

指導老師:許嘉裕

110378042 楊承諺

110378049 張禕君

110378053 劉云韜

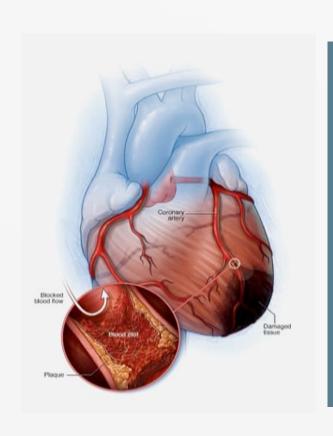


- 01 研究背景
- 02 研究方法&結果
- 03 結論



研究背景

研究背景

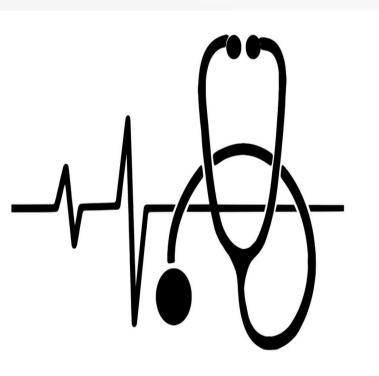


心臟病被認為是當今世界的主要死亡原因之一。目標是通過數據集上應用數據挖掘和機器學習技術來提取隱藏的模式,幫助醫療保健專業人員預測心臟病。



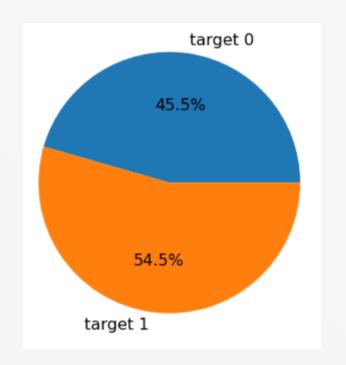


數據資料介紹



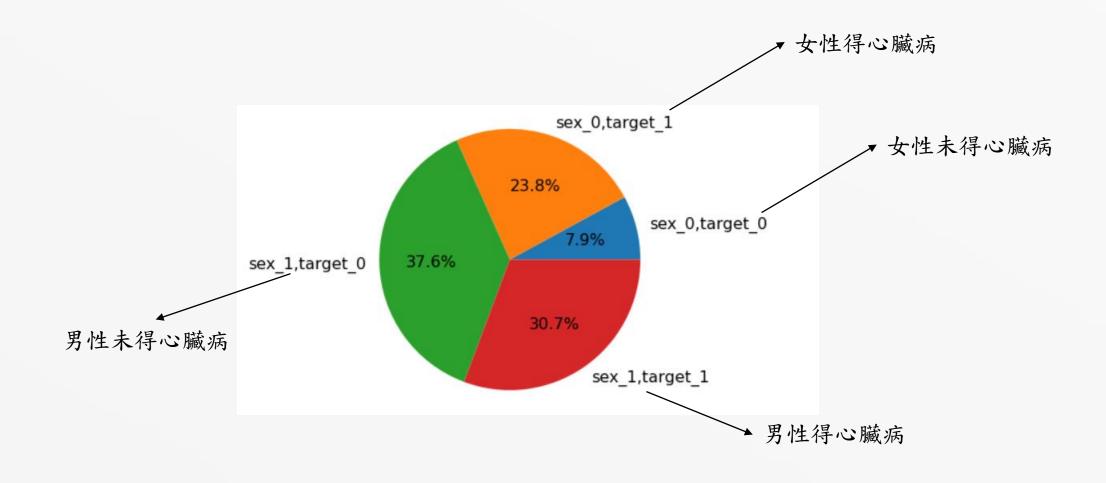
304 筆 資 料 , 14 個 特 徵 , 例 如 : 年 龄 (age)、鈣(ca)、膽 固 醇(chol)、性 別、最大心率等

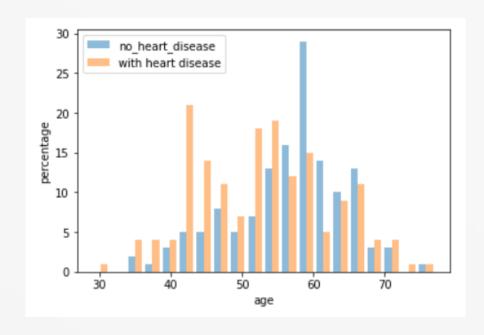
	age	sex	ср	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	са	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1



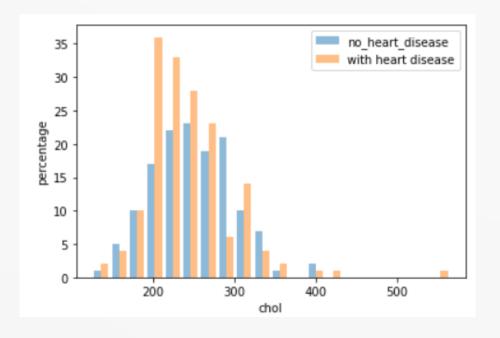
1:診斷出得心臟病(165人)

0:診斷未得心臟病(138人)

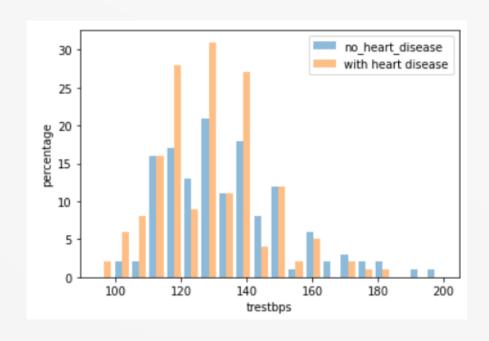




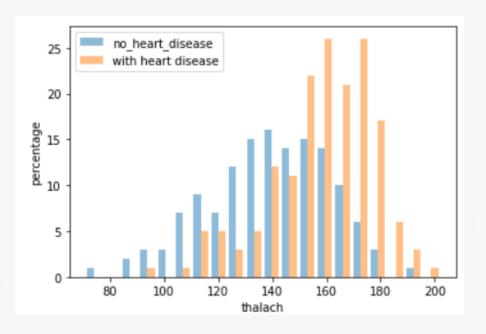
40歲以上的人患有心臟病的風險更高



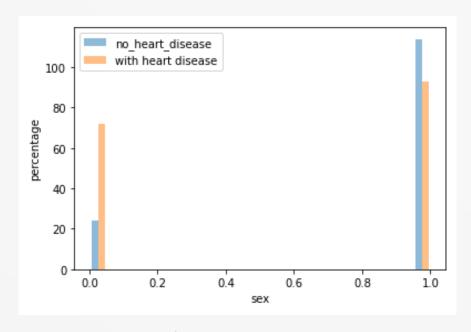
患有心臟病的人膽固醇含量超過 200 毫克/分升。 根據研究,膽固醇 (chol) 的正常值應低於200mg/dl。



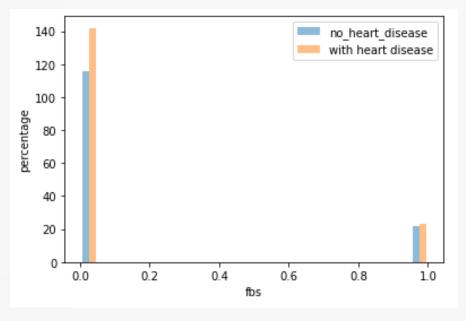
理想的血壓應低於120 mmHg。 無論患者是否患有心臟病,50%以上的患者血壓較高。



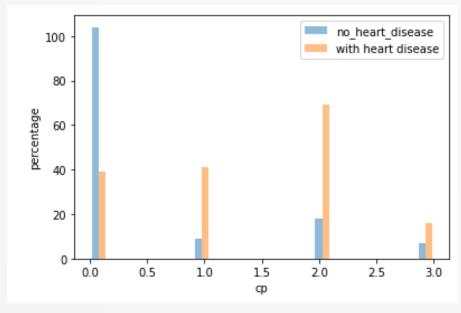
最大心率與年齡呈負相關,結果心臟 病患者的中位數更高。



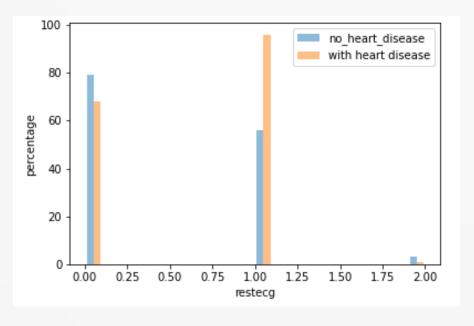
男性患心臟病高於女性



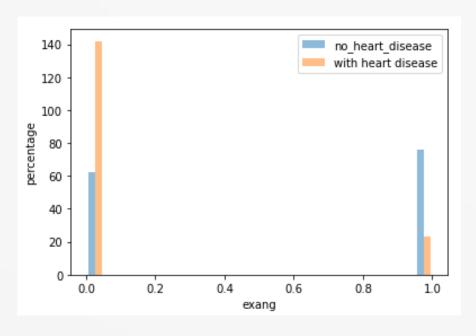
空腹血糖 (> 120 mg/dl , 1 = true; 0 = false)



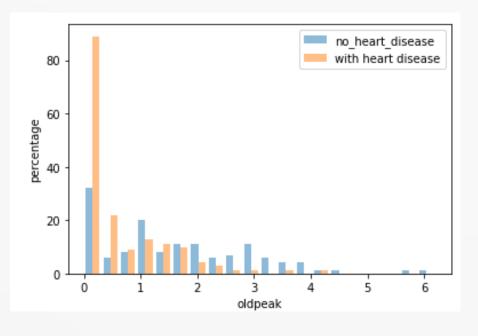
心絞痛,患有非典型絞痛的人患心臟病較高



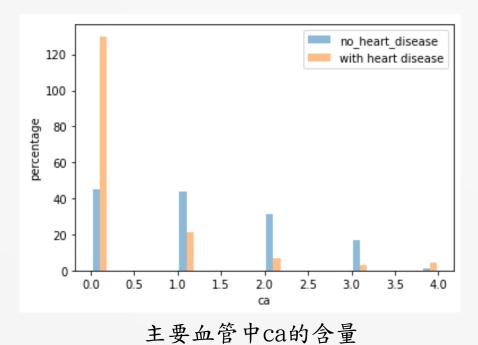
静息心電圖測量(0 = 正常,1 = ST-T 波異常,2 = 根據 Estes 標準顯示可能或明確的左心室肥大), ST-T波異常較易診斷出心臟病



運動誘發的心絞痛(1 = 是;0 = 否)



相對於休息,運動引起的 ST 段壓低



no_heart_disease 120 with heart disease 100 percentage 80 60 40 20 1.0 2.0 0.0 0.5 1.5 2.5 3.0

一種稱為地中海貧血的血液疾病 (1 = 正常; 2 = 固定缺陷; 3 = 可 逆缺陷)

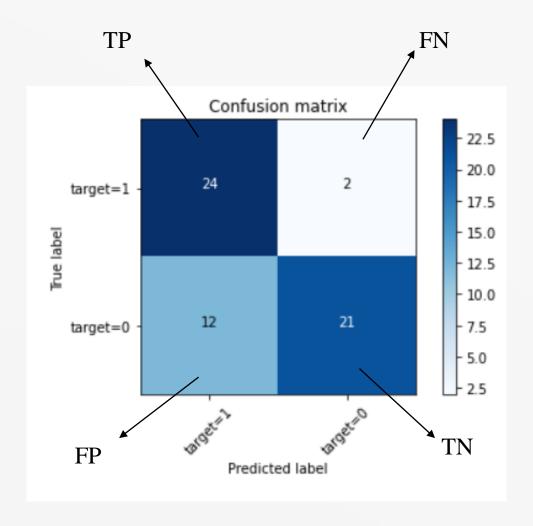


	age	trestbps	chol	fbs	restecg	thalach	oldpeak	ca	target	sex_Male cp	_atypical angina	cp_non- anginal pain	cp_typical angina	exang_exercise induced angina	slope_upsloping	thal_normal	thal_reversable defect
0	63	145	233	1	0	150	2.3	0	1	1	0	1	0	0	0	1	0
1	37	130	250	0	1	187	3.5	0	1	1	17	0	0	0	0	0	0
2	41	130	204	0	0	172	1.4	0	1	0	0	0	1	0	0	0	0
3	56	120	236	0	1	178	0.8	0	1	1	0	0	1	0	0	0	0
4	57	120	354	0	1	163	0.6	0	1	0	0	0	0	1	0	0	0
5	57	140	192	0	1	148	0.4	0	1	1	0	0	0	0	1	1	0
6	56	140	294	0	0	153	1.3	0	1	0 /	0	0	1	0	1	0	0
7	44	120	263	0	1	173	0.0	0	1	1	0	0	1	0	0	0	1
8	52	172	199	1	1	162	0.5	0	1	1/	1	0	0	0	0	0	1
9	57	150	168	0	1	174	1.6	0	1	/1	1	0	0	0	0	0	0
										/							

為了更清楚了解每個特徵意涵我們將資料進行擴充,例如心絞痛分成 1. 非典型心絞痛2. 非心絞痛3. 典型心絞痛



研究方法&結果



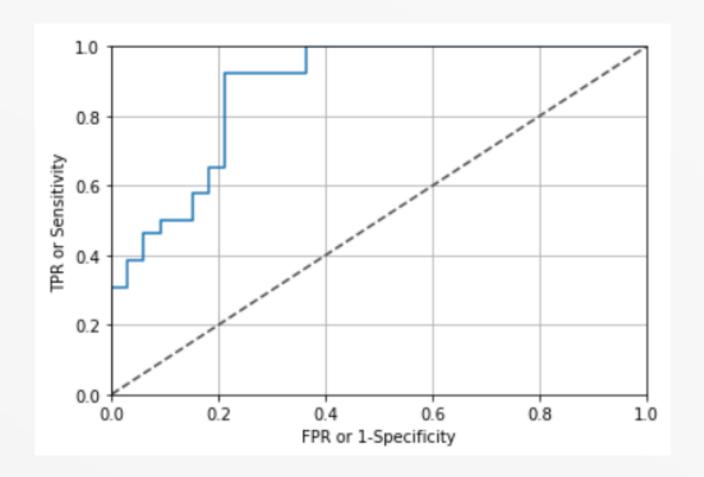
將資料切割成訓練集測試集8:2 使用XGB Classifier

n_estimators:100; max_depth:3; learning_rate:0.1

假陰性數量為2,代表召回率再90%以上

The precision rate=67%; The recall rate=92%; The f1_score rate=77%

ROC curve



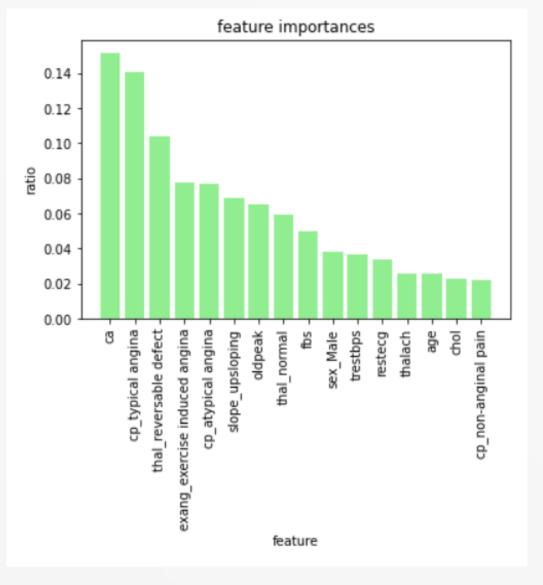
AUC=0.88→AUC在0.5~1的範圍內,說明模型具有預測價值。

Permutation importance

Weight	Feature
0.0475 ± 0.0254	sex_Male
0.0441 ± 0.0460	thal_reversable defect
0.0237 ± 0.0973	ca
0.0034 ± 0.0136	exang_exercise induced angina
0 ± 0.0000	thal_normal
0 ± 0.0000	cp_typical angina
0 ± 0.0000	cp_non-anginal pain
0 ± 0.0000	fbs
-0.0000 ± 0.0303	trestbps
-0.0102 ± 0.0166	cp_atypical angina
-0.0136 ± 0.0332	restecg
-0.0169 ± 0.0214	slope_upsloping
-0.0203 ± 0.0691	oldpeak
-0.0339 ± 0.0303	chol
-0.0373 ± 0.0332	thalach
-0.0441 ± 0.0460	age

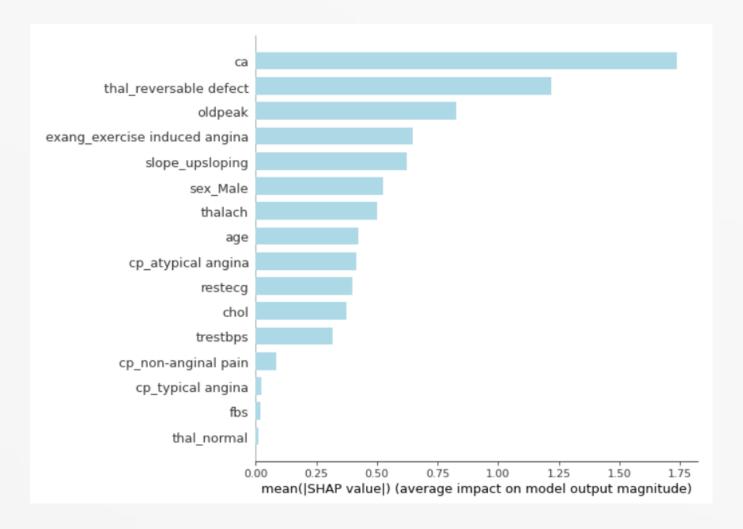
透過可解釋性來了解每個特徵的權重,也可得知 sex_Male是對結果影響最大,age並未有太大影響。

Feature importance



由此圖可以得知,可逆缺陷和ca特徵較其他高

SHAP值



當f $(x_{i,j})>0$,說明該特徵對目標值的預測起到了正向作用;反之,該 特徵使與目標預測值有相反作用。因此Shap不僅給出特徵影響力的大小, 也反映出每一個樣本中的特徵的影響力的正負性。

21



◆結論

透過Feature importance和SHAP值得比較,發現鈣(ca)是所有特徵中影響心臟病的最主要原因,和我們當初預期是膽固醇(chol)、年齡、最大心律和血壓是主要原因有所差異。

根據國民健康署的「國民營養健康狀況變遷調查」鈣離子可以控制肌肉的緊繃、收縮。在缺乏鈣離子的情況下,肌肉的收縮、放鬆就會變得不聽使喚,同樣也包含心臟,所以心臟會開始過度收縮、引起心悸。(https://www.hpa.gov.tw/home/index.aspx)

◆未來想法

透過前面的結果,發現有四個特徵 (cp_typical agina、fbs、cp_non-aginal pain和thal_normal) 的SHAP值未超過0.25,表示特徵影響力不大,應當刪除後再做一次建模。

謝謝聆聽~