

# Practica 1. Titanic

En esta práctica vamos a encontrar el mejor clasificador para los datos del Titanic publicados en Kaggle (<https://www.kaggle.com/c/titanic>), que también podemos encontrarlos en el Campus Virtual.

La forma en la que vamos a abordar el problema es la siguiente:

En primer lugar, realizaremos un examen del dataset y estudiaremos los atributos y si aparecen datos faltantes. Como final de la fase de preprocesamiento obtendremos un dataset en el que se hayan eliminado los atributos innecesarios y se haya solucionado los atributos faltantes.

En segundo lugar, entrenaremos a cada uno de los clasificadores (Rpart, nnet, e1071) usando validación cruzada. Obtendremos el accuracy y el área bajo la curva para cada clasificador.

En tercer lugar, vamos a entrenar a cada uno de los clasificadores, usando validación cruzada y modificando sus parámetros como CP, size o el tipo de kernel, con el objetivo de encontrar el clasificador mejor de su clase.

Por último, obtendremos el mejor clasificador de todos y realizaremos una predicción con el mismo usando los datos de test de Titanic.