC: tipo real, indicando el porcentaje medio anual (tanto por ciento) de incremento del precio del producto en los últimos 10 años
- Nivel competitivo (NC): puede ser muy agresivo, agresivo, normal, suave

4. Valores: indica ciertos valores de los productos

- Identificador del producto (ID-P)
- Precio de venta al público (PVP): valor entero (en euros)
- Coste del producto (C): valor entero (en euros)
- Trimestre del año (T): primer, segundo, tercer, y cuarto trimestres. Indica a qué trimestre del año corresponden los valores del coste y del trimestre
- Percepción Cliente (PC): caro, barato, normal. Especifica la percepción que el cliente tuvo sobre si el producto era caro, barato, o tendrá un precio normal. Este valor es obtenido a través de encuestas.

En las tablas 1,2,3 y 4 se muestran algunos de los registros contenidos en la base de datos:

EMPRESA

ID-E	C	NT	GP
1	3	alto	bajo
2	2	medio	alto
3	3	alto	bajo
4	2	alto	alto

Tabla 1: Datos de las empresas

PRODUCTOS

ID-P	NP	ID-M	ID-E
1	Tomates Caracol	1	1
2	Tomates de la Huerta	1	2
3	Lechugas Caracol	2	1
4	Lechugas Garrido	2	3

Tabla 2: Datos de los productos

MERCADOS

Id-M	NM	IPC	NC
1	tomates	4	alto
2	lechugas	2	medio

Tabla 3: Datos de los mercados

VALORES

ID-P	PVP	C	T	PC
1	2,1	0,9	primer	barato
1	1,7	0,3	segundo	normal
1	1,1	0,2	tercer	caro
1	1,4	0,25	cuarto	caro
2	2,2	0,95	primer	barato
2	1,6	0,2	segundo	caro
2	1,0	0,15	tercer	normal
2	1,6	0,4	cuarto	barato
3	3,1	1,95	primer	caro
3	2,6	1,25	segundo	barato
3	2,4	1,19	tercer	caro
3	2,5	1,4	cuarto	normal
4	2,2	1,5	primer	barato
4	1,6	0,8	segundo	caro
4	1,4	0,7	tercer	barato

Tabla 4: Valores de los productos

- 1) **(2,5 puntos)** Discretiza el atributo PVP de la tabla 4 en dos intervalos según el cambio de clase. Para ello utiliza sólo para los ejemplos que sean tomates (ID-P 1 o 2).
- 2) **(0,5 puntos)** Discretiza el atributo C de la tabla 4 en dos intervalos iguales. En este caso, hay que tener en cuenta todos los valores de la tabla 4.
- 3) **(2,5 puntos)** ¿Sería percibido por los usuarios como caro un producto “Lechugas Garrido” que se vendiera en el cuarto trimestre a 2,4 euros, y que se fabrica a 1,5 euros? Para ello aplica el algoritmo Naive Bayes únicamente a la tabla 4.
- 4) **(2,5 puntos)** Aplica el algoritmo ID3 sólo teniendo en cuenta la tabla 4, y en particular solamente los ejemplos con ID-P 1 y 2.
- 5) A partir de dicha base de datos, y teniendo en cuenta todos los datos, responde a las siguientes preguntas:
 1. **(0,5 puntos)** ¿Cómo construiríamos el conjunto de entrenamiento? Escribe la descripción completa de un posible ejemplo del conjunto de entrenamiento.
 2. **(1,5 puntos)** Responde de forma precisa: ¿Cómo evaluarías si un árbol de decisión construido mediante el algoritmo C4.5 aplicado a esta base de datos es un buen modelo? ¿Qué dos medidas nos darían una idea cuantitativa y cualitativa de lo bueno que es? ¿Cómo calcularías cada uno de ellas?