

Titanic - Modelo árbol decisión

¿qué atributos/features elegimos para aprender y poder decidir si un pasajero sobrevivió o no?

De todos los atributos el nombre del pasajero no nos sirve para generalizar podríamos usar edad, clase y sexo

Entrenar un modelo

El objetivo es hacer una función que pasándole un pasajero nos diga si sobrevivió o no
¿cómo generamos un árbol de decisión a partir de los datos de entrenamiento?

Podemos utilizar como criterio de clasificación la **entropía** que mide el grado de incertidumbre que tengo, de forma que al construir el árbol de decisión lo hace viendo qué atributos van a ordenar mejor el conjunto de datos para conseguir una menor entropía y de esta forma reducir el grado de incertidumbre. Esto es lo que se llama ganancia de información, de todos los atributos elegir el que mejor va a ordenar el conjunto de datos

arbol de 3 niveles/max_depth=3
con al menos 5 nodos hoja/min_samples_leaf=5

```
from sklearn import tree

clf = tree . DecisionTreeClassifier(criterion = 'entropy' , max_depth =3 ,
min_samples_leaf =5 )

clf = clf . fit(X_train,y_train)
```