

Conjunto de datos

Para esta actividad se utilizará un dataset reducido en características con el objetivo de entrenar modelos para predecir si un paciente tiene riesgo de cardiopatía. En el mismo se almacenan medidas de diagnóstico sobre pacientes y una variable de salida indicando el riesgo alto o bajo.

Práctica

Utilizando Python, sin recurrir a otros paquetes externos que no sean los vistos en la cátedra:

- 1) Realizar un análisis exploratorio de datos y presentar los resultados más importantes.
- 2) Preprocesar el conjunto de datos para que pueda utilizarse en los distintos tipos de modelos vistos en clase.
Tener en cuenta si es necesario:
 - Análisis de las variables.
 - Valores erróneos/faltantes, identificación de variable objetivo.
 - Valores atípicos.
 - Correlación entre variables.
 - Procesamiento de variables categóricas.
 - Balance del conjunto de datos.
 - Normalización del conjunto de datos.
- 3) Entrenar y mostrar los resultados obtenidos con al menos dos de los modelos vistos.
Tener en cuenta si es necesario:
 - División de conjunto de datos para entrenamiento y prueba.
 - Instanciación de modelos.
 - Entrenamiento del modelo.
 - Evaluación de modelos (Accuracy y F1-Score).

Entregar

- 1) Un archivo .py o .ipynb conteniendo la solución a los incisos 1, 2 y 3.
- 2) Un informe que documente y justifique los distintos pasos realizados, resultados obtenidos, herramientas y técnicas utilizadas.
- 3) El grupo deberá exponer en un coloquio con ayuda de una presentación, de no más de 10 minutos, del trabajo realizado. Luego de la exposición se realizarán preguntas.

La fecha y hora del coloquio será publicada oportunamente.