

# 1. Descripción del dataset

El dataset a tratar en la práctica es

<https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset>.

Se trata de un dataset con 14 variables, las cuales están asociadas a pacientes con síntomas coronarios y cuyo objetivo es la predicción de la probabilidad (mayor o menor) de tener un ataque al corazón bajo esos síntomas. El dataset es de mucha relevancia, dado que poder entrenar un modelo que anticipe esta circunstancia podría ayudar a salvar muchas vidas.

El dataset consta de las siguientes variables:

- age : Edad del paciente
- sex: Sexo del paciente codificado como 0 o 1. Se desconoce su traducción a Hombre o Mujer.
- cp : Chest Pain type chest pain type. Tipo de dolor en el pecho. Puede tomar los valores:
  - Valor 1: typical angina
  - Valor 2: atypical angina
  - Valor 3: non-anginal pain
  - Valor 4: asymptomatic
- trtbps : resting blood pressure (in mm Hg). Presión sanguínea en reposo.
- chol : cholesterol in mg/dl fetched via BMI sensor. Colesterol en sangre.
- fbs : (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false). Azúcar en sangre en ayunas por encima de 129 mg/dl. Codificado como 1 Verdadero, 0 Falso.
- restecg : resting electrocardiographic results. Resultados del electrocardiograma en reposo. Puede tomar los siguientes valores:
  - Valor 0: normal
  - Valor 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
  - Valor 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
- thalachh : maximum heart rate achieved. Frecuencia cardíaca máxima.
- exng: "exercise induced angina". Angina inducida por el ejercicio. Codificado como 1 "Si" 0 "No".
- caa: number of major vessels. Número de vasos sanguíneos mayores. Codificado de 0 a 3
- target : Variable objetivo. 0 = menor posibilidad de ataque al corazón 1 = mayor posibilidad de ataque al corazón.

El dataset consta de otras tres variables que no están descritas y no conocemos su significado (oldpeak, slp y thall) que no utilizaremos por prudencia.

A continuación, vamos a visualizar los primeros datos del dataset

```
In [84]: import pandas as pd
```

```
In [85]: df = pd.read_csv("./datos/heart.csv")
```

```
In [86]: df.head()
```

```
Out[86]:
```

	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Borramos las columnas que no vamos a utilizar

```
In [87]: df = df.drop(["oldpeak", "slp", "thall"], axis=1)
```

## 2. Integración y selección

Vamos a integrar los ficheros que contienen las descripciones de los campos categoricos del dataset, de manera que sea más fácil su interpretación y trabajo con los datos.

Después de la integración de cada fichero, borraremos las columnas originales de cruce, para dejar sólo la nueva columna descriptiva

### Exang

```
In [88]: df_exang = pd.read_csv("./datos/exang.csv")
```

```
In [89]: df = df.merge(df_exang, how="left", left_on="exng", right_on="id_exang").drop(["
```

### Chest pain

```
In [90]: df_cp = pd.read_csv("./datos/chest_pain.csv")
```

```
In [91]: df = df.merge(df_cp, left_on="cp", right_on="id_cp", how="left").drop(["cp", "id
```

Visualizamos el dataset final

```
In [92]: df
```

Out[92]:

	age	sex	trtbps	chol	fbs	restecg	thalachh	caa	output	desc_exang	desc_cp
<b>0</b>	63	1	145	233	1	0	150	0	1	no	non-anginal pain
<b>1</b>	37	1	130	250	0	1	187	0	1	no	atypical angina
<b>2</b>	41	0	130	204	0	0	172	0	1	no	typical angina
<b>3</b>	56	1	120	236	0	1	178	0	1	no	typical angina
<b>4</b>	57	0	120	354	0	1	163	0	1	yes	NaN
<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>
<b>298</b>	57	0	140	241	0	1	123	0	0	yes	NaN
<b>299</b>	45	1	110	264	0	1	132	0	0	no	non-anginal pain
<b>300</b>	68	1	144	193	1	1	141	2	0	no	NaN
<b>301</b>	57	1	130	131	0	1	115	1	0	yes	NaN
<b>302</b>	57	0	130	236	0	0	174	1	0	no	typical angina

303 rows × 11 columns

### 3. Limpieza de los datos.

### 4. Análisis de los datos.

### 5. Representación de los resultados

Se realizará durante toda la práctica

### 6. Resolución del problema

In [ ]: