

# 預測肥胖等級

報告者: 潘建豪



# 目錄



**01**

資料概述

**02**

圖表說明

**03**

問題與解決

**04**

總結

01

# 資料概述

# 01 資料概述



## 為何選此主題?



自己和身邊的朋友都隨著年紀的增長，有越來越肥胖的趨勢，所以想深入探討引響肥胖等級的因素，並做預測。

---

資料來源:

<https://www.kaggle.com/datasets/fateme/mehmehrparvar/obesity-levels>

# 01 資料概述

## 背景說明



特徵名稱	研究問題	答案	資料的值
Age	請問你的年齡?	年齡	Numeric value
Gender	請問你的性別?	男, 女	Male,Female
Height	請問你的身高?	單位:公分	Numeric value
Weight	請問你的體重?	單位:公斤	Numeric value
BMI	請問你的BMI?	BMI	Numeric value
CALC	請問你經常喝酒?	沒有, 偶爾, 經常, 總是	no,Sometimes, Frequently,Always
FAVC	請問你經常吃高熱量的食物嗎?	是, 否	yes,no
FCVC	請問你吃飯時經常會搭配蔬菜嗎?	沒有, 偶爾, 總是	1,2,3

參考資料:  
<https://www.mdpi.com/2076-3417/13/6/3875>

# 01 資料概述

## 背景說明



特徵名稱	研究問題	答案	資料的值
NCP	請問你一天吃幾頓正餐?	1頓或沒有，2頓，3頓，3頓以上	1,2,3,4
SCC	請問你會監測你一天吃多少卡路里嗎?	是，否	yes,no
SMOKE	請問你抽菸嗎?	是，否	yes,no
CH2O	請問你一天喝多少水?	少於1公升，1~2公升，大於2公升	1,2,3
FHWO	請問你的家庭成員中是否有人超重?	是，否	yes,no
FAF	請問你一個禮拜運動幾天?	沒有，1~2天，2~4天，4天以上	0,1,2,3
TUE	請問你一天花多少時間使用3C產品?	0~2小時，3~5小時，大於5小時	0,1,2

# 01 資料概述

## 背景說明



特徵名稱	研究問題	答案	資料的值
CAEC	請問你經常在兩餐之間吃任何食物嗎？	沒有，偶爾，經常，總是	no, Sometimes, Frequently, Always
MTRANS	請問你通常使用哪種交通工具？	汽車，腳踏車，機車，大眾交通運輸，走路	Automobile, Bike, Motorbike, Public_Transportation, Walking
NObeyesded	-	-	Insufficient_Weight, Normal_Weight, Overweight, Obesity_Type_I, Obesity_Type_II, Obesity_Type_III

# 01 資料概述

## 流程圖



### 資料蒐集

使用Kaggle下載資料

1

### 建立機器學習模型

分別使用LinearRegression，  
DecisionTree，RandomForest等

3

### 資料分析

利用圖表得出相關性

5

### 資料處理

使用Python的  
Pandas整理資料

2

### 視覺化

使用Python的  
matplotlib繪圖

4

### 與使用者互動

透過django讓使用者輸入個人資訊，並輸出預測的肥胖等級及BMI

6





02

## 圖表說明

## 02 圖表說明

## 準確率



未標準化

Model: LinearRegression  
score: 0.3355

Model: RandomForestClassifier  
Accuracy\_score: 0.9905

Model: LGBMClassifier  
Accuracy\_score: 0.9943

Model: DecisionTreeClassifier  
Accuracy\_score: 0.9924

Model: ExtraTreesClassifier  
Accuracy\_score: 0.9602

Model: KNN  
Accuracy\_score: 0.8996

Model: KMeans(標籤已修正)  
Accuracy\_score: 0.5656

標準化

Model: LinearRegression  
score: 0.3355

Model: RandomForestClassifier  
Accuracy\_score: 0.9905

Model: LGBMClassifier  
Accuracy\_score: 0.9943

Model: DecisionTreeClassifier  
Accuracy\_score: 0.9924

Model: ExtraTreesClassifier  
Accuracy\_score: 0.9602

Model: KNN  
Accuracy\_score: 0.7652

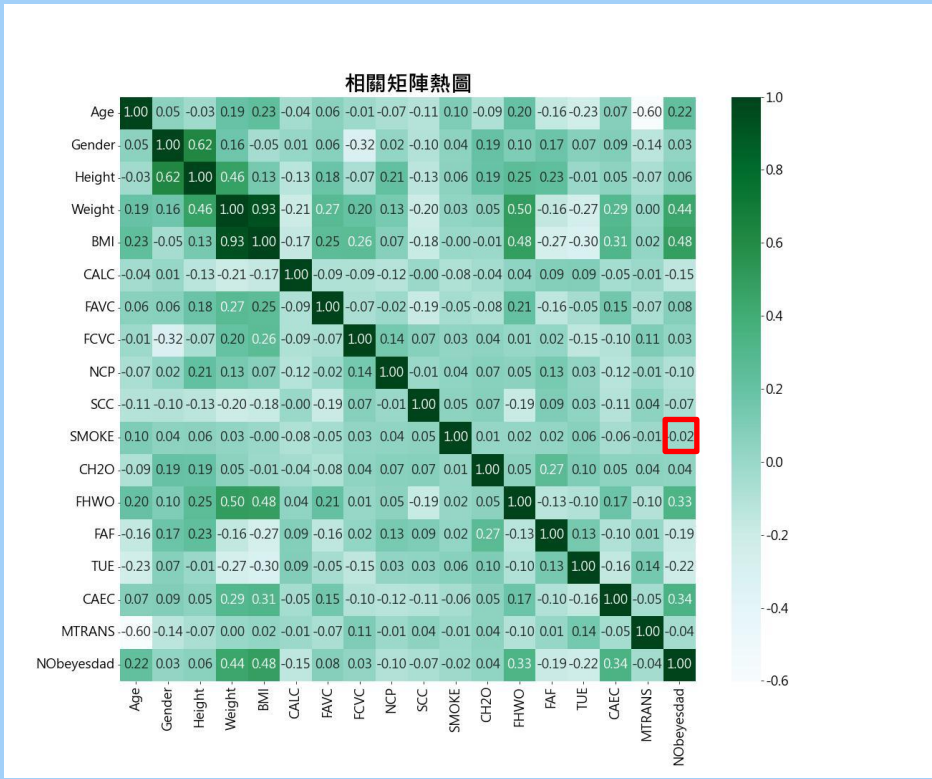
Model: KMeans(標籤已修正)  
Accuracy\_score: 0.4311

# 02 圖表說明

## 特徵篩選



SMOKE(請問你抽菸嗎?)



新的X未標準化

Model: LinearRegression  
score: 0.3351  
-----

Model: RandomForestClassifier  
Accuracy\_score: 0.9886  
-----

Model: LGBMClassifier  
Accuracy\_score: 0.9943  
-----

Model: DecisionTreeClassifier  
Accuracy\_score: 0.9924  
-----

Model: ExtraTreesClassifier  
Accuracy\_score: 0.9602  
-----

Model: KNN  
Accuracy\_score: 0.8996

## 02 圖表說明

## 使用者互動



肥胖等級預測	
年齡	<input type="text"/>
性別	<input checked="" type="radio"/> 男 <input type="radio"/> 女
身高	<input type="text"/> 單位:公分
體重	<input type="text"/> 單位:公斤
請問你經常喝酒?	<input type="radio"/> 總是 <input type="radio"/> 經常 <input checked="" type="radio"/> 偶爾 <input type="radio"/> 沒有
請問你經常吃高熱量的食物嗎?	<input type="radio"/> 是 <input checked="" type="radio"/> 否
請問你吃飯時經常會搭配蔬菜嗎?	<input checked="" type="radio"/> 總是 <input type="radio"/> 偶爾 <input type="radio"/> 沒有
請問你一天吃幾頓正餐?	<input type="radio"/> 1頓或沒有 <input type="radio"/> 2頓 <input checked="" type="radio"/> 3頓 <input type="radio"/> 3頓以上
請問你會監測你一天吃多少卡路里嗎?	<input type="radio"/> 是 <input checked="" type="radio"/> 否
請問你抽菸嗎?	<input type="radio"/> 是 <input checked="" type="radio"/> 否
請問你一天喝多少水?	<input type="radio"/> 少於1公升 <input checked="" type="radio"/> 1~2公升 <input type="radio"/> 大於2公升
請問你的家庭成員中是否有人超重?	<input type="radio"/> 是 <input checked="" type="radio"/> 否
請問你一個禮拜運動幾天?	<input checked="" type="radio"/> 沒有 <input type="radio"/> 1~2天 <input type="radio"/> 2~4天 <input type="radio"/> 4天以上
請問你一天花多少時間使用3C產品?	<input type="radio"/> 0~2小時 <input type="radio"/> 3~5小時 <input checked="" type="radio"/> 大於5小時
請問你經常在兩餐之間吃任何食物嗎?	<input type="radio"/> 總是 <input type="radio"/> 經常 <input checked="" type="radio"/> 偶爾 <input type="radio"/> 沒有
請問你通常使用哪種交通工具?	<input type="radio"/> 汽車 <input type="radio"/> 腳踏車 <input type="radio"/> 機車 <input checked="" type="radio"/> 大眾交通運輸 <input type="radio"/> 走路
<input type="button" value="送出"/> <input type="button" value="重設"/>	

## 02 圖表說明

## 使用者互動



肥胖等級預測	
年齡	<input type="text" value="32"/>
性別	<input checked="" type="radio"/> 男 <input type="radio"/> 女
身高	<input type="text" value="176.51"/>
體重	<input type="text" value="65.65"/>
請問你經常喝酒?	<input type="radio"/> 總是 <input type="radio"/> 經常 <input checked="" type="radio"/> 偶爾 <input type="radio"/> 沒有
請問你經常吃高熱量的食物嗎?	<input type="radio"/> 是 <input checked="" type="radio"/> 否
請問你吃飯時經常會搭配蔬菜嗎?	<input checked="" type="radio"/> 總是 <input type="radio"/> 偶爾 <input type="radio"/> 沒有
請問你一天吃幾頓正餐?	<input type="radio"/> 1頓或沒有 <input type="radio"/> 2頓 <input checked="" type="radio"/> 3頓 <input type="radio"/> 3頓以上
請問你會監測你一天吃多少卡路里嗎?	<input type="radio"/> 是 <input checked="" type="radio"/> 否
請問你抽菸嗎?	<input type="radio"/> 是 <input checked="" type="radio"/> 否
請問你一天喝多少水?	<input type="radio"/> 少於1公升 <input checked="" type="radio"/> 1~2公升 <input type="radio"/> 大於2公升
請問你的家庭成員中是否有人超重?	<input type="radio"/> 是 <input checked="" type="radio"/> 否
請問你一個禮拜運動幾天?	<input checked="" type="radio"/> 沒有 <input type="radio"/> 1~2天 <input type="radio"/> 2~4天 <input type="radio"/> 4天以上
請問你一天花多少時間使用3C產品?	<input type="radio"/> 0~2小時 <input type="radio"/> 3~5小時 <input checked="" type="radio"/> 大於5小時
請問你經常在兩餐之間吃任何食物嗎?	<input type="radio"/> 總是 <input type="radio"/> 經常 <input checked="" type="radio"/> 偶爾 <input type="radio"/> 沒有
請問你通常使用哪種交通工具?	<input type="radio"/> 汽車 <input type="radio"/> 腳踏車 <input type="radio"/> 機車 <input checked="" type="radio"/> 大眾交通運輸 <input type="radio"/> 走路
<input type="button" value="送出"/> <input type="button" value="重設"/>	

## 預測結果

Normal\_Weight(正常體重)

BMI: 21.07

[返回輸入頁面](#)



03

## 問題與解決

## 03 問題與解決

### 準確率不一樣？



未標準化

Model: LinearRegression  
score: 0.3355

Model: RandomForestClassifier  
Accuracy\_score: 0.9905

Model: LGBMClassifier  
Accuracy\_score: 0.9943

Model: DecisionTreeClassifier  
Accuracy\_score: 0.9924

Model: ExtraTreesClassifier  
Accuracy\_score: 0.9602

Model: KNN  
Accuracy\_score: 0.8996

Model: KMeans(標籤已修正)  
Accuracy\_score: 0.5656

未標準化

Model: LinearRegression  
score: 0.3355

Model: RandomForestClassifier  
Accuracy\_score: 0.9867

Model: LGBMClassifier  
Accuracy\_score: 0.9943

Model: DecisionTreeClassifier  
Accuracy\_score: 0.9924

Model: ExtraTreesClassifier  
Accuracy\_score: 0.9545

Model: KNN  
Accuracy\_score: 0.8996

Model: KMeans(標籤已修正)  
Accuracy\_score: 0.5656

## 03 問題與解決

## 如何解決?



```
tree_models = [RandomForestClassifier(),LGBMClassifier(verbose=-1),
               DecisionTreeClassifier(),ExtraTreesClassifier()]

#一般會設定test_size=0.2~test_size=0.25(預設值)，因此選擇test_size=0.25
#random_state則隨便選
for name, model in zip(tree_name,tree_models):
    XTrain,XTest,yTrain,yTest = train_test_split(x,y,test_size=0.25, random_state=30)
    Model = model
    Model.fit(XTrain, yTrain)
    pred = Model.predict(XTest)
    accuracy = accuracy_score(yTest, pred)
    print('Model: %s'%name)
    print('Accuracy_score: %.4f'%accuracy)
    print('-----')
```

```
tree_models = [RandomForestClassifier(random_state=30),LGBMClassifier(verbose=-1),
               DecisionTreeClassifier(),ExtraTreesClassifier(random_state=30)]

#一般會設定test_size=0.2~test_size=0.25(預設值)，因此選擇test_size=0.25
#random_state則隨便選
for name, model in zip(tree_name,tree_models):
    XTrain,XTest,yTrain,yTest = train_test_split(x,y,test_size=0.25, random_state=30)
    Model = model
    Model.fit(XTrain, yTrain)
    pred = Model.predict(XTest)
    accuracy = accuracy_score(yTest, pred)
    print('Model: %s'%name)
    print('Accuracy_score: %.4f'%accuracy)
    print('-----')
```



## 03 問題與解決

### 如何特徵篩選？



原本是想說看迴歸係數，刪除係數小的，再看準確率。但是每個人對小的定義不一樣，沒辦法量化，所以這方法行不通。

後來在網路上找到相關矩陣，而且可以使用seaborn搭配matplotlib來畫熱力圖(heatmap)。

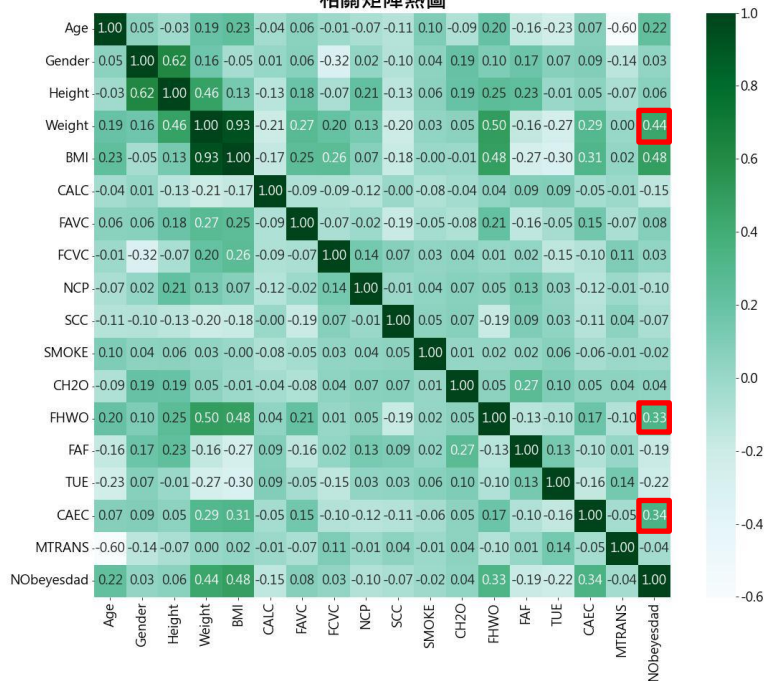
04

總結

# 04 總結



相關矩陣熱圖



- 從與肥胖等級的相關係數來看，前三名分別為
1. 體重(Weight)的0.44
  2. 請問你的家庭成員中是否有人超重?(FHWO)的0.33
  3. 你在兩餐吃間有沒有吃東西?(CAEC)的0.34
- 呈正相關(由於肥胖等級是由BMI分類而來的所以排除)，所以好好控制體重並且肚子餓不要再吃點心。

# 謝謝聆聽!

此專案GitHub的QRCode



CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.

