

Autor: Alfredo Rubio Navarro

1. Descripción del dataset.

¿Por qué es importante y qué pregunta/problema pretende responder?

El dataset elegido para la práctica estudia la presencia o ausencia de enfermedad cardíaca basándose en los datos de 1026 pacientes de cuatro centros hospitalarios distintos: Cleveland, Hungary, Switzerland, and Long Beach V: <https://www.kaggle.com/johnsmith88/heart-disease-dataset/data>

La pregunta que intenta responder es: ¿Cuáles son los factores que debemos tener en cuenta a la hora de clasificar un nuevo paciente con síntomas de enfermedad coronaria?

Está claro que cuanto antes se diagnostique correctamente y se trate al paciente, mayores posibilidades tiene de recuperarse o de no sufrir secuelas.

El conjunto de datos incluye sujetos sanos y pacientes con enfermedades cardíacas, de 34 a 77 años. El dataset contiene un total de 14 características clínicas para cada caso.

Los atributos del dataset son:

1. age: Edad del paciente.
2. sex: Sexo.
3. cp: Tipo de dolor en el pecho.
4. trestbps: Presión sanguínea en reposo al ser admitido en el hospital.
5. chol: Nivel de colesterol en sangre en mg/dl.
6. fbs: Nivel de azúcar en sangre en ayunas > 120 mg / dl.
7. restecg: Resultados electrocardiográficos en reposo.
8. thalach: Frecuencia cardíaca máxima alcanzada.
9. exang: Angina inducida por el ejercicio.
10. oldpeak: Depresión onda ST inducida por el ejercicio relativo al descanso.
11. slope: Pendiente del segmento ST del ejercicio pico.
12. ca: Número de vasos principales (0-3) coloreados por flourosopía.
13. thal: Exploración cardíaca de talio.
14. target: Diagnóstico de enfermedad cardíaca.

2. Integración y selección de los datos de interés a analizar.
3. Limpieza de los datos.
 - 3.1. ¿Los datos contienen ceros o elementos vacíos? ¿Cómo gestionarías cada uno de estos casos?
 - 3.2. Identificación y tratamiento de valores extremos.
4. Análisis de los datos.
 - 4.1. Selección de los grupos de datos que se quieren analizar/comparar (planificación de los análisis a aplicar).
 - 4.2. Comprobación de la normalidad y homogeneidad de la varianza.
 - 4.3. Aplicación de pruebas estadísticas para comparar los grupos de datos. En función de los datos y el objetivo del estudio, aplicar pruebas de contraste de hipótesis, correlaciones, regresiones, etc. Aplicar al menos tres métodos de análisis diferentes.
5. Representación de los resultados a partir de tablas y gráficas.
6. Resolución del problema. A partir de los resultados obtenidos, ¿cuáles son las conclusiones? ¿Los resultados permiten responder al problema?