

Lectura 6

Construir y evaluar un árbol de decisión

SI3015 - Fundamentos de Aprendizaje Automático

2026

1. Ejercicio en clase:

Descargar uno de estos dos datasets:

Link dataset de películas

Link dataset del Titanic

Nota: también puede probar con un conjunto de datos propio en el que pueda aplicar las actividades a continuación.

Cargar el conjunto de datos y explorelo gráficamente para realizar las siguientes actividades:

- Limpie y/o transforme el conjunto de datos.
- Determine las columnas de características y la columna objetivo.
- Divida el dataset en entrenamiento y prueba.
- Defina el pipeline.
- Entrene el modelo aplicando **Random Forest** (bagging) y **Gradient Boosting**.
- Obtenga y compare las métricas de desempeño (matriz de confusión, accuracy, precision, recall, F1 score) para Random Forest y Gradient Boosting.

2. Sugerencias:

- Cerciorese de realizar la mayor limpieza, exploración y transformación posible del conjunto de datos. Recuerde que un conjunto de datos suficientemente limpio y transformado aumenta significativamente la probabilidad de obtener muy buenos resultados al aplicar el modelo de aprendizaje automático.
- Recuerde basarse en los ejemplos `ej_arboles.ipynb` para realizar el ejercicio.