

利用身體指數 預測是否罹患肝病

第一組期末報告

4108040027 許喬淇 4109012008 紀懿佳

指導教授：蔡孟勳 教授

Content

1 資料集介紹

2 研究動機與目的

3 使用的SDGs

4 魚骨圖

5 資料預處理與分析

6 模型評估與結果

7 知識發現

8 文獻參考

9 結論

資料集介紹

- 資料集：Liver Disease Patient Dataset
- 來源：kaggle
- 資料集說明：
 - 具有一個 training dataset 和 testing dataset
 - training dataset 為已標註資料集
 - 資料集具有21917個正樣本(患病)、8774個負樣本(無患病)
 - 身體紀錄包含年齡、性別、膽紅素、鹼性磷酸酶、轉氨酶等

資料集介紹

- 共有十項身體檢查紀錄和是否患病的欄位(Result)
- 是否患病是由專家標註，1為患病、2為沒有患病

	Age of the patient	Gender of the patient	Total Bilirubin	Direct Bilirubin	lkphos Alkaline Phosphatase	gpt Aminotransferase	Alamine Aminotransferase	Sgot Aspartate Aminotransferase	Total Protiens	LB Albumin	A/G Ratio Albumin and Globulin Ratio	Result
0	65.0	Female	0.7	0.1	187.0		16.0	18.0	6.8	3.3	0.90	1
1	62.0	Male	10.9	5.5	699.0		64.0	100.0	7.5	3.2	0.74	1
2	62.0	Male	7.3	4.1	490.0		60.0	68.0	7.0	3.3	0.89	1
3	58.0	Male	1.0	0.4	182.0		14.0	20.0	6.8	3.4	1.00	1
4	72.0	Male	3.9	2.0	195.0		27.0	59.0	7.3	2.4	0.40	1
5	46.0	Male	1.8	0.7	208.0		19.0	14.0	7.6	4.4	1.30	1
6	26.0	Female	0.9	0.2	154.0		NaN	12.0	7.0	3.5	1.00	1
7	29.0	Female	0.9	0.3	202.0		14.0	11.0	6.7	3.6	1.10	1
8	17.0	Male	0.9	0.3	202.0		22.0	19.0	7.4	4.1	1.20	2
9	55.0	Male	0.7	0.2	290.0		53.0	58.0	6.8	3.4	1.00	1

資料集的前十列

研究動機與目的

背景

- 在台灣，肝病有國病之稱。衛福部死因統計，每年約13,000人死於慢性肝病、肝硬化及肝癌。
- 肝臟是個沉默的器官，沒有明顯患病初期症狀。



目的

- 利用機器學習分析身體各項指數，預測罹患肝病的可能性。

使用的SDGs

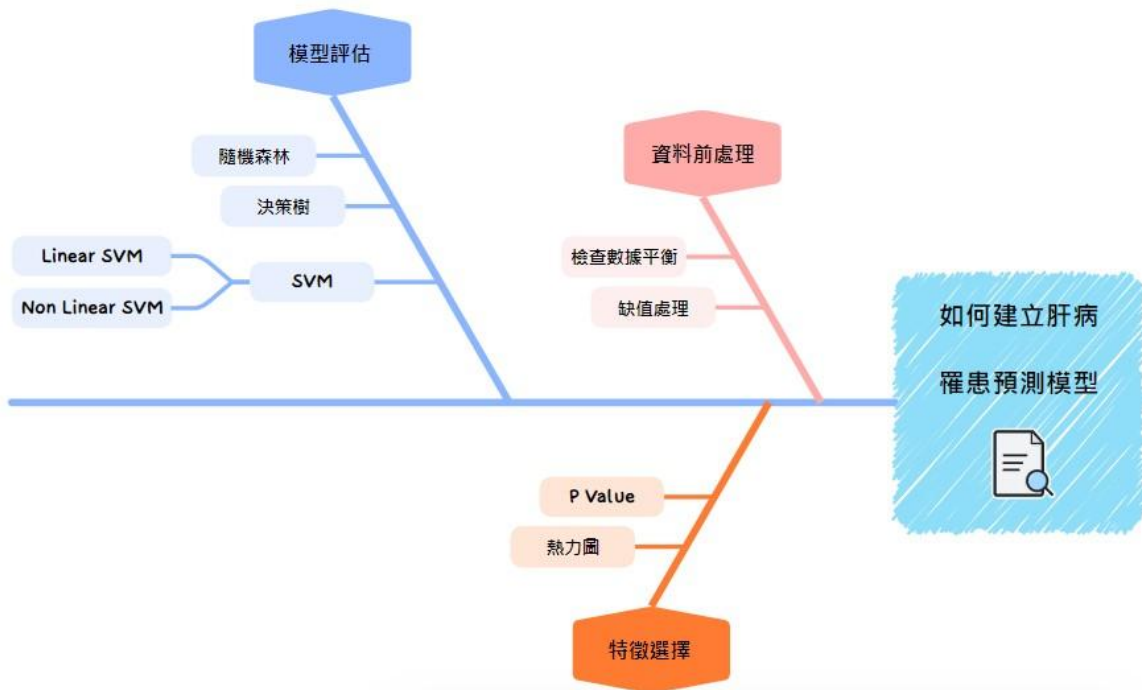
3 健康與福祉



減少非傳染性疾病造成的死亡率並促進心理健康

在西元2030年前，透過預防與治療，將非傳染性疾病的未成年死亡數減少三分之一，並促進心理健康。

魚骨圖



資料預處理與分析

資料預處理

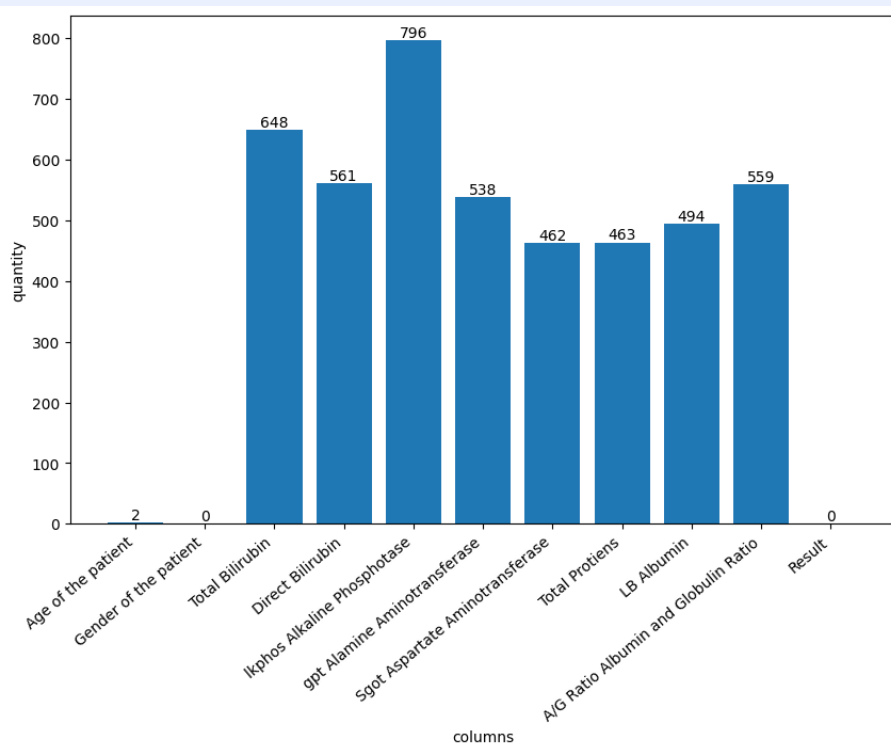
- 數據化Gender of the patient : 0 (female) 、1 (male)

	Age of the patient	Gender of the patient	Total Bilirubin	Direct Bilirubin	lkphos Alkaline Phosphatase	gpt Alamine Aminotransferase	Sgot Aspartate Aminotransferase	Total Protiens	LB Albumin	A/G Ratio Albumin and Globulin Ratio	Result
0	65.0	Female	0.7	0.1	187.0	16.0	18.0	6.8	3.3	0.90	1
1	62.0	Male	0	5.5	699.0	64.0	100.0	7.5	3.2	0.74	1
2	62.0	Male	1	4.1	490.0	60.0	68.0	7.0	3.3	0.89	1
3	58.0	Male	1	0.4	182.0	14.0	20.0	6.8	3.4	1.00	1
4	72.0	Male	1	2.0	195.0	27.0	59.0	7.3	2.4	0.40	1
5	46.0	Male	1	0.7	208.0	19.0	14.0	7.6	4.4	1.30	1
6	26.0	Female	1	0.2	154.0	NaN	12.0	7.0	3.5	1.00	1
7	29.0	Female	0	0.3	202.0	14.0	11.0	6.7	3.6	1.10	1
8	17.0	Male	0	0.3	202.0	22.0	19.0	7.4	4.1	1.20	2
9	55.0	Male	1	0.2	290.0	53.0	58.0	6.8	3.4	1.00	1
			1								

資料預處理與分析

資料預處理

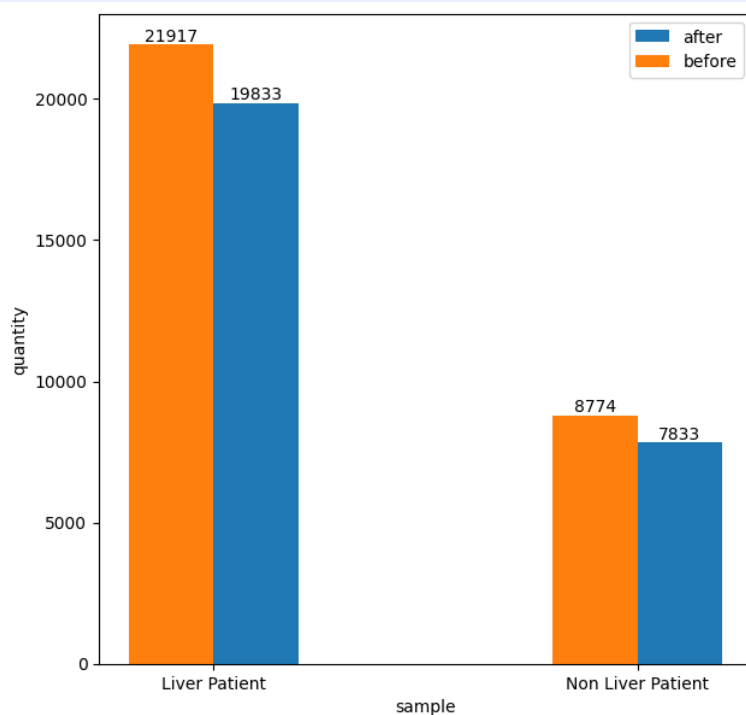
- 檢視具有缺失值之列的數量，並刪除



資料預處理與分析

資料分析

- 刪除缺失資料資料前、後的正負樣本數量



- 刪除前的正負樣本比值為

$$\frac{21917}{8774} = \frac{2.50}{1}$$

- 刪除後的正負樣本比值為

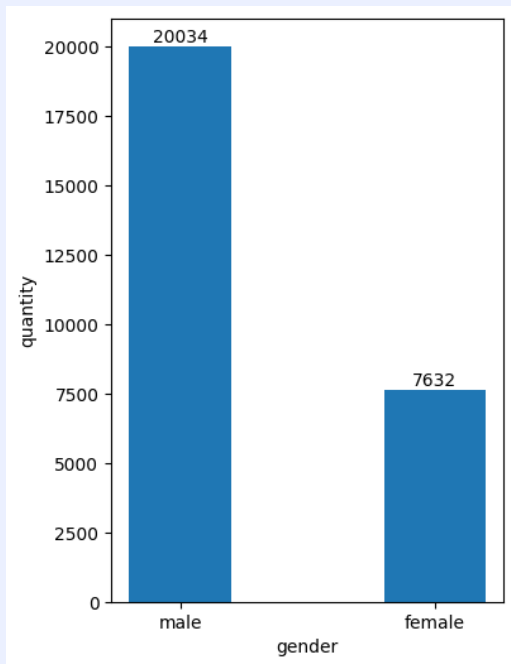
$$\frac{19833}{7833} = \frac{2.53}{1}$$

- 比值 < 3 ，較無數據不平衡之問題

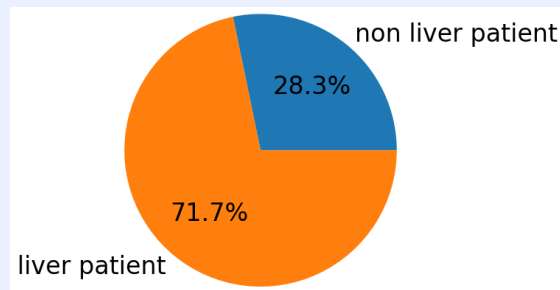
資料預處理與分析

資料分析

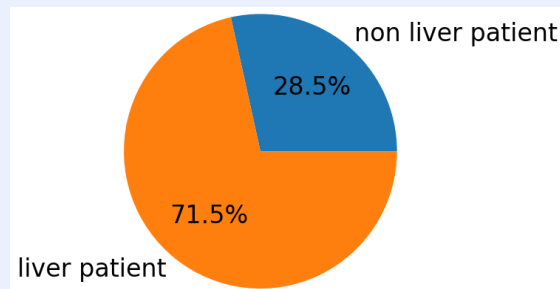
- 資料集中的男女數量



- 資料集中的男女患病比例



male

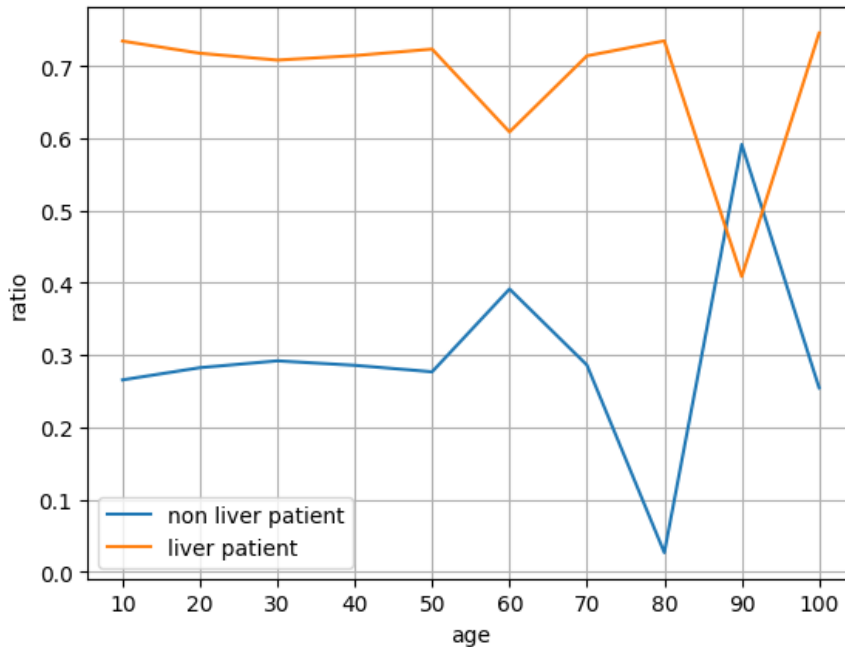


female

資料預處理與分析

資料分析

- 觀察年齡與是否患病的關係

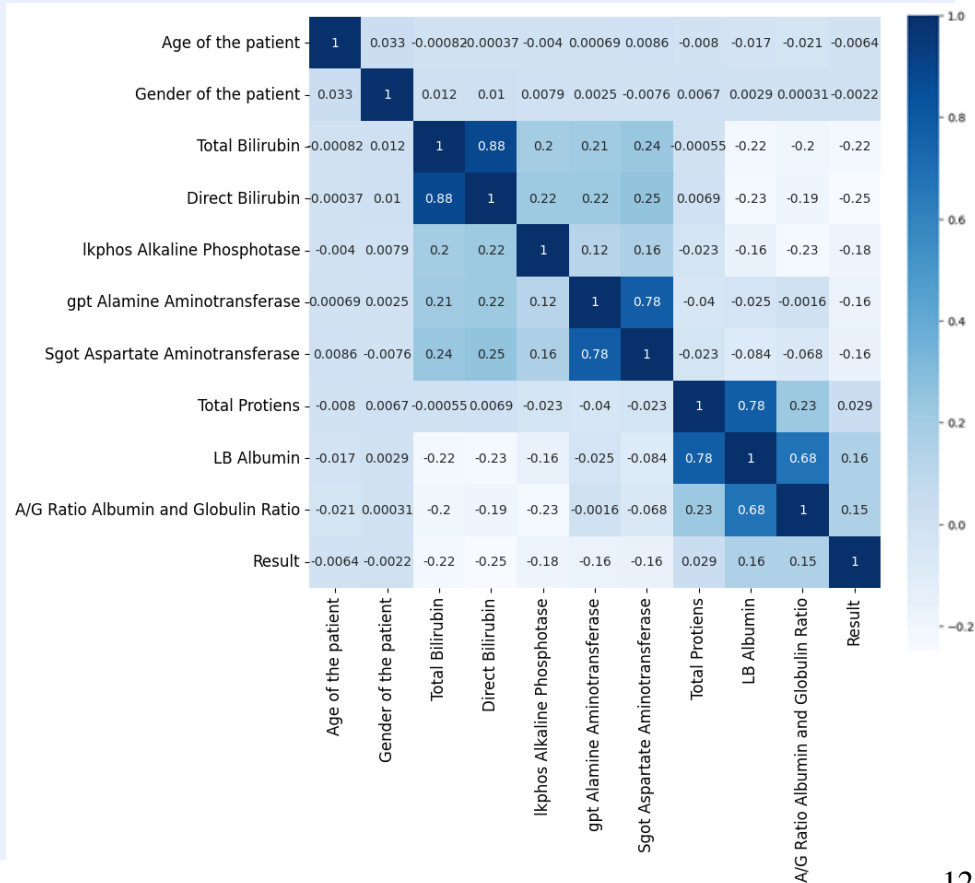


發現在50歲前，年齡
的上升不會對患病比
例造成影響

資料預處理與分析

特徵篩選

- 使用 seaborn 函式庫之熱力圖分析特徵
- 年齡、性別與是否患病的關係度較低

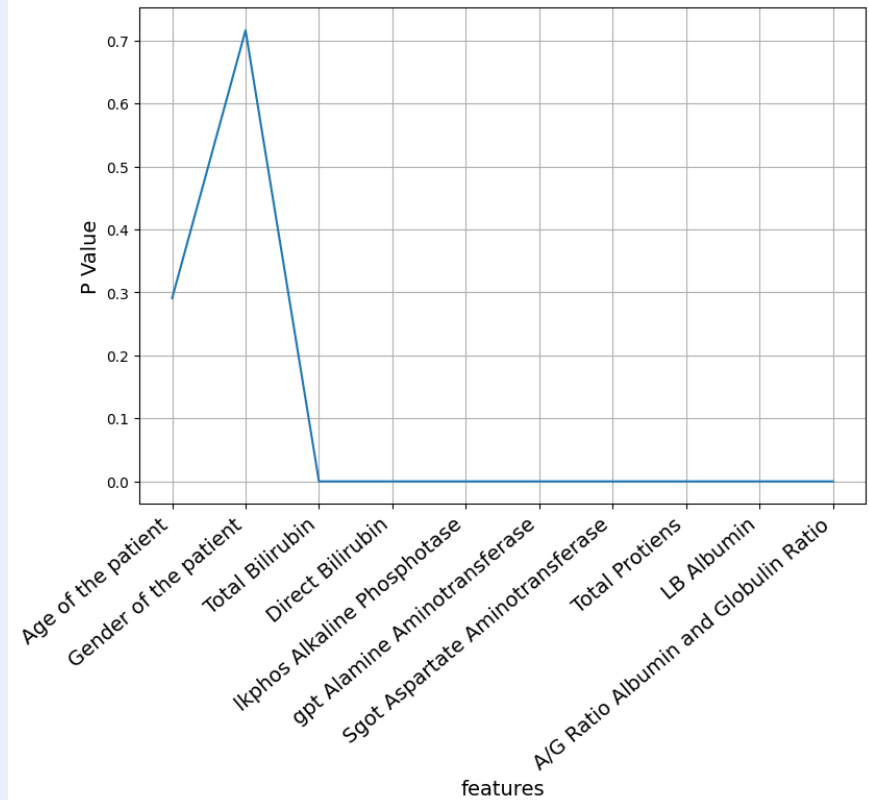


資料預處理與分析

特徵篩選

- 使用 scikit-learn 的 `feature_selection` 函式庫分析特徵
- $P \text{ Value} < 0.05$ 為顯著特徵
- 年齡、性別與是否患病的關係不顯著

使用模型評估時，不將年齡和性別納入考量



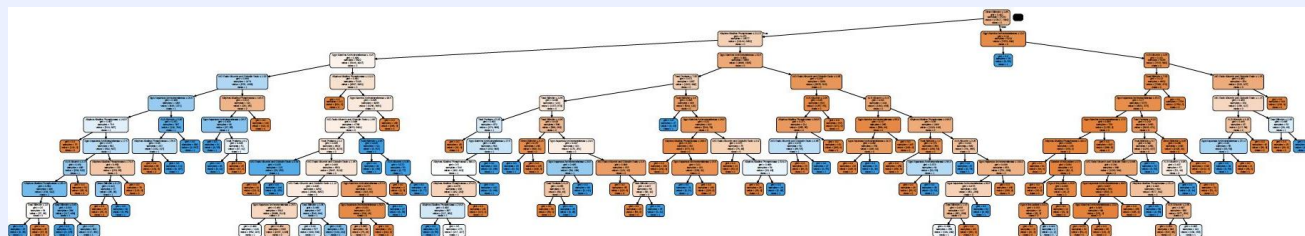
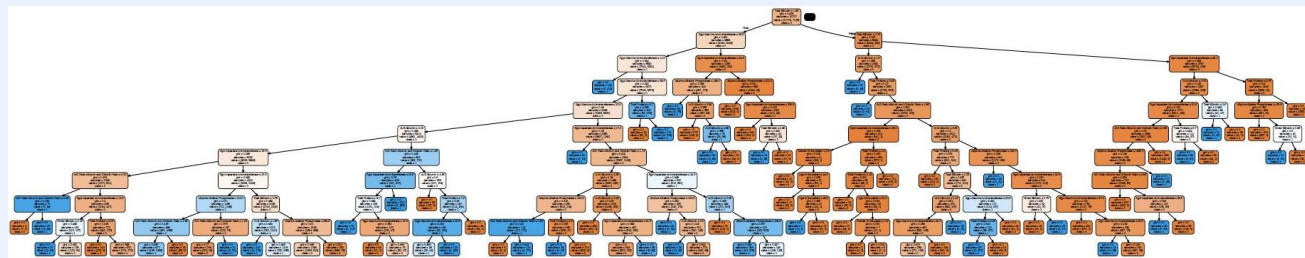
模型評估與結果

	01	02	03	04
model	Linear SVM	Non Linear SVM	Decision Tree	Random Forest
score	0.711	0.719	0.900	0.983

模型評估與結果-Random Forest

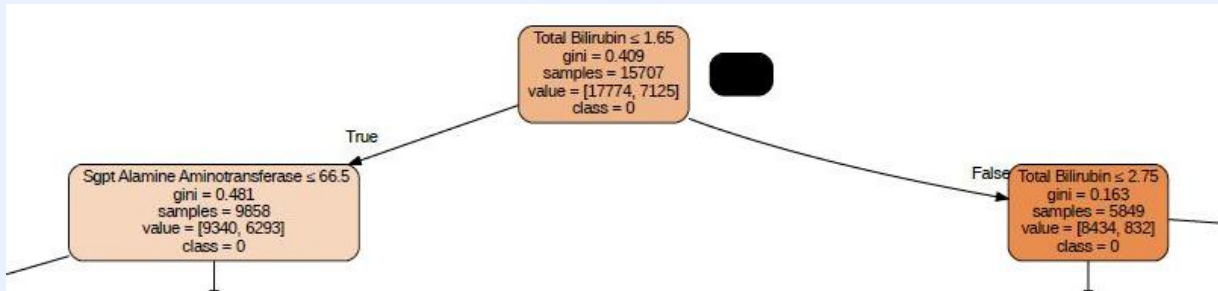
	Precision	Recall	F1-score
1	0.90	0.97	0.93
2	0.91	0.72	0.80
Accuracy			0.90
macro avg	0.90	0.85	0.87
weighted avg	0.90	0.90	0.90

模型評估與結果-Random Forest

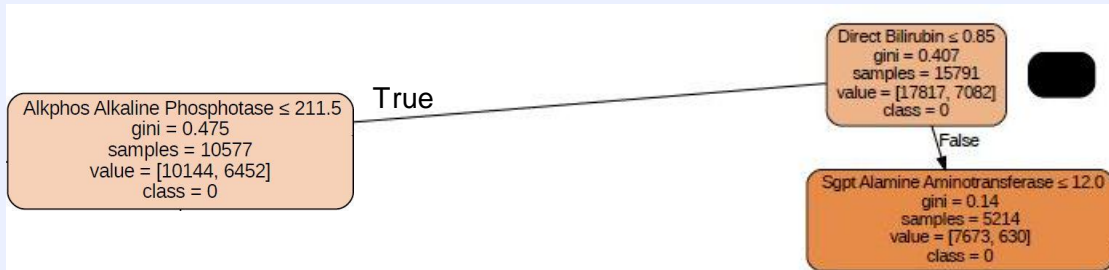


模型評估與結果-Random Forest

- 第一棵樹的第一、二階層的節點



- 第二棵樹的第一、二階層的節點



模型評估與結果-Random Forest

重要特徵

- Total Bilirubin：膽紅素，膽紅素值如果上升，就可能會有膽管阻塞、肝細胞受傷或紅血球破壞過多等情形發生。
- Sgpt Alanine Aminotransferase：gpt轉氨酶
- Sgot Aspartate Aminotransferase：got轉氨酶

GOT、GPT轉安酶存在於肝細胞內，當細胞壞死破壞之後，會釋
出到血液中，所以肝發炎時GOT、GPT轉安酶之值都可能上升。

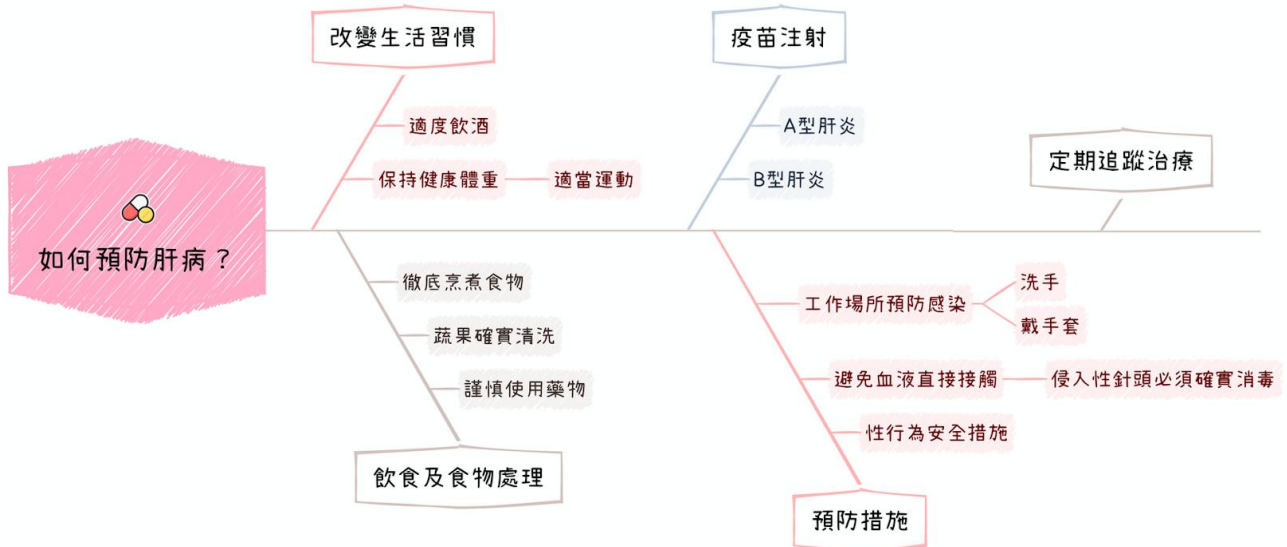
知識發現

指數變化原因：

1. 病毒感染：A、B、C、D、E肝炎
2. 飲食不當、生活作息不佳、仰賴酒精、藥物等多種因素

其實工作勞累不一定會爆肝，反而是“暴飲暴食”
更易有爆肝風險！

知識發現



結論

- 透過資料分析我們找到了推測肝病的重要指數(膽紅素、gpt轉氨酶、got轉氨酶)，並進一步地查出造成這些指數變動的原因。
- 找到原因就可以做出解決方法。
- 肝臟是沉默的器官，希望透過這個報告可以提高人們對於肝病的重視。

文獻參考

1. 中山醫學戴學附設醫院肝病防治中心：B型肝炎檢驗判讀
(<http://web.csh.org.tw/web/224020/index.php/123-4/7-2/>)
2. 聯安醫周刊：預防「肝苦」人生！五種常見肝病一次搞懂
(<https://www.lianan.com.tw/drliananepaper/Article/410-1>)
3. 康健：肝真的會「爆」嗎？過勞爆肝有沒有道理？
(<https://www.commonhealth.com.tw/article/85845>)
4. 17項永續發展目標
(<https://globalgoals.tw/3-good-health-and-well-being>)
5. 常見B肝迷思大破解！
(<https://today.line.me/tw/v2/article/x2XwjNw>)



Thank You

