

## Trabajo Práctico Final: Titanic

El siguiente consiste en uno de los ejercicios más clásicos de las competencias de “Machine Learning”. Para más ejercicios de esta naturaleza, se sugiere acceder al sitio “Kaggle” en donde se efectúan diversas competencias de análisis de datos similares a los propuestos por este ejercicio.

El RMS Titanic fue en su momento el mayor barco de pasajeros del mundo, hundiéndose en su viaje inaugural de Southampton a Nueva York en el año 1912. En el evento perecieron 1514 de las 2223 personas que iban a bordo, entre tripulación y pasajeros. El objetivo es determinar un mecanismo de clasificación que en función de diversos datos de cada pasajero (clase, edad, sexo, entre otros atributos) pueda determinar si el pasajero sobrevivió o no a la catástrofe.

El archivo “titanicAlumno.txt” contiene una muestra de pasajeros, los cuales deberán usar para entrenar su mecanismo de clasificación. Posteriormente, se evaluará la performance del clasificador propuesto con un conjunto de datos “secretos” que conserva la cátedra. A continuación, ofrecemos un listado de los atributos presentes:

- survival: supervivencia (0 No, 1 Sí).
- pclass: clase del pasajero (1,2 o 3).
- name: nombre del pasajero (texto).
- sex: sexo del pasajero (“male”, “female”).
- age: edad del pasajero (número, ojo que hay datos faltantes).
- sibsp: cantidad de hermanos y cónyuges (totalizado) embarcados (número entero).
- parch: cantidad de padres e hijos (totalizado) embarcados (número entero).
- ticket: código del boleto (texto).
- fare: tarifa del pasaje (número real).
- cabin: cabina (texto).
- embarked: puerto de embarque (S= Southampton, Q=Queenstown, C = Cherbourg).

Para la entrega, se deberá tener un archivo con el código del “clasificador”. El mismo será evaluado con los datos “secretos” utilizando el validador que se encuentra en el archivo **validacionTitanic.r**. Se tomará en cuenta la capacidad general predictiva (tasa de aciertos general) y el puntaje F. Se armará un “ranking” con ambos resultados.

El informe final deberá constar con una explicación de la metodología utilizada para la clasificación y un estudio de los resultados obtenidos con la misma.

Es importante que su clasificador sea compatible con el validador en **validacionTitanic.r**, procure probarlo con el mismo antes de entregarlo.