

Vraag 1: Lees de dataset heart_ex.csv in, en noem de lijst 'heart'.

Bespreking van de belangrijkste features

age

sex (0 = vrouw, 1 = man)

cp : chest pain type

trtbps: resting blood pressure (in mm Hg)

chol: cholestoral in mg/dl fetched via BMI sensor

fbs : (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)

restecg : resting electrocardiographic results

thalachh : maximum heart rate achieved

exng : exercise induced angina (1 = yes; 0 = no)

oldpeak : Previous peak

slp : Slope

caa : number of major vessels (grote bloedvaten van het hart) (0-4)

thall : Thal rate (hartslag)

output : geeft een indicatie naar een hoog of laag risico op een hartaanval (1 = hoog risico, 0 = laag risico)

Het is niet belangrijk om de juiste betekenis van de verschillende features te kennen.

Vraag 2: Toon de 10 eerste rijen van de dataset "heart".

Vraag 3: Wat voor soort (categorisch ordinaal - categorisch nominaal - numeriek discreet - numeriek continu) variabele is

exng:

chol:

Vraag 4: Hoeveel types chest pain (cp) zijn er? Geef per type weer hoeveel keer het voorkomt (geen grafiek, enkel de cijfers).

Vraag 5: Maak een grafiek van de aantallen patiënten naar cholesterolwaarden? Bepaal zelf welk type grafiek hier het best voor in aanmerking komt.

Vraag 6: Maak een cirkeldiagram van de slope (slp).

Vraag 7: Ga na of er een verband is tussen het geslacht en het risico op een hartaanval.

Vraag 8: Ga na of er een verband is tussen de cholesterol-waarde (chol) en de hoogst bereikte waarde voor hartslag (thalachh).

Vraag 9: Welk aantal grote bloedvaten van het hart (caa) komt het meest voor bij vrouwen? Hoeveel procent van de vrouwen heeft dit?

We willen voorspellen of een patiënt een verhoogd risico op een hartaanval heeft of niet.

Vraag 10: Waarom kunnen we de feature 'chest pain' (cp) in de huidige staat niet laten meespelen bij het voorspellen van een hartaanval?

Vraag 11: Maak een nieuwe dataset heart_ML met de variabele output als target en waarbij je volgende variabelen als feature selecteert:

age

sex (0 = vrouw, 1 = man)

trtbps: resting blood pressure (in mm Hg)

chol: cholesterol in mg/dl fetched via BMI sensor

restecg : resting electrocardiographic results

exng : exercise induced angina (1 = yes; 0 = no)

caa : number of major vessels (grote bloedvaten van het hart) (0-3)

thall : Thal rate (hartslag)

Vraag 12: Toon de laatste 5 rijen van dataset heart_ML

Vraag 13: Zet input (features) in X en output in y

Vraag 14: Neem 75% als training data en 25% als testing data

Vraag 15: Pas één van de geziene ML algoritmes toe om y_pred te bepalen. Zorg ervoor dat de voorspelling minstens 75% nauwkeurig is

Vraag 16: Print de eerste 20 rijen van de voorspelde y-waarde (y_pred) en de werkelijke y-waarde

Vraag 17: Hoe nauwkeurig is deze voorspelling?