


MALALTIES CARDIAQUES EN EL PACIENT

PROJECTE FINAL

—
BIG DATA

SILVIA AMPUERO | IT ACADEMY – JUNY 2024



PRESENTACIÓ DEL CONJUNT DE DADES

El conjunt de dades escollit és el "Conjunto de datos de enfermedades cardíacas" de Kaggle, en aquest enllaç <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>.

Aquest conjunt de dades, recopilat el 1988, inclou informació de pacients de quatre bases de dades diferents: Cleveland, Hongria, Suïssa i Long Beach V.

Tot i que el conjunt de dades conté 76 atributs, els experiments publicats solen utilitzar un subconjunt de 14 atributs. L'atribut principal és la presència o absència de malaltia cardíaca en el pacient, amb un valor enter de 0 (sense malaltia) o 1 (amb malaltia).



CARACTERÍSTIQUES GENERALES

El conjunt de dades utilitzat en aquest projecte correspon a dades de salut pública relacionades amb malalties cardiaques que pertanyen al sector de la salut i la medicina i inclouen tant dades quantitatives com qualitatives.

El conjunt de dades es recopila a partir de pacients de quatre bases de dades diferents: Cleveland, Hongria, Suïssa i Long Beach V.

Aquestes dades són principalment per a estudis i investigacions mèdiques amb l'objectiu de predir la presència de malaltia cardíaca en pacients.

DEFINICIÓ DE LES VARIABLES

Tenim definides 14 variables, que són les següents:

1. **Edat:** Edat del pacient en anys.
2. **Sexe:** Sexe del pacient (1 = masculí, 0 = femení).
3. **Tipus de Dolor Toràcic:** Classificació en 4 tipus de dolor toràcic.
4. **Pressió Arterial en Repos:** Mesura de la pressió arterial en repos en mm Hg.
5. **Colesterol Sèric:** Nivell de colesterol en sèrum en mg/dl.
6. **Glucèmia en dejuni:** Nivell de glucosa en sang en dejuni superior a 120 mg/dl.
7. **Resultats Electrocardiogràfics en Repos:** Valors de 0, 1 o 2 segons els resultats de l'ECG en repòs.
8. **Freqüència Cardíaca Màxima Assolida:** Freqüència cardíaca màxima assolida durant una prova d'esforç.
9. **Angina Induïda per l'Exercici:** Presència d'angina induïda per l'exercici (1 = sí, 0 = no).
10. **Oldpeak:** Depressió del segment ST induïda per l'exercici en relació amb el repòs.
11. **Pendents del Segment ST:** Inclinatoria del segment ST en el màxim exercici.

12. Nombre de Vasos Principals Colorejats per

Fluoroscòpia: Nombre de vasos principals (0-3) colorejats.

13. Thal: 0 = normal; 1 = defecte reparat; 2 = defecte reversible.

14. Objectiu: Presència de malaltia cardíaca (0 = no, 1 = sí).



PRESENTACIÓ DELS OBJECTIUS

L'objectiu principal d'aquest projecte és desenvolupar un model predictiu que permeti determinar la presència de malaltia cardíaca en pacients utilitzant el conjunt de dades de malalties cardíques.

Els objectius secundaris són; realitzar un anàlisi exploratori de dades, selecció de característiques, construcció del model predictiu, validació del model i finalment la interpretació dels resultats.