FIFA 2022 WORLD CUP

Group 7

B094020001林芳宇 | B094020007陳文薇 | B094020018王穎誠
B094020020林禹丞 | B094020041吳芃儀 | B094020039林峰瑋

TABLE OF CONTENTS

01

研究背景 & 資料集介紹

一夜致富只差一哩路!!

阿芳在今年2022FIFA比賽中輸得慘兮兮，就算看了多少報導，還是怎麼賭怎麼輸.....，所以她下定決心，準備在明年絕地大反攻!她想請求碰碰車團隊的協助，該團隊透過今年2022年的FIFA資料中各變數的變化，來對製造射門動作(SCA)進行預測。

團隊認為，想獲勝就要得分;想得分就是要射門，但射門得分與否，攸關於選手本身技巧，不過無論技巧優劣，都必須要射門!所以我們最終想預測哪些因素會影響到選手出腳射門的行為

資料集介紹

資料來源

Kaggle
FBREF

樣本數

669筆

欄位

168欄

資料集介紹

檔案	內容
player_defense	紀錄球員的防守行為
player_gca	紀錄射門成功的行為
player_misc	紀錄球員的犯規行為
player_passing	紀錄球員的傳球行為
player_passing_types	紀錄球員的傳球類型
player_playingtime	紀錄球員的出場資訊
player_possession	紀錄球員的盤球行為
player_shooting	紀錄球員的射門行為
player_stats	紀錄球員的基本資料

欄位介紹

欄位	說明
Position	紀錄球員的位置
touches_att_pen_area	在禁區觸球次數
passes_received	接到傳球次數
dribbles_complete	盤球成功次數
average _shot_distance	球員的平均射門距離
progressive_passes_received	球員接到漸進式傳球次數
dribbles_completed	球員完成盤球的次數
shot_free_kick	自由球
tackle	鏟球次數

02

資料前處理

合併資料集

Player_defense,player_gca,
player_misc, player_passing,.....

KNN補NA值

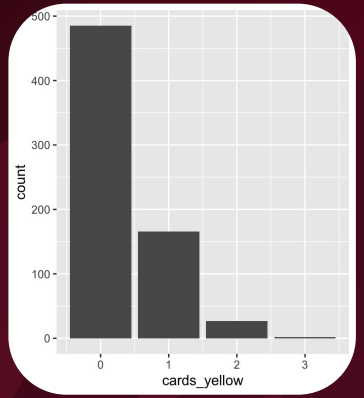
Scale, center

03

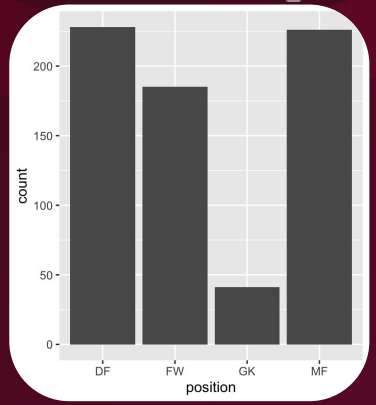
資料探索

單變量分析

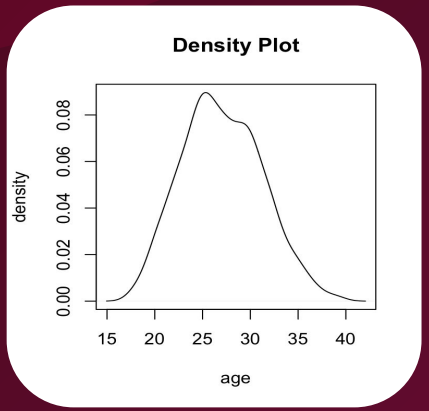
黃牌數量



位置



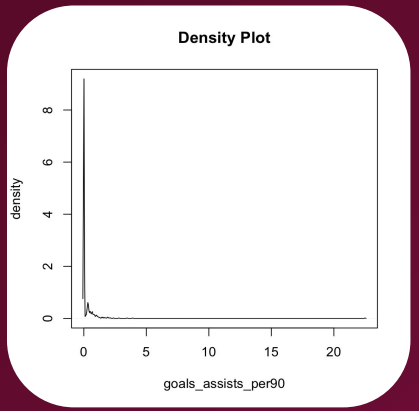
年齡




```
        Lilliefors (Kolmogorov-Smirnov) normality test

data:  player$age
D = 0.07924, p-value = 6.43e-11
```

```
> fivenum(player$age)
[1] 18 24 27 30 39
```

每場的進球與進攻次數




```
        Lilliefors (Kolmogorov-Smirnov) normality test

data:  player$goals_assists_per90
D = 0.41323, p-value < 2.2e-16
```

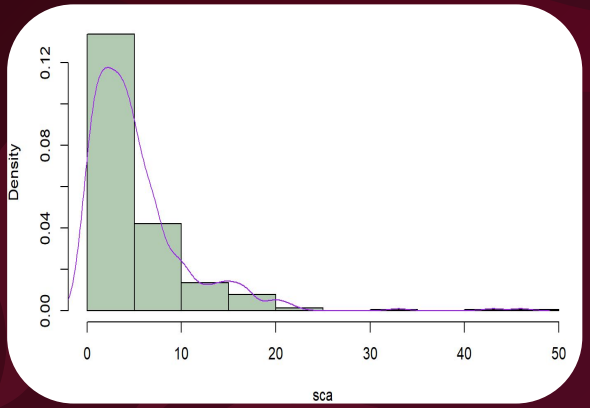
```
> fivenum(player$goals_assists_per90)
[1]  0.00  0.00  0.00  0.14 22.50
```

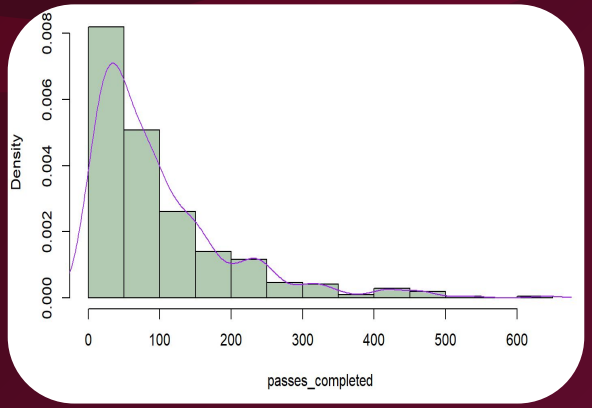
單變量分析

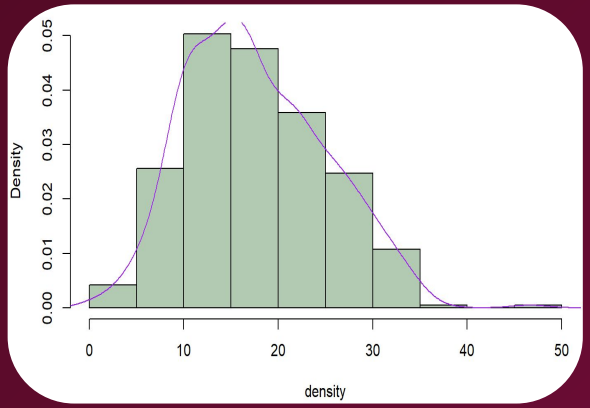
直接或間接導致
射門的動作

傳球成功

平均射程








```
        Lilliefors (Kolmogorov-Smirnov) normality test

data:  player$sca
D = 0.18902, p-value < 2.2e-16
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	2.000	4.000	5.273	7.000	46.000

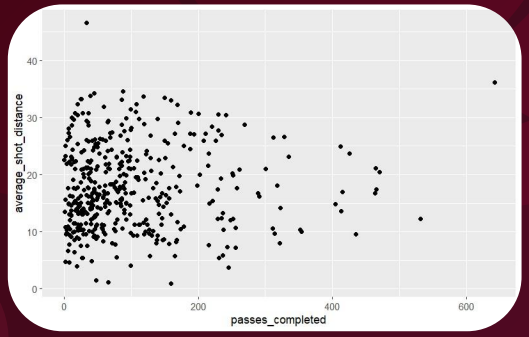
```
        Lilliefors (Kolmogorov-Smirnov) normality test

data:  player$passes_completed
D = 0.16193, p-value < 2.2e-16
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	33.00	69.00	99.41	136.00	642.00

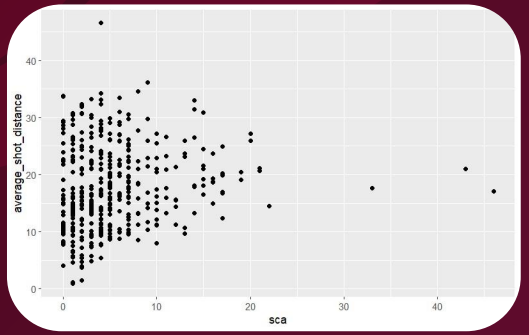
```
        Lilliefors (Kolmogorov-Smirnov) normality test

data:  player$average_shot_distance
D = 0.060234, p-value = 0.0007695
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.90	12.00	16.70	17.66	22.60	46.60

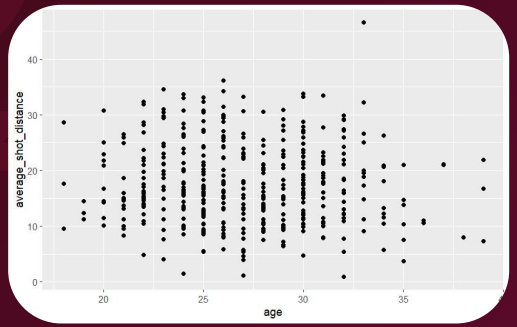
雙變量分析



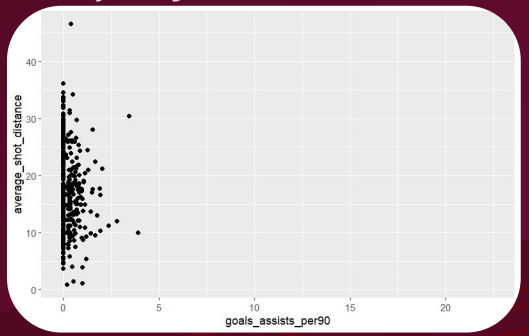
x=Passes_completed
y=avg_shot_distance



x=sca
y=avg_shot_distance

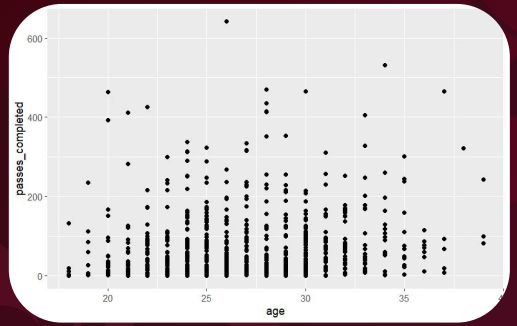


x=age
y=avg_shot_distance

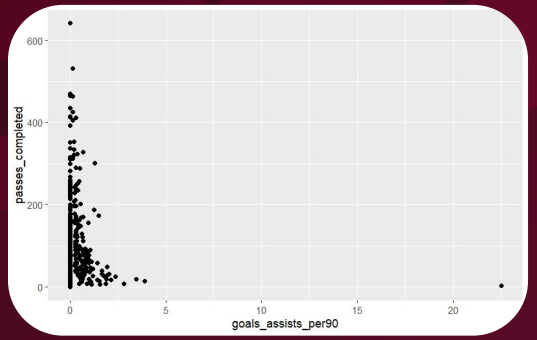


x=goals_assists_per90
y=avg_shot_distance

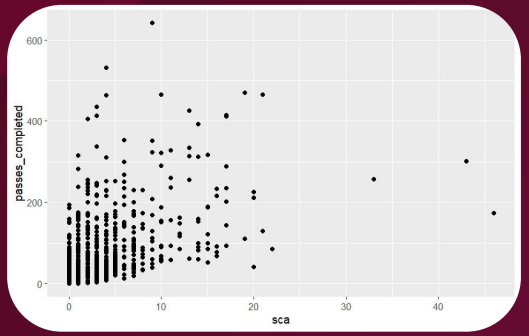
雙變量分析



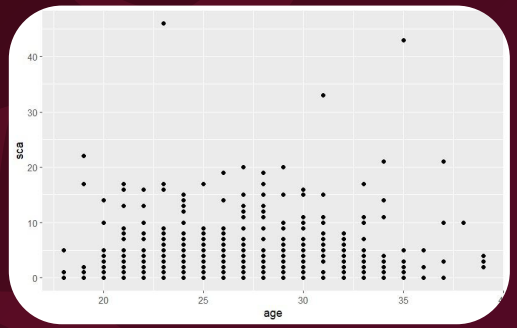
x=age
y=passes_completed



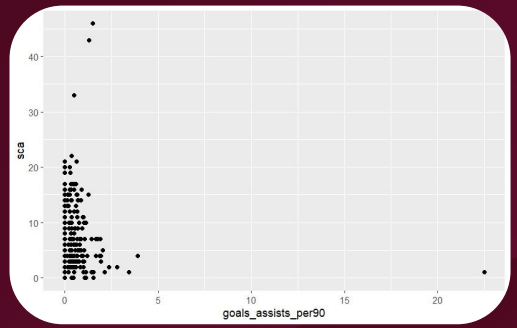
x=goal_assists_per90
y=passes_completed



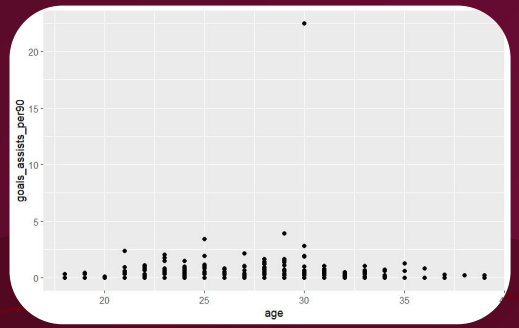
x=sca
y=passes_completed



x=sca
y=passes_completed

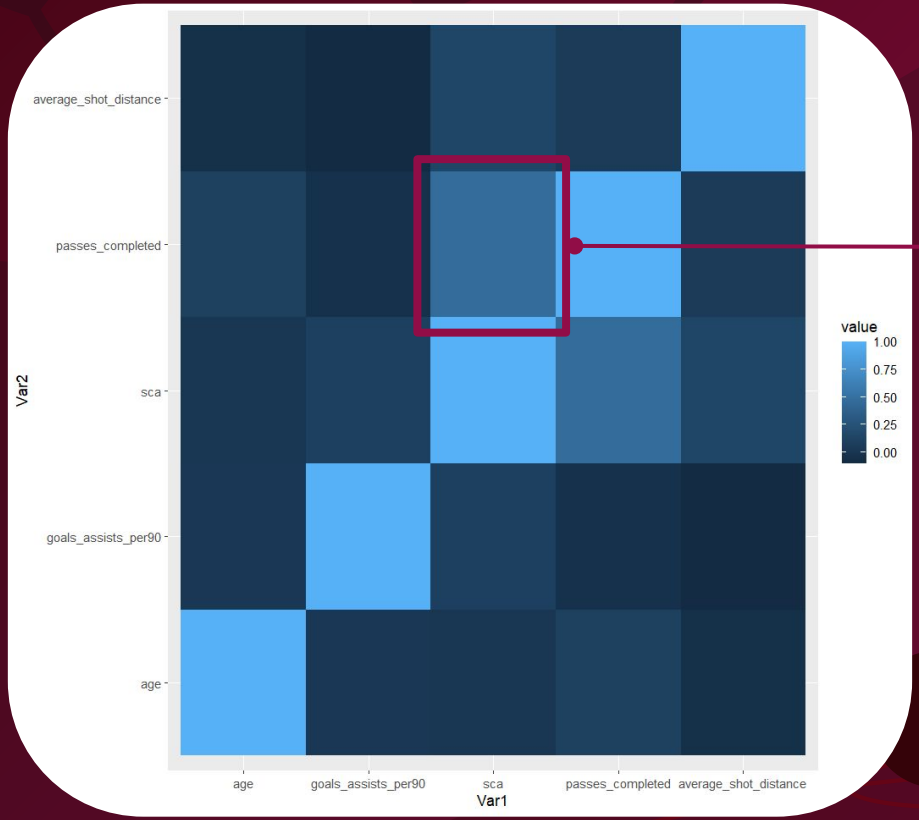


x=sca
y=passes_completed



x=age
y=goals_assists_per90

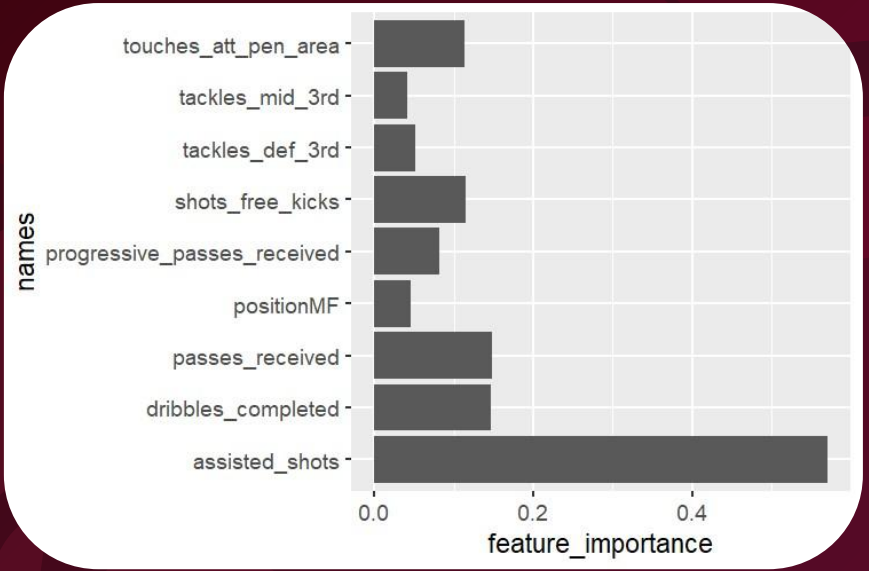
相關性分析


average_shot_distance
passes_completed
sca
goals_assists_per90
age
Var2
age
goals_assists_per90
sca
passes_completed
average_shot_distance
Var1
value
1.00
0.75
0.50
0.25
0.00
完成傳球次數

04

建立模型

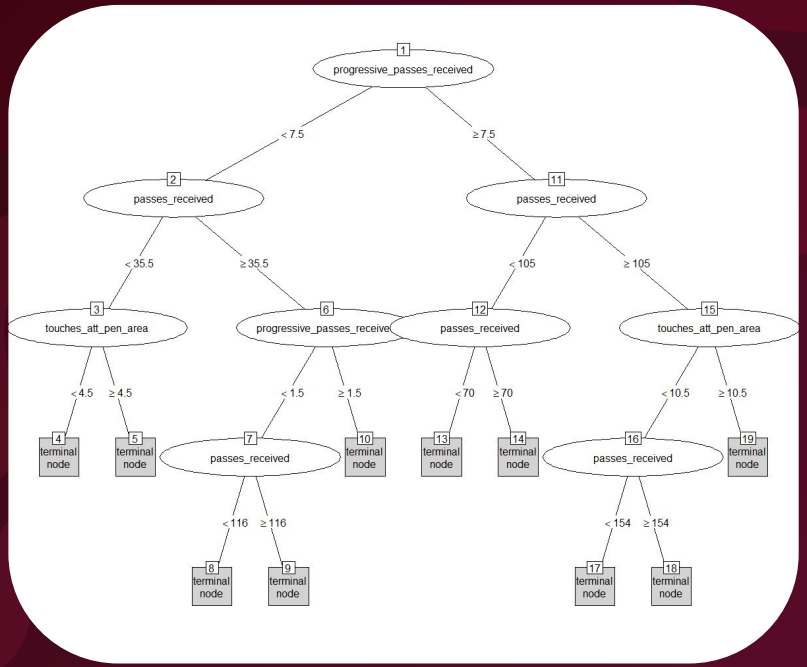
線性回歸 LM

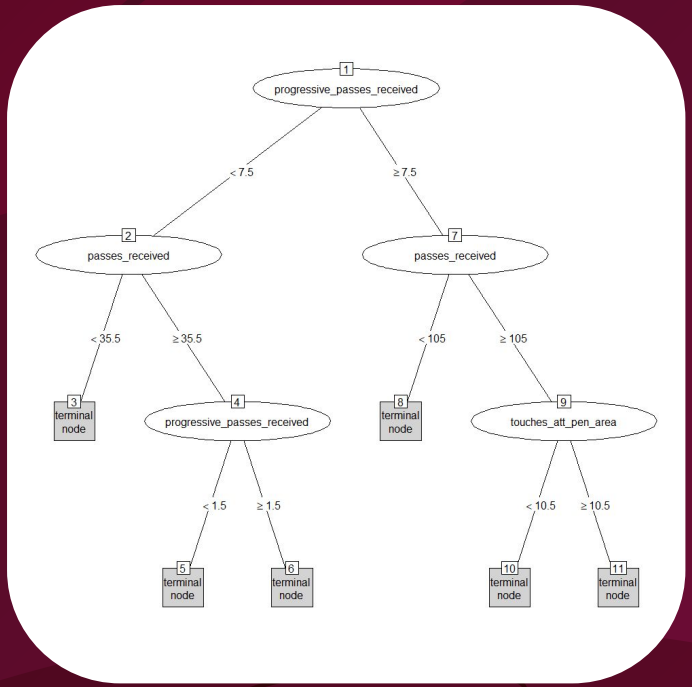



```
esiduals:
    Min      1Q  Median      3Q     Max
-6.4936 -0.8575 -0.1400  0.5914 11.0604

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                 3.34121    0.05939  56.261  < 2e-16 ***
positionMF                  0.17261    0.06322   2.730  0.00650 **
tackles_def_3rd             0.21004    0.06806   3.086  0.00211 **
tackles_mid_3rd             0.15318    0.06847   2.237  0.02561 *
assisted_shots              2.25287    0.08263  27.264  < 2e-16 ***
touches_att_pen_area        0.45439    0.10528   4.316 1.84e-05 ***
dribbles_completed          0.57181    0.07321   7.811 2.26e-14 ***
passes_received             0.56901    0.07186   7.918 1.04e-14 ***
progressive_passes_received 0.33377    0.10537   3.168  0.00161 **
shots_free_kicks            0.44525    0.06209   7.171 2.01e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.531 on 655 degrees of freedom
Multiple R-squared:  0.8517,    Adjusted R-squared:  0.8497
-statistic: 417.9 on 9 and 655 DF,  p-value: < 2.2e-16
```

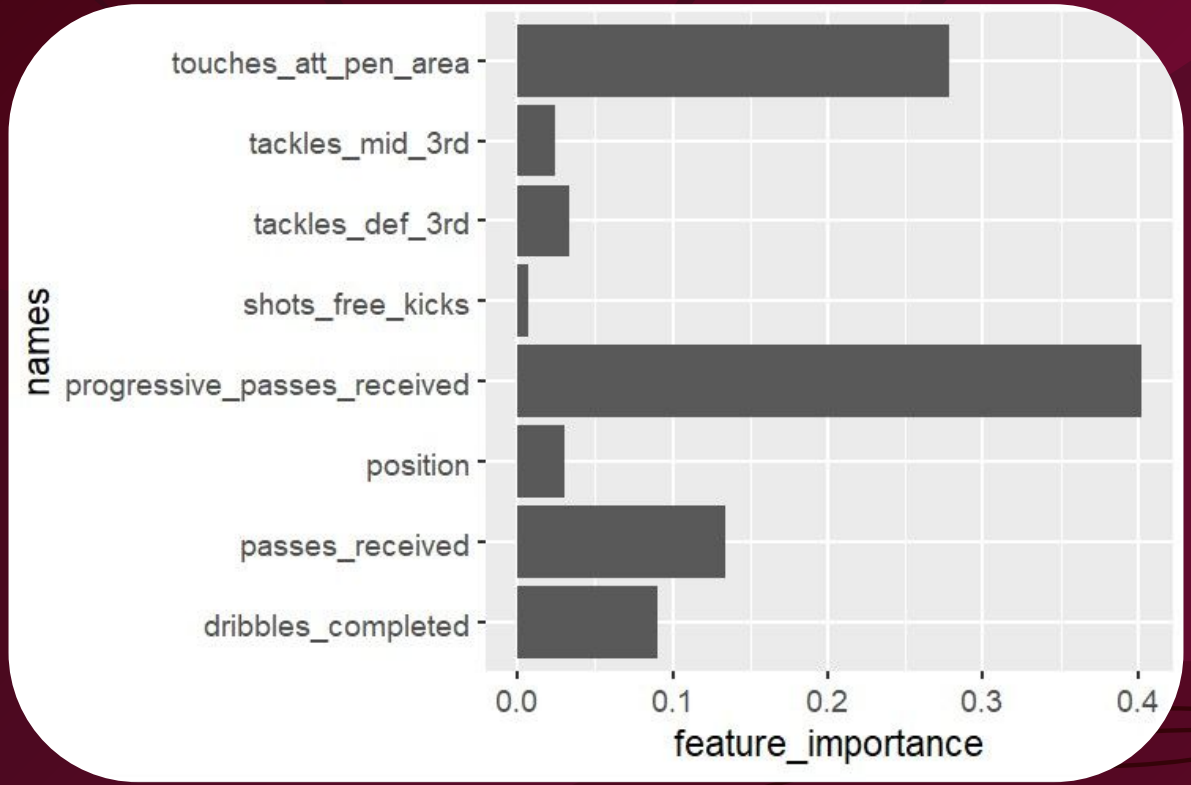
#SCA

決策樹

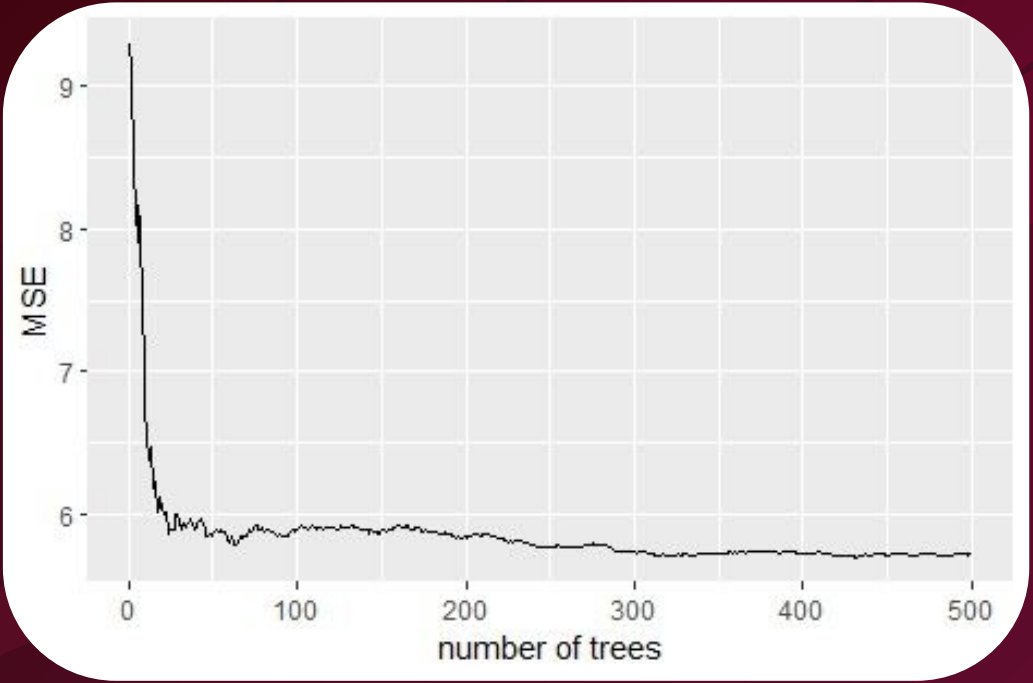

1
progressive_passes_received
< 7.5
≥ 7.5
2
passes_received
11
passes_received
< 35.5
≥ 35.5
< 105
≥ 105
3
touches_att_pen_area
6
progressive_passes_receive
12
passes_received
15
touches_att_pen_area
< 4.5
≥ 4.5
< 1.5
≥ 1.5
< 70
≥ 70
< 10.5
≥ 10.5
4
terminal node
5
terminal node
7
passes_received
10
terminal node
13
terminal node
14
terminal node
16
passes_received
19
terminal node
< 116
≥ 116
< 154
≥ 154
8
terminal node
9
terminal node
17
terminal node
18
terminal node


1
progressive_passes_received
< 7.5
≥ 7.5
2
passes_received
7
passes_received
< 35.5
≥ 35.5
< 105
≥ 105
3
terminal node
4
progressive_passes_received
8
terminal node
9
touches_att_pen_area
< 1.5
≥ 1.5
< 10.5
≥ 10.5
5
terminal node
6
terminal node
10
terminal node
11
terminal node

決策樹

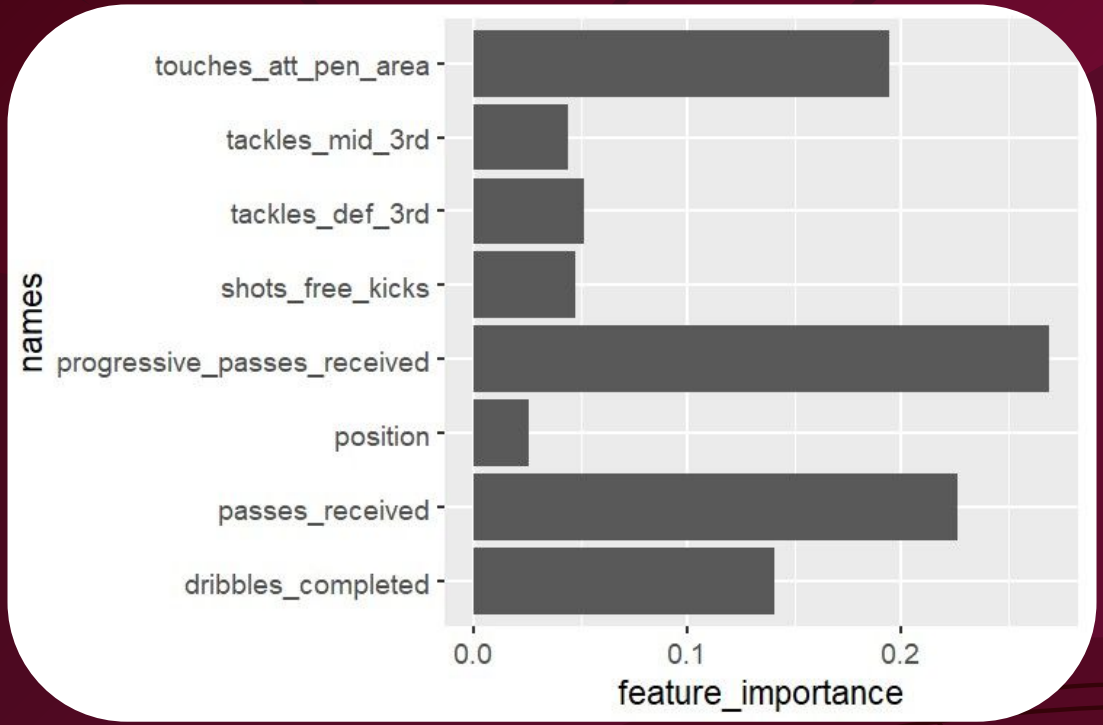

touches_att_pen_area
tackles_mid_3rd
tackles_def_3rd
shots_free_kicks
progressive_passes_received
position
passes_received
dribbles_completed
names
0.0
0.1
0.2
0.3
0.4
feature_importance

隨機森林



樹的數量大約在
300棵時趨於穩定

隨機森林


touches_att_pen_area
tackles_mid_3rd
tackles_def_3rd
shots_free_kicks
progressive_passes_received
position
passes_received
dribbles_completed
names
0.0
0.1
0.2
feature_importance

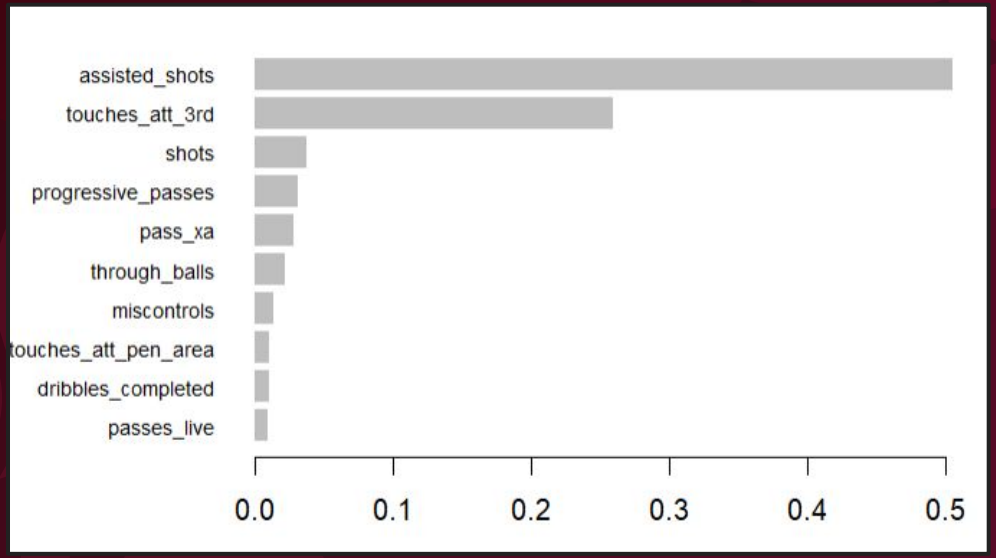
隨機森林

```
> train_player_RMSE
$RF_50
[1] 1.304398

$RF_150
[1] 1.272153

$RF_200
[1] 1.2564

$RF_300
[1] 1.25535

$RF_350
[1] 1.254881

$RF_400
[1] 1.261285
```

```
> test_player_RMSE
$RF_50
[1] 1.341737

$RF_80
[1] 1.335853

$RF_100
[1] 1.324001

$RF_200
[1] 1.305567

$RF_300
[1] 1.305142

$RF_350
[1] 1.297036

$RF_400
[1] 1.300944
```

模型訓練結果顯示
RF在350棵時
的RMSE最低

XGBOOST 極限梯度提升



	eta	max_depth	min_child_weight	subsample	colsample_bytree	nrounds	RMSE
1	0.05	3	3	0.65	0.8	85	1.716030
2	0.05	3	3	0.65	0.9	90	1.720187
3	0.10	3	3	0.80	0.9	41	1.727919
4	0.05	3	7	0.80	0.8	76	1.735950
5	0.05	3	5	0.65	0.8	76	1.737663
6	0.05	3	3	0.80	0.8	78	1.739665
7	0.10	3	5	0.65	1.0	46	1.740054
8	0.05	3	7	0.65	0.9	76	1.740452
9	0.05	3	7	0.80	0.9	78	1.741311
10	0.10	3	5	0.65	0.9	49	1.741320

eta = 0.05, max_depth = 3, min_child_weight = 3, subsample = 0.65, colsample_bytree = 0.8, nrounds = 85 時RMSE最小

#SCA

模型表現評估 *在此以RMSE做評估

Model	Train	Test
Linear regression	1.471359	1.626082
XGBoost	1.236659	1.975852
Regression Tree	2.47325	2.947111
Random Forest ($RF_350)	1.254881	1.305142

共同重要特徵

重要性	Random Forest	Regression Tree	XGBoost	Linear regression
1	Progressive_passes_received	Progressive_passes_received	assisted_shots	assisted_shots
2	touches_att_pen_area	touches_att_pen_area	touches_att_3rd	passes_received
3	passes_received	passes_received	shots	dribbles_completed
4	dribbles_completed	dribbles_completed	Progressive_passes	shot_free_kicks
5	tackles_mid_3rd	tackles_def_3rd	pass_xa	touches_att_pen_area

05

結論

如何增加射門次數?

根據共同重要特徵結果，要增加球隊的
射門行為，我們可以...

- **提高傳接球精準度**
- **加強盤球穩定度**
- **積極鏟球**

我們原本以為有影響的，但實際沒有影響的

1. **年齡**
2. **完成傳球次數**

THANKS

2022