

# PRÁCTICA: ÁRBOLES DE DECISIÓN

## PARTE I: (Python/WEKA)

Tome los datos de ejercicio de teoría en el que se discretizaba un valor de la presión. Reproduzca en Python los resultados obtenidos mediante árboles de decisión. Para ello, simplemente construya los árboles gráficamente.

## PARTE II: (WEKA)

En el siguiente enlace del repositorio de la UCI:

<http://archive.ics.uci.edu/ml/datasets/Thoracic+Surgery+Data>

Se encuentra un ejemplo de clasificación de pacientes relacionados con enfermedades torácicas, en particular, con el riesgo a un año de padecerlas.

Responder a las siguientes cuestiones:

1. Por qué no se puede aplicar directamente ID3.
2. Pase el algoritmo J48 aplicando el método de retención o Hold-Out. Analice el árbol obtenido sin poda. ¿Se podría prescindir de algún atributo? Si es así, hágalo y compare los resultados de nuevo.
3. Habrá notado que cuando se usa algún atributo numérico, implícitamente se aplica una discretización al plantear las diferentes ramas del árbol a partir de él. ¿Por qué es más eficiente esta técnica que la aplicada en PARTE I?
4. Plantee, entonces, una discretización basada en el punto anterior, aunque no resulte ser binaria, sino en tantos tramos como induzcan los valores usados al formar las ramas del árbol con J48 sin poda (Hold-Out). Introduzca este fichero de nuevo al algoritmo ID3. Compare los resultados con el J48 de 2.