

MÁSTER UNIVERSITARIO EN
LÓGICA, COMPUTACIÓN E INTELIGENCIA ARTIFICIAL
Aprendizaje Automático

Apellidos:

Nombre :

En este cuestionario se espera que el alumno profundice en algunos aspectos de la implementación de los árboles de decisión **que no se han visto en clase**. Se espera que el alumno presente **el mejor análisis posible** dependiendo de su formación básica y de su experiencia con sistemas de aprendizaje automático.

Para ello, cada alumno debe elegir una base de datos del repositorio de la Universidad de California (accesible en <https://archive.ics.uci.edu/ml/datasets.php>) y realizar sobre él el siguiente estudio.

1. Reproducir los pasos que se han visto en clase sobre el conjunto iris, en particular:
 - a) Carga del fichero en python y probablemente, eliminación de las filas que no tengan todos los datos. Esto último no se ha visto en clase. Se sugiere mirar la documentación de la librería *pandas*.
 - b) Realiza un pequeño examen exploratorio creando la *scatter matrix* y el mapa de calor (*heatmap*) de los coeficientes de correlación de Pearson entre las variables.
 - c) Divide la base de datos en conjunto de entrenamiento y prueba y crea el correspondiente árbol de decisión sobre el conjunto de entrenamiento.
 - d) Representa gráficamente el árbol obtenido.
 - e) Encuentra la media de rendimiento (*score*) del árbol sobre los conjuntos de entrenamiento, prueba y total.
2. La parte realmente interesante de este cuestionario empieza ahora, donde se espera que el alumno demuestre su **autonomía** y **capacidad de profundización** en la materia a partir de los conceptos básicos. Se espera que el alumno continúe con el análisis del conjunto de datos y presente el estudio **más completo posible**. ¿Qué otros métodos podríamos utilizar para obtener el *mejor* árbol de decisión posible?

Nota: Los métodos *ensemble* basados en árboles de decisión (*Random forest*, *Gradient boosted*, etc) se estudiarán en la asignatura *Inteligencia Artificial para la Ciencias de los Datos* del segundo cuatrimestre.