

Práctica: Detección de anomalías

Objetivo

Aplicar e interpretar los resultados de un algoritmo de detección de anomalías (se recomienda isolation forest) aplicado sobre el conjunto de datos. El objetivo es clasificar la existencia de tumor o no.

Pasos a seguir (orientativo)

- Análisis descriptivo de los datos
- Determinar el conjunto de modelización y el de validación
- Tratamiento de missing si se considera pertinente
- Calcular las métricas de evaluación de ajuste adecuadas
- Comentar los resultados obtenidos
- Otros comentarios que parezcan adecuados

Información sobre los datos

El enlace: <https://archive.ics.uci.edu/ml/datasets/Breast%2BCancer%2BWisconsin%2B%28Diagnostic%29> contiene la información necesaria sobre los datos.

El objetivo es que sin utilizar la información de clase (diagnosis) ni el id se clasifique la muestra como tumor o no tumos.

Comentarios sobre la evaluación

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

.