



# 心臟數據分析

GET STARTED

# 組員介紹



**潘玠佑**

成功大學工程科學系

**王安琪**

逢甲大學財金系

**黃晰婕**

東海大學財務金融系

**黃晟旺**

逢甲大學資訊工程系



# 目錄



緣由



預計方法



成果建議



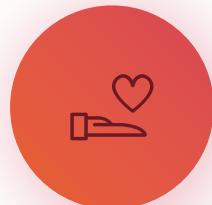
# 緣由

---

Heart disease is a very crucial and sudden attack as it may result in the loss of life. Therefore, it is important to verify the symptoms, to know what affects the beginning of this disease to prevent sickness.



# 預期方法



## 分群法

遮住class項（是否有心臟病），  
以其他屬性來判斷分群關係



## 關聯規則

使用關聯與分類來看  
會發現什麼特別之處

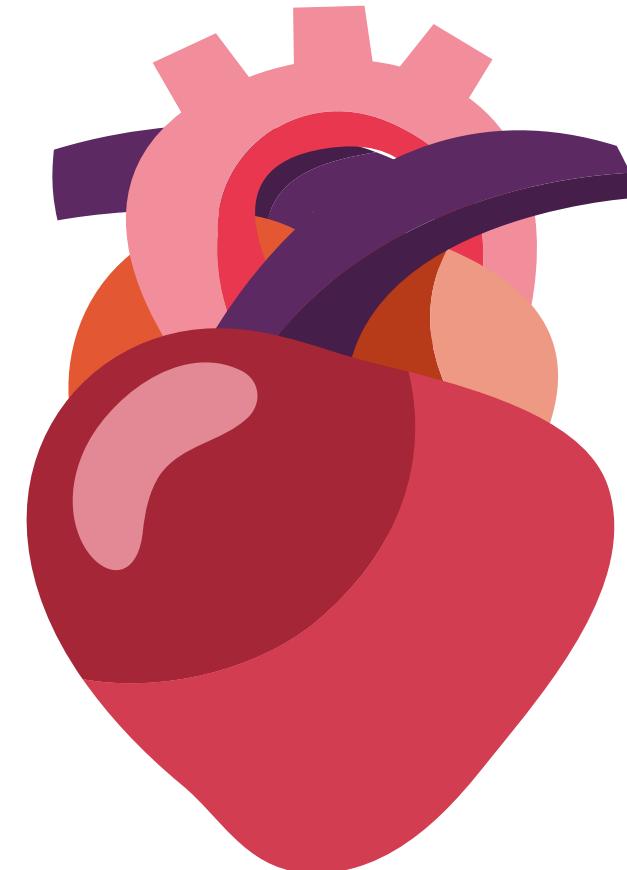


## 分類法



年齡
性別
胸痛類型 ( 4 個值 )
靜息血壓
以 mg/dl 為單位的血清膽甾醇
空腹血糖 $> 120 \text{ mg/dl}$
靜息心電圖結果 ( 值 0、1、2 )
達到的最大心率
運動性心絞痛
oldpeak = 運動相對於休息引起的 ST 段壓低
峰值運動ST段的斜率
熒光染色的主要血管數量 ( 0-3 )
塔爾 : 0 = 正常 ; 1 = 固定缺陷 ; 2 = 可逆缺陷 患者的姓名和社會安全號碼最近從數據庫中刪除，取而代之的是虛擬值

# Data Set





# 資料預處理

離散化

降維

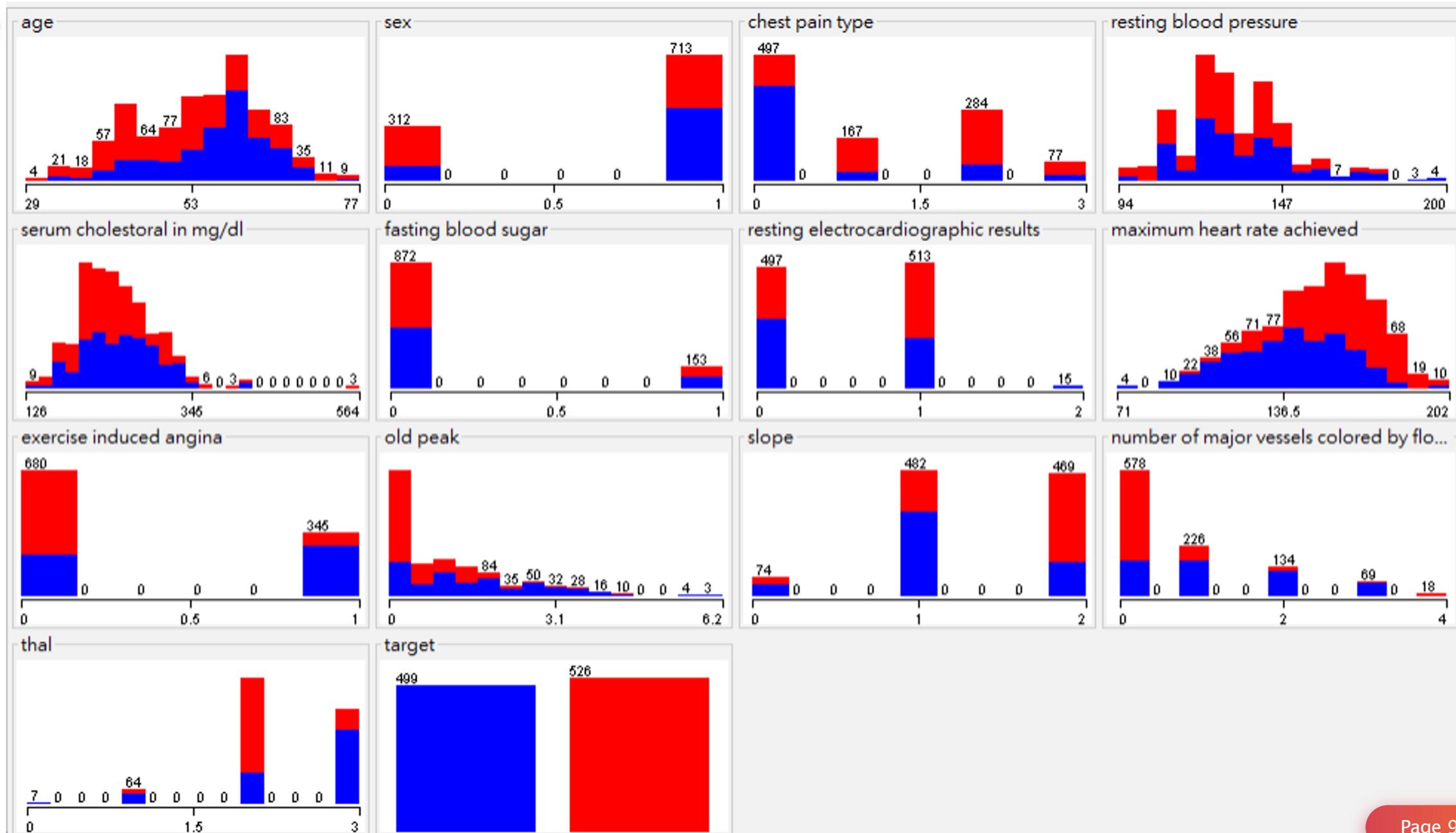


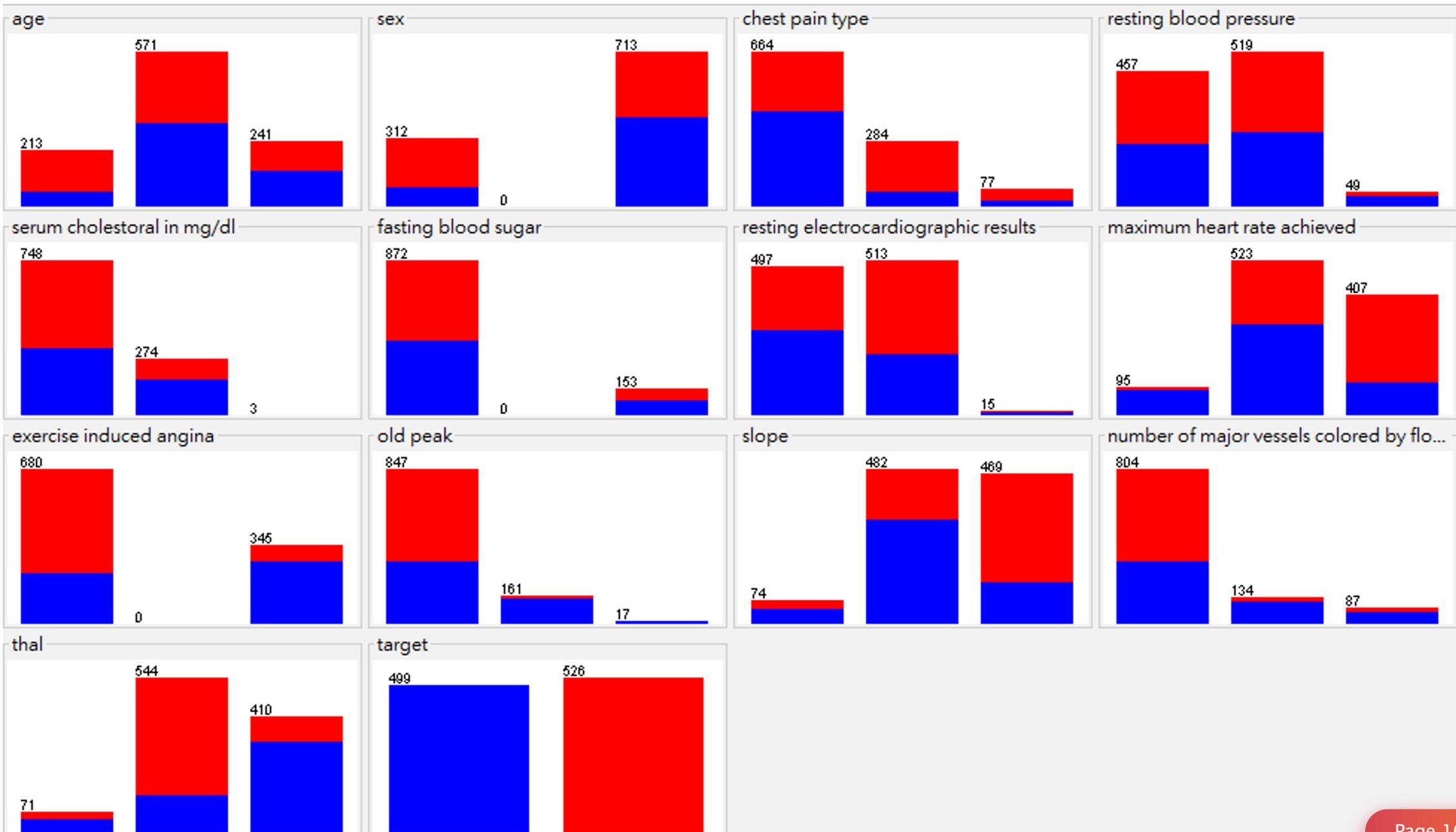
# 離散化

舉例

年齡: 0~100









# 降維

remove

Ranked attributes:

```
0.2047 3 chest pain type
0.1621 13 thal
0.1574 9 exercise induced angina
0.1272 12 number of major vessels colored by flourosopy
0.113 11 slope
0.1014 10 old peak
0.0882 8 maximum heart rate achieved
0.0768 1 age
0.0654 2 sex
0.0516 5 serum cholestoral in mg/dl
0.0388 4 resting blood pressure
0.0233 7 resting electrocardiographic results
0 6 fasting blood sugar
```

Selected attributes: 3,13,9,12,11,10,8,1,2,5,4,7,6 : 13



# 分群法 cluster

未知群集（類別），找到群集（類別）。  
「分群」。所有數據進行分組，相似數據歸類於同一組，一筆數據只屬於某一組，每一組稱作一個「群集 Cluster」。

simple  
K-means



# 分群法 cluster

2 group eular distance  
3 groups eular distance

Class attribute: target

Classes to Clusters:

0 1 <-- assigned to cluster

274 225 | F

71 455 | T

2 group

Cluster 0 <-- F

Cluster 1 <-- T

Incorrectly clustered instances : 296.0 28.878 %

Class attribute: target

Classes to Clusters:

0 1 2 <-- assigned to cluster

274 82 143 | F

71 304 151 | T

3 group

Cluster 0 <-- F

Cluster 1 <-- T

Cluster 2 <-- No class

Incorrectly clustered instances : 447.0 43.6098 %



# 分群法 cluster

2 groups manhattan distance

Class attribute: target

Classes to Clusters:

0	1	<-- assigned to cluster
349	150	F
63	463	T

Cluster 0 <-- F

Cluster 1 <-- T

Incorrectly clustered instances :

213.0 20.7805 %

# 分群結論

只分兩群是對的



# 關聯規則

Min Support  
Min Confidence



# 關聯規則



# 關聯規則結論

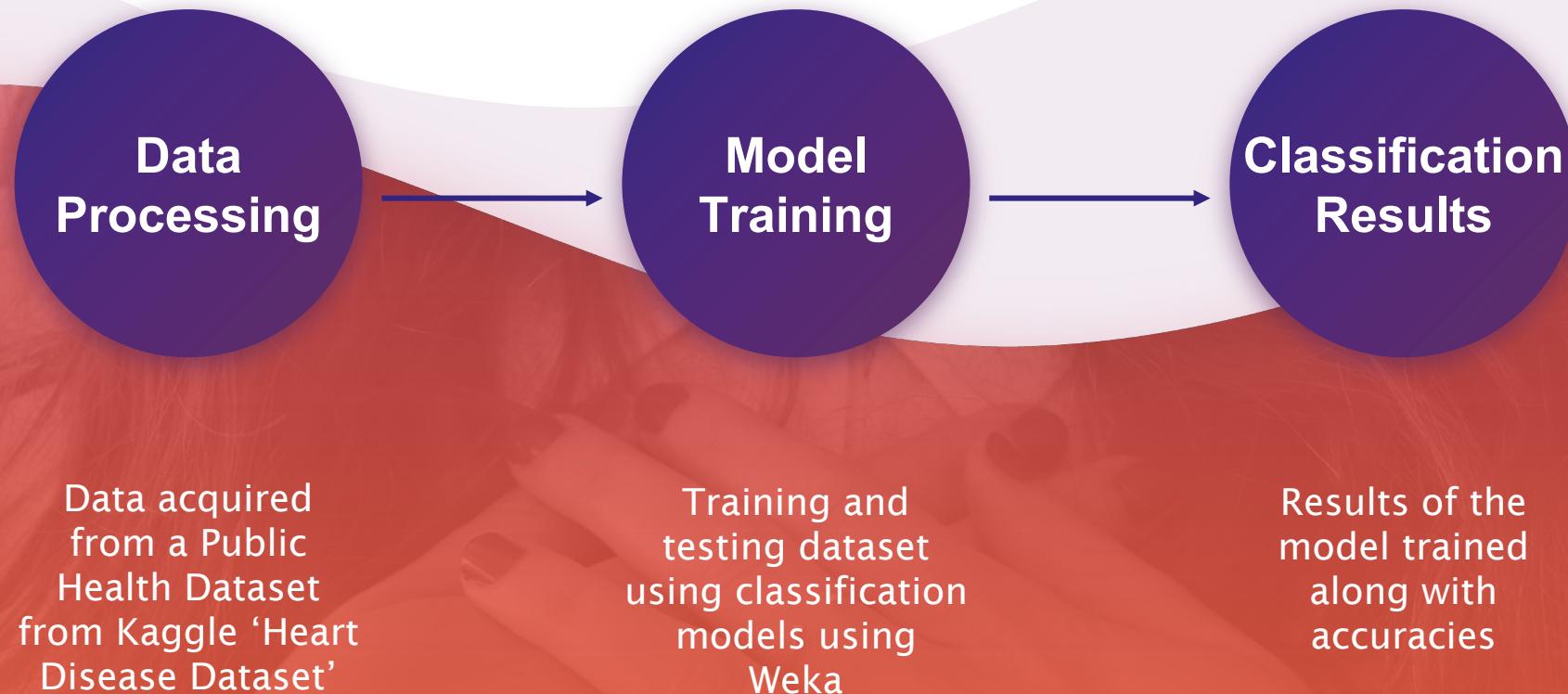
勞發性心絞痛 & X光下可視主血管數=0 ==> T  
<sup:(0.5) conf:(0.81)>

勞發性心絞痛 & Peak<0.3 ==> T  
<sup:(0.5) conf:(0.72)>

勞發性心絞痛 ==> T  
<sup:(0.4) conf:(0.68)>



# 分類法





# 分類法 結果

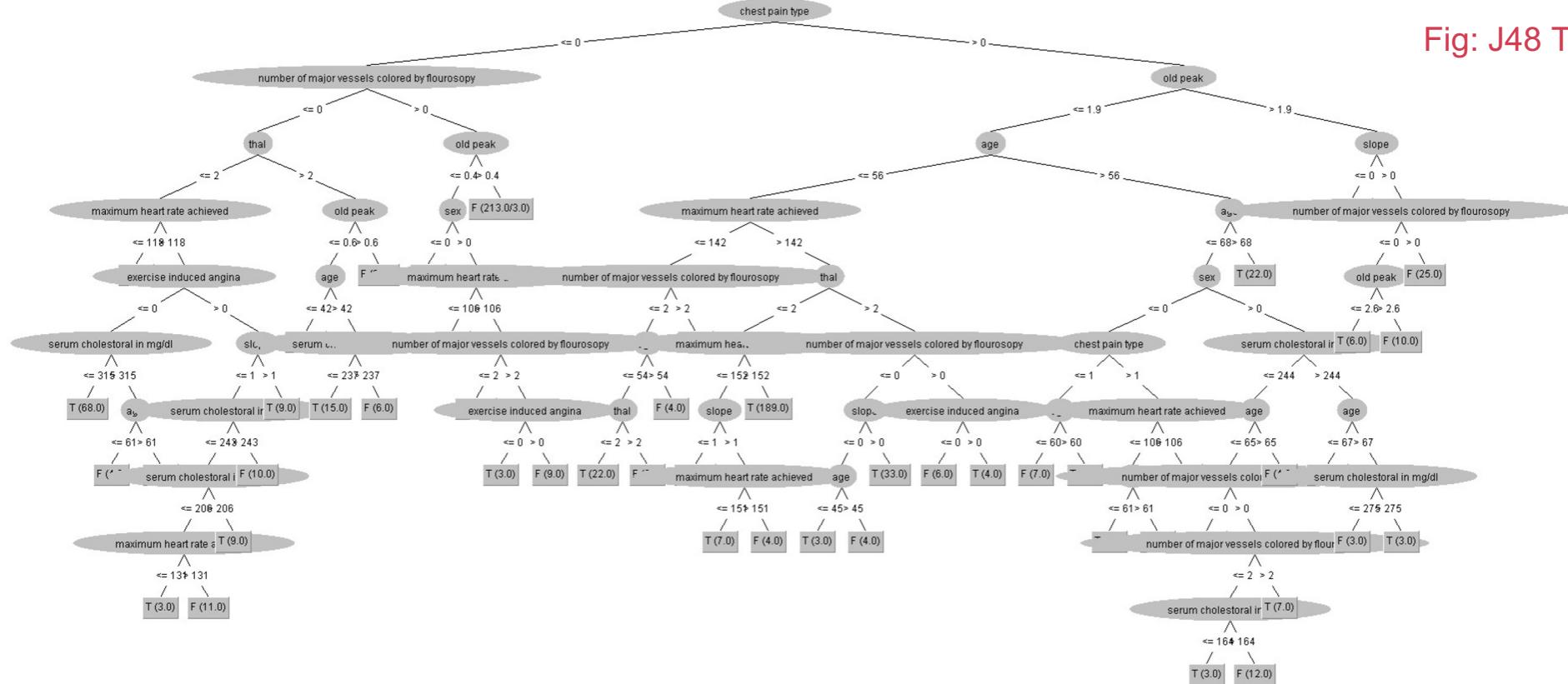


Fig: J48 Tree

<i>Model</i>	<i>Accuracy (%)</i>
Tree J48 (C4.5)	97.65
<b>Random Forest Tree</b>	<b>100</b>

# 分類結論

由於樹分支過多，不易看出直接成因。

但仍可發現最頂部影響分支為：

1. 透視下可視血管數量
2. 心臟跳動最高點
3. 是否有心絞痛



# 預期成果

成因



防治

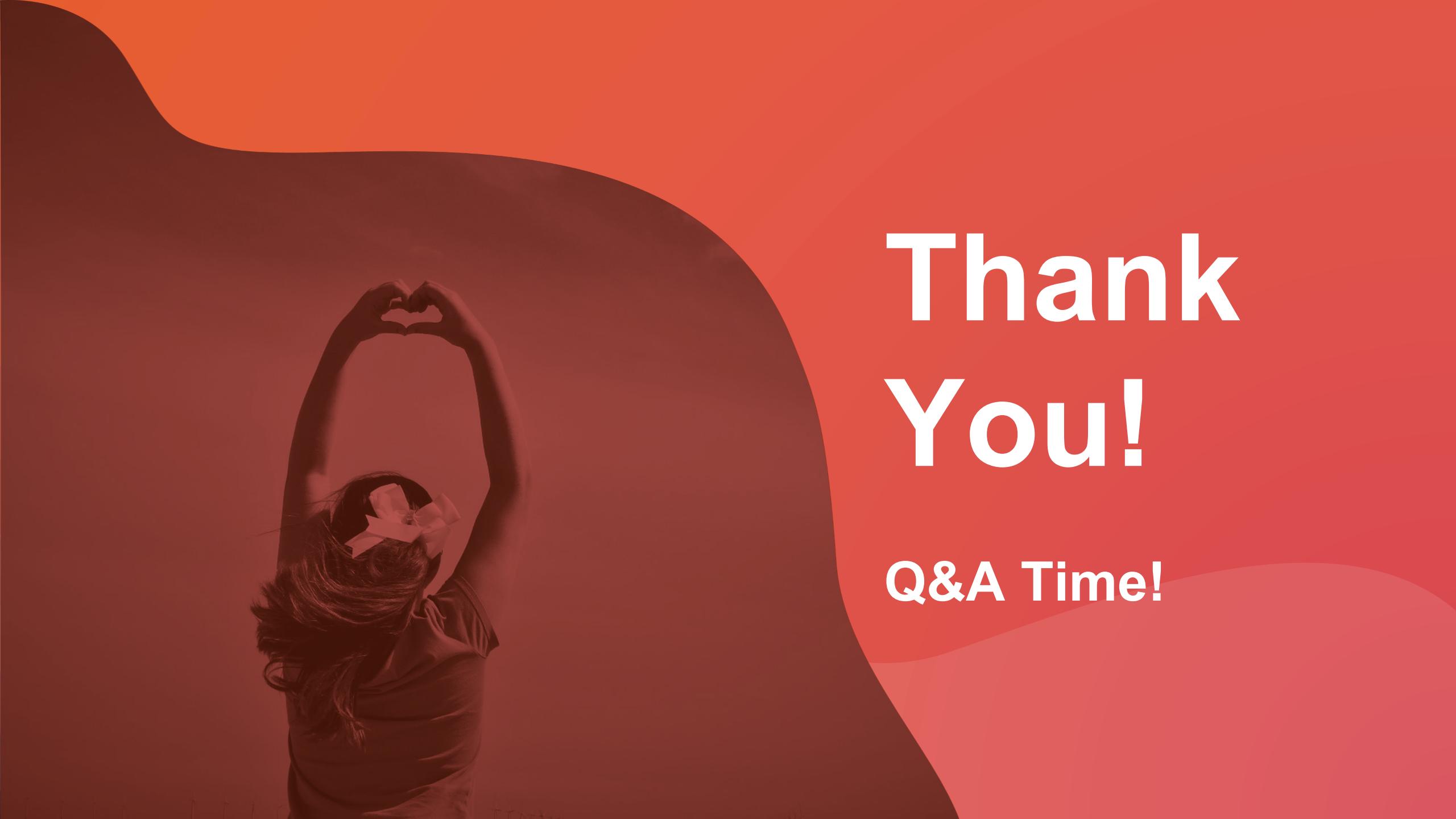


過勞 肥胖



# 建議

- Compare different models to find model with the best accuracy.
- Enlarge dataset and parameter optimizations to create more accurate models



Thank  
You!

Q&A Time!