

Práctica 2 (25% nota final)

1. Descripción del dataset. ¿Por qué es importante y qué pregunta/problema pretende responder?

Para la realización de esta práctica contamos con dos conjuntos de datos parecidos que contienen información acerca de los pasajeros que viajaban a bordo del Titanic cuando naufragó. Estos datasets se llaman 'train.csv' y 'test.csv'. El objetivo es conseguir predecir si los individuos del conjunto de test sobreviven a partir de los patrones que hayamos detectado en el conjunto de entrenamiento.

Train.csv cuenta con información de 891 pasajeros acerca de las siguientes variables:

- PassengerId: identificador único, type Integer.
- Survived: variable objetivo, 0 = No sobrevivió y 1 = sobrevivió, type Integer.
- Pclass: clase del ticket, 1 = 1ª clase, 2 = 2ª clase y 3 = 3ª clase, type Integer.
- Name: nombre del pasajero, type string.
- Sex: sexo, type string.
- Age: edad, type Integer.
- Sibsp: número de familiares cercanos a bordo, donde se incluyen marido/mujer, hermanos y hermanastros.
- Parch: número de padres/hijos a bordo, para niños que viajaban con una niñera se les asigna el valor 0, type Integer.
- Ticket: número del ticket, type string.
- Fare: tarifa del pasajero, type float.
- Cabin: número de cabina, type string.
- Embarked: puerto de embarque, C = Cherbourg, Q = Queenstown y S = Southampton, type string.

El conjunto de test cuenta con 418 registros y todas las variables del conjunto de entrenamiento a excepción de Survived.

El resto de preguntas de la práctica basadas en la limpieza y el análisis del conjunto de datos están resueltas en el notebook del repositorio.

Contribuciones	Firma
Investigación previa	EPM, MLM
Redacción de las respuestas	EPM, MLM
Desarrollo del código	EPM, MLM