

## Rapport Python AI

### 1. Beskrivning av datasetet (källa och innehåll).

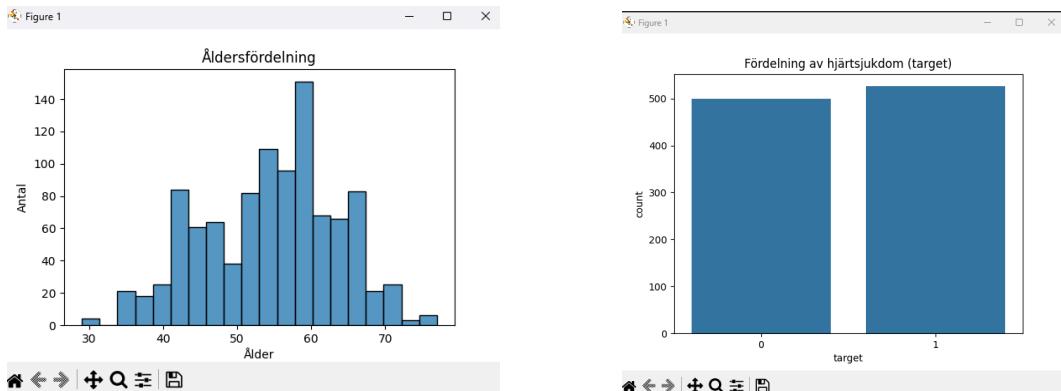
Datasetet i projektet är Heart Disease Dataset vilket är ett dataset som innehåller data om patienter med olika medicinska attribut såsom ålder, kön, blodtryck, kolesterolnivåer och EKG. I datasetet så representeras varje individ rad per rad och variabeln **target** anger ifall individen har hjärtsjukdom (1) eller inte (0).

### 2. Förläring av vald modell och varför den valdes

Modellen som användes i projektet är Logistic Regression. Jag valde denna modell då användningsområdet för denna modell beskrivs vara inom statistik och binära problem, det vill säga hjärtsjukdom (1) - frisk (0). Detta gav mig intrycket att denna modell var väldigt bra lämpad till uppgiften då huvuduppgiften av mitt projekt var att få två möjliga utfall.

### 3. Kortfattad presentation av resultat (t.ex. tabell, diagram eller nyckelvärden).

I början av projektet skapades grafer för att visualisera datan i datasetet både för åldersfördelning och sjukdoms fördelning för att få en överblick på datan.



Efter träning och köring av modellen uppnådde den en accuracy på 80% vilket betyder att i majoriteten av fallen kan modellen korrekt indikera om hjärtsjukdom eller inte. I applikationen testades även modellen med olika inmatningar där jag observera hur sannolikheten för hjärtsjukdom förändrades beroende på vilka värden som matades in.

Resultat av att köra applikationen och inmatad data:

```

Accuracy: 0.7951219512195122
precision    recall   f1-score   support
          0       0.85      0.72      0.78     102
          1       0.76      0.87      0.81     103

   accuracy           0.80      205
   macro avg       0.80      0.79      0.79     205
weighted avg       0.80      0.80      0.79     205

Mata in patientdata:
Ålder: 63
Kön (0 = kvinna, 1 = man): 0
Bröstmärta (0-3): 1
Vilo-blodtryck: 120
Kolesterol: 130
Fastande blodsocker > 120 (0 = nej, 1 = ja): 0
EKG-resultat (0-2): 0
Maxpuls: 179
Ansträngningsutlöst angina (0 = nej, 1 = ja): 0
ST-depression (oldpeak): 0
Lutning ST (0-2): 0
Antal färgade kärl (0-3): 1
Thal (1-3): 1
Sannolikhet för hjärtsjukdom: 0.99
● Hög risk för hjärtsjukdom

```

```

Accuracy: 0.7951219512195122
Mata in patientdata:
Accuracy: 0.7951219512195122
Mata in patientdata:
Ålder: 50
Ålder: 50
Kön (0 = kvinna, 1 = man): 1
Bröstmärta (0-3): 2
Kön (0 = kvinna, 1 = man): 1
Bröstmärta (0-3): 2
Vilo-blodtryck: 160
Vilo-blodtryck: 160
Kolesterol: 200
Kolesterol: 200
Fastande blodsocker > 120 (0 = nej, 1 = ja): 0
Fastande blodsocker > 120 (0 = nej, 1 = ja): 0
EKG-resultat (0-2): 0
Maxpuls: 90
Maxpuls: 90
Ansträngningsutlöst angina (0 = nej, 1 = ja): 0
ST-depression (oldpeak): 3
Lutning ST (0-2): 2
Thal (1-3): 1
Sannolikhet för hjärtsjukdom: 0.43
Resultat:
Måttlig risk för hjärtsjukdom

```

#### 4. Den etiska reflektionen

När man först hör om en AI modell som kan vara till hjälp med att diagnostisera patienter gällande hjärtsjukdom så låter det väldigt bra men det är viktigt att betona hur modellen fungerar.

Modellen baserar sina gissningar på ett statistiskt mönster i datasetet och inte på några medicinska riktlinjer eller expertis.

Datan är också begränsad vilket innebär att det i största sannolikhet kan vara missvisande för vissa individer som inte representeras i datan.

Dilemmat med denna typ av metod inom sjukvård är att något sånt här skulle kunna vara väldigt hjälpsamt i teorin, men ifall det visar sig att en sån här modell har begått ett misstag eller att det finns en bug i den så är det svårt att veta hur man ska hantera det då skadan redan är gjord. Att råka feldiagnostisera patienter är något som troligtvis kommer att ske med en sådan modell.

Att använda AI till att diagnostisera ifall patienter har tecken på hjärtsjukdom är något som inte bör göras. Det kan enbart göras den dagen program och modeller inte kan ha fel och det är högst osannolikt att det någonsin kommer att ske. Frågan om vem som bär ansvaret när AI har fel är fortfarande olöst.