

Tarea 2

Problema

Considere al conjunto de datos QSAR¹, un juego de datos con 1025 atributos, 1024 de entrada y uno de salida, todos binarios.

Se pide:

- a) Implemente el algoritmo ID3 visto en el teórico. Evalúe sus resultados sobre QSAR.
- b) Implemente el algoritmo Random Forest, utilizando como base su implementación de la parte (a), realizándole las modificaciones necesarias. Evalúe sus resultados sobre QSAR.
- c) Aplique las implementaciones de Scikit-learn para (a) y (b) sobre QSAR y compare los resultados con los de sus implementaciones.

En las partes (a) y (b) se podrá utilizar pandas y scikit-learn para la carga del dataset y la generación de archivos de entrenamiento, testeo, etc.

Entregables

- Informe con las pruebas realizadas y los resultados obtenidos.
- El informe a entregar debe ser un Jupyter Notebook.
- Código escrito para resolver el problema.

Fecha límite de entrega

Lunes 13 de abril (inclusive)

¹ <https://archive.ics.uci.edu/ml/datasets/QSAR+oral+toxicity>