

Laboratorio 2: Árboles y Random Forest para regresión

Se recomienda hacer la práctica en grupos de 4 o 5 alumnos.

Objetivo

Aplicar al conjunto de datos proporcionado los métodos o algoritmos Árboles y Random Forest para predecir (regresión) el número de ventas de bicicletas por hora.

Pasos a seguir (orientativo)

- Análisis descriptivo de los datos
- Determinar el conjunto de modelización y el de validación
- Tratamiento de missing (si los hay)
- Tratamiento de variables categóricas
- Calcular las métricas de evaluación de ajuste adecuadas
- Comparar mediante las medidas que parezcan adecuadas la capacidad predictiva de ambos métodos
- Comentar los resultados obtenidos
- Otros comentarios que parezcan adecuados
- Repite el ejercicio haciendo el siguiente grupo y desarrollando un algoritmo de clasificación: grupo 1 si $cnt \leq 20$, grupo 2 en otro caso.

Información sobre los datos

El enlace: <https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset> contiene los datos, la variable respuesta es “cnt”.

Comentarios sobre la evaluación

- Máximo de 3 páginas
- Se puede usar R o Python
- Se deben comentar los resultados obtenidos y el código

Deadline

Los resultados se tendrán que entregar antes del 15 de Febrero por la intranet.