預測肥胖等級

報告者:潘建豪



目錄

資料概述

02 圖表說明

03 問題與解決

94 總結

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/fate mehmehrparvar/obesity-levels

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
ВМІ	請問你的BMI?	ВМІ	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.m dpi.com /2076-3417/13 /6/3875

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

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

流程圖

000

資料蒐集

使用Kaggle下載資料

建立機器學習模型

分別使用LinearRegression , DecisionTree , RandomForest等 資料分析

利用圖表得出相關性

資料處理

使用Python的 Pandas整理資料 視覺化

使用Python的 matplotlib繪圖 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

圖表說明



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

特徵篩選



相關矩陣熱圖 Age 1.00 0.05 -0.03 0.19 0.23 -0.04 0.06 -0.01 -0.07 -0.11 0.10 -0.09 0.20 -0.16 -0.23 0.07 -0.60 0.22 Gender- 0.05 1.00 0.62 0.16 -0.05 0.01 0.06 -0.32 0.02 -0.10 0.04 0.19 0.10 0.17 0.07 0.09 -0.14 0.03 Height -0.03 0.62 1.00 0.46 0.13 -0.13 0.18 -0.07 0.21 -0.13 0.06 0.19 0.25 0.23 -0.01 0.05 -0.07 0.06 Weight - 0.19 0.16 0.46 1.00 0.93 -0.21 0.27 0.20 0.13 -0.20 0.03 0.05 0.50 -0.16 -0.27 0.29 0.00 0.44 BMI - 0.23 -0.05 0.13 0.93 1.00 -0.17 0.25 0.26 0.07 -0.18 -0.00 -0.01 0.48 -0.27 -0.30 0.31 0.02 0.48 CALC -0.04 0.01 -0.13 -0.21 -0.17 1.00 -0.09 -0.09 -0.12 -0.00 -0.08 -0.04 0.04 0.09 0.09 -0.05 -0.01 -0.15 FAVC - 0.06 0.06 0.18 0.27 0.25 -0.09 1.00 -0.07 -0.02 -0.19 -0.05 -0.08 0.21 -0.16 -0.05 0.15 -0.07 0.08 FCVC --0.01 -0.32 -0.07 0.20 0.26 -0.09 -0.07 1.00 0.14 0.07 0.03 0.04 0.01 0.02 -0.15 -0.10 0.11 0.03 NCP -- 0.07 0.02 0.21 0.13 0.07 -0.12 -0.02 0.14 1.00 -0.01 0.04 0.07 0.05 0.13 0.03 -0.12 -0.01 -0.10 -0.2 SCC -0.11 -0.10 -0.13 -0.20 -0.18 -0.00 -0.19 0.07 -0.01 1.00 0.05 0.07 -0.19 0.09 0.03 -0.11 0.04 -0.07 SMOKE 0.10 0.04 0.06 0.03 -0.00 -0.08 -0.05 0.03 0.04 0.05 1.00 0.01 0.02 0.02 0.06 -0.06 -0.01 0.02 CH2O --0.09 0.19 0.19 0.05 -0.01 -0.04 -0.08 0.04 0.07 0.07 0.01 1.00 0.05 0.27 0.10 0.05 0.04 0.04 FHWO - 0.20 0.10 0.25 0.50 0.48 0.04 0.21 0.01 0.05 -0.19 0.02 0.05 1.00 -0.13 -0.10 0.17 -0.10 0.33 FAF --0.16 0.17 0.23 -0.16 -0.27 0.09 -0.16 0.02 0.13 0.09 0.02 0.27 -0.13 1.00 0.13 -0.10 0.01 -0.19 --0.2 TUE -0.23 0.07 -0.01 -0.27 -0.30 0.09 -0.05 -0.15 0.03 0.03 0.06 0.10 -0.10 0.13 1.00 -0.16 0.14 -0.22 CAEC - 0.07 0.09 0.05 0.29 0.31 -0.05 0.15 -0.10 -0.12 -0.11 -0.06 0.05 0.17 -0.10 -0.16 1.00 -0.05 0.34 MTRANS --0.60 -0.14 -0.07 0.00 0.02 -0.01 -0.07 0.11 -0.01 0.04 -0.01 0.04 -0.10 0.01 0.14 -0.05 1.00 -0.04 NObeyesdad - 0.22 0.03 0.06 0.44 0.48 -0.15 0.08 0.03 -0.10 -0.07 -0.02 0.04 0.33 -0.19 -0.22 0.34 -0.04 1.00

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
```

使用者互動



肥胖等級預測		
年齢		
性別	●男 ●女	
身高	單位·公分	
體重	單位:公斤	
請問你經常喝酒?	●總是 ●經常 ●偶爾 ●沒有	
請問你經常吃高熱量的食物嗎?	●是 o 否	
請問你吃飯時經常會搭配蔬菜嗎?	●總是 ●偶爾 ●沒有	
請問你一天吃幾頓正餐?	●1頓或沒有 ●2頓 ●3頓 ●3頓以上	
請問你會監測你一天吃多少卡路里嗎?	●是 ●否	
請問你抽菸嗎?	●是 ●否	
請問你一天喝多少水?	●少於1公升 ●1~2公升 ●大於2公升	
請問你的家庭成員中是否有人超重?	●是 ●否	
請問你一個禮拜運動幾天?	○ 沒有 ●1~2天 ●2~4天 ●4天以上	
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	
請問你經常在兩餐之間吃任何食物嗎?	●總是 ●經常 ●偶爾 ●沒有	
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	
	送出 重設	

使用者互動



肥胖等級預測		
年齢	32	
性別	●男 ●女	
身高	176.51	
體重	65.65	
請問你經常喝酒?	●總是 ●經常 ●偶爾 ●沒有	
請問你經常吃高熱量的食物嗎?	●是 ●否	
請問你吃飯時經常會搭配蔬菜嗎?	●總是 ●偶爾 ●沒有	
請問你一天吃幾頓正餐?	●1頓或沒有 ●2頓 ●3頓 ●3頓以上	
請問你會監測你一天吃多少卡路里嗎?	●是 ●否	
請問你抽菸嗎?	●是 ●否	
請問你一天喝多少水?	●少於1公升 ●1~2公升 ●大於2公升	
請問你的家庭成員中是否有人超重?	●是 •否	
請問你一個禮拜運動幾天?	●沒有 ●1~2天 ●2~4天 ●4天以上	
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	
請問你經常在兩餐之間吃任何食物嗎?	●總是 ●經常 ●偶爾 ●沒有	
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	
速出 重設		

預測結果

Normal_Weight(正常體重)

BMI: 21.07

返回輸入頁面

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?

03 問題與解決

如何特徵篩選?

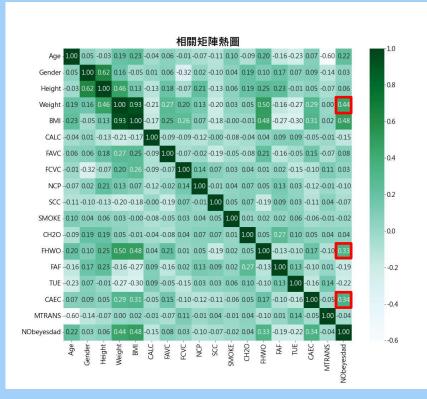
000

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)。

總結

04 總結



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.

