

# 機器學習作業3

H54084078

陳詠君

資訊112乙

# INTRODUCTION

作業題目：snake

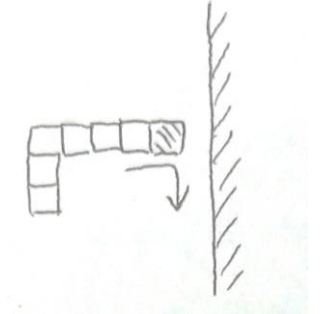
利用Artificial Intelligence技術讓蛇學習如何自己吃到食物，並且達到蛇身長度為30的目標。

觀察到的現象：蛇會死亡的原因為：1.碰到自己的身體 2.撞牆，因此寫rule的時候要考慮這兩個狀況，讓蛇不死亡比一味地讓蛇一直往食物的方向移動重要。另外，這次可以拿來訓練的模型跟作業1，2很不一樣，因此此次作業的重頭戲為模型的選擇。

# METHOD

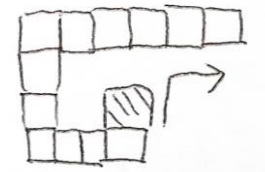
## 1. Rule :

首先，考慮撞牆，如果頭與牆的座標只差10，則return使蛇轉彎；

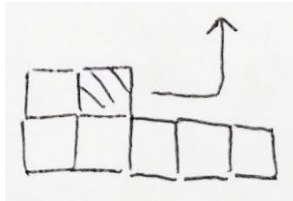


其次，考慮碰到身體，可以分成兩種情況：

第1種情況：比如說蛇往上時，如果頭的x座標跟某段身體的x座標一樣，而頭的y座標與該段身體的y座標相差10，則return。



第2種情況：比如說蛇往右時，如果頭的x座標與某段身體的x座標一樣，而頭的y座標與該段身體的y座標只差10，則return，可以避免頭與身體距離太近導致撞到身體。



最後，才是讓頭往食物的方向前進。

# METHOD

2.提取之特徵(9個)：(1)頭的x座標、(2)頭的y座標、(3)蛇的前進方向、(4)頭距離牆壁的x向量距離、(5)頭距離牆壁的y向量距離、(6)食物的x座標、(7)食物的y座標、(8)body、(9)wall 共9個。

其中，頭的x座標、頭的y座標、食物的x座標、食物的y座標，只要從scene\_info提取即可，其他特徵則要另外算。

3.pickle檔的蒐集： score > 30的pickle檔(150筆)。

# METHOD

## 4.機器學習模型：

### (1)Knn

在knn分類中，輸出是一個分類族群。一個物件的分類是由其鄰居的「多數表決」決定的， $k$ 個最近鄰居( $k$ 為正整數，通常較小)中最常見的分類決定了賦予該物件的類別。Ex： $k=5$ ，該物件的類別由五個節點賦予。

### (2)Extra Tree

extra trees對於每個決策樹的訓練集，不採用隨機採樣，即每個決策樹採用原始訓練集，且會隨機的選擇一個特徵值來劃分決策樹。

由於隨機選擇了特徵值的劃分點位，而不是最優點位，這樣會導致生成的決策樹的規模一般會較大。也就是說，模型的方差較少，但是偏倚較大。

# METHOD

## (3) Decision Tree

**Decision tree**是一個預測模型，他代表的是對象屬性與對象值之間的一種映射關係。樹中每個節點表示某個對象，而每個分叉路徑則代表某個可能的屬性值，而每個葉節點則對應從根節點到該葉節點所經歷的路徑所表示的對象的值。決策樹僅有單一輸出，若欲有複數輸出，可以建立獨立的決策樹以處理不同輸出。

## (4) Multilayer Perceptron (MLP)

**MLP Classifier**是遵循人類神經系統原理，學習並進行數據預測。它首先學習，然後使用權重存儲數據，並使用算法來調整權重並減少訓練過程中的偏差，即實際值和預測值之間的誤差。

由於**MLP Classifier**運用算法來減少實際值與預測值的偏差，因此選用**MLP Classifier**訓練。

# RESULT

MLP每局跑5次，共5局的結果

Score: 68  
Score: 43  
Score: 35  
Score: 43  
Score: 37

Score: 20  
Score: 61  
Score: 43  
Score: 10  
Score: 50

Score: 49  
Score: 20  
Score: 51  
Score: 22  
Score: 61

Score: 46  
Score: 58  
Score: 52  
Score: 68  
Score: 43

Score: 36  
Score: 50  
Score: 23  
Score: 75  
Score: 54

# RESULT

score \	第1局	第2局	第3局	第4局	第5局
<30	0	2	2	2	1
>=30	5	3	3	3	4

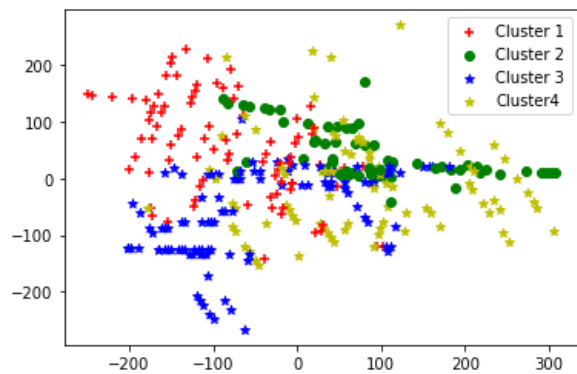


用MLP跑的結果



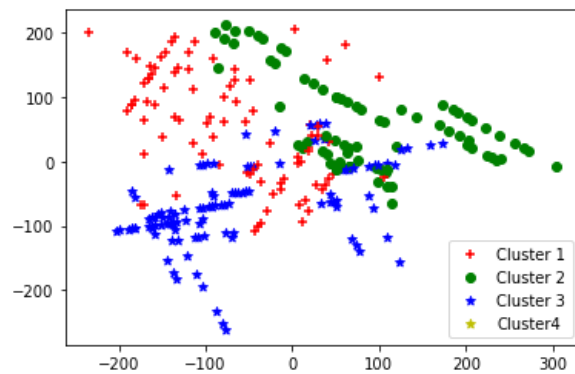
# DISCUSSION

問題一：以不同的機器學習演算法處理同一筆data會有甚麼差異？



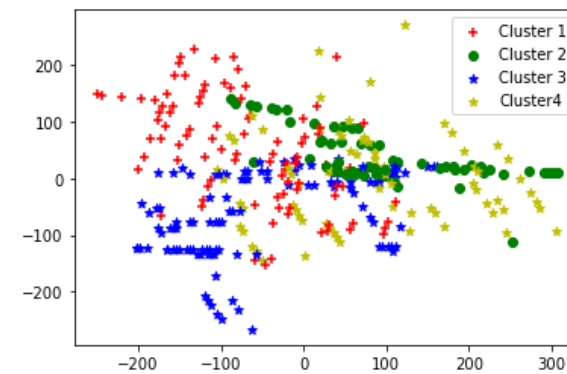
MLP

Precision=0.891



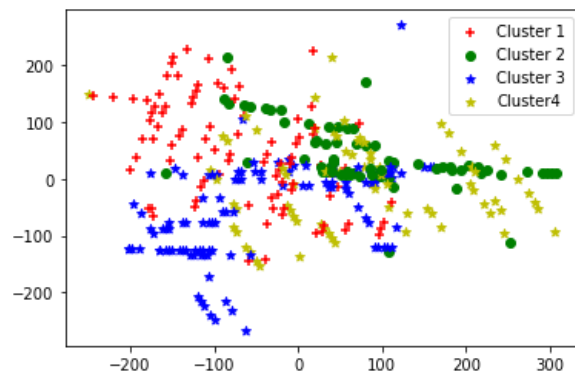
KNN

precision=0.92



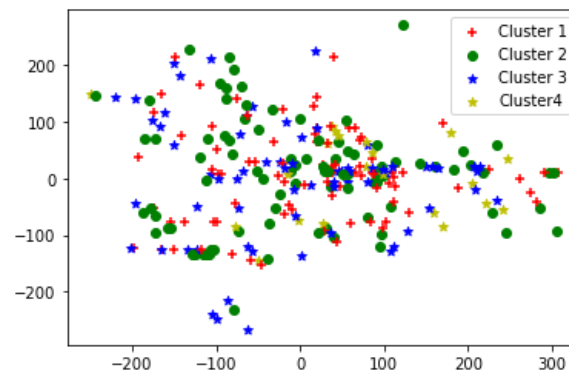
Extra Tree

precision=0.93



Decision Tree

Precision=0.92



Actual Label

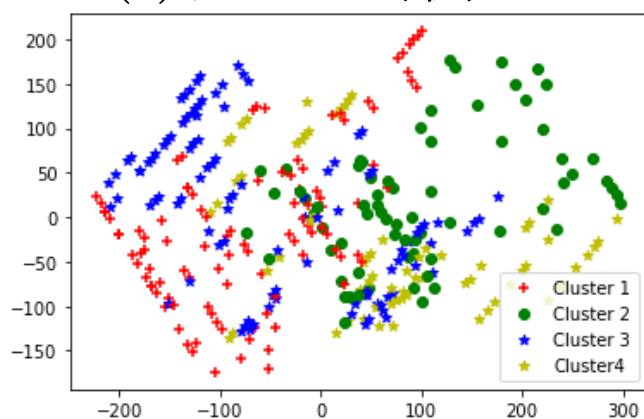
# DISCUSSION

問題一：以不同的機器學習演算法處理同一筆data會有甚麼差異？

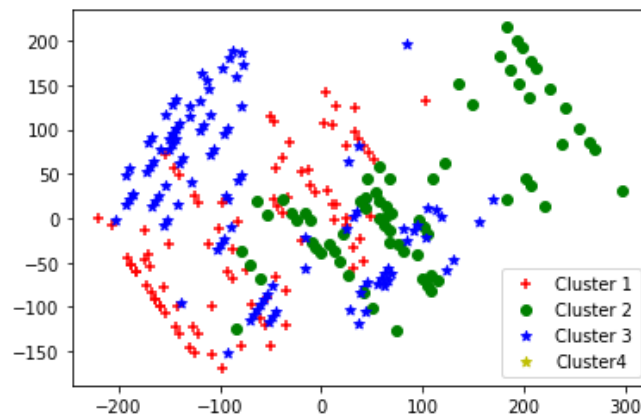
上頁投影片的5張圖，是將所選取的特徵參數以PCA的方式投影到二維畫面所畫出的點，分別以紅、綠、藍、黃將模型預測答案為comment = 0(“UP”)、comment = 1(“DOWN”)、comment = 2 (“LEFT”)、comment = 3 (“RIGHT”)。MLP、KNN、Extra Tree、Decision Tree皆為監督式學習，學習的過程與我們所選的答案有關，且四者正確率都超過0.89，但是knn預測起來幾乎沒有黃色的點，所以即使accuracy高達0.92但卻與正確label會有差異，而MLP、Extra Tree、Decision Tree呈現出來的圖長的差不多，因此三者都是很好的模型，經過反覆試驗，決定選用MLP做為這次訓練貪食蛇的模型。

問題二：所選的特徵數量多寡、是否具有代表性，也會影響到模型學習的成效。

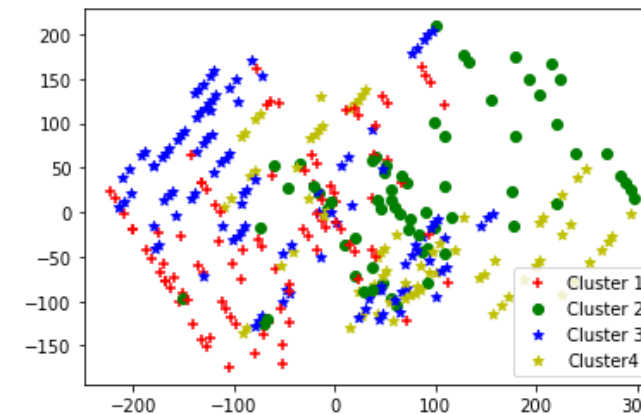
(1)以 wall 為例：



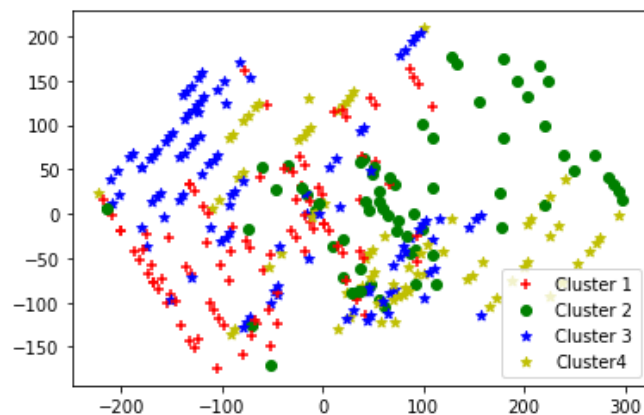
MLP



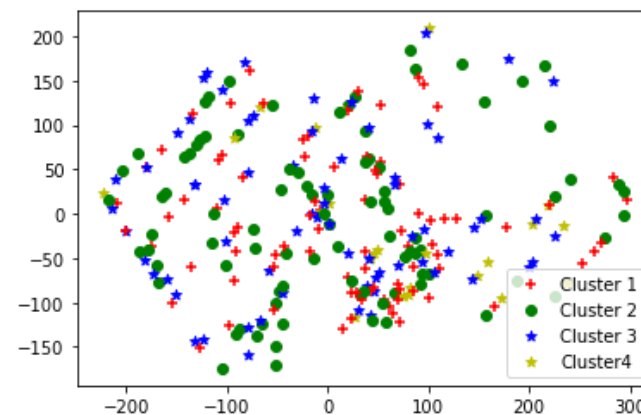
KNN



Extra Tree



Decision Tree

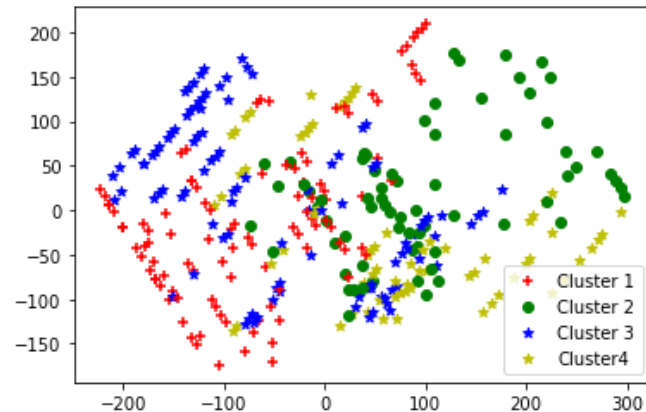


Actual Label

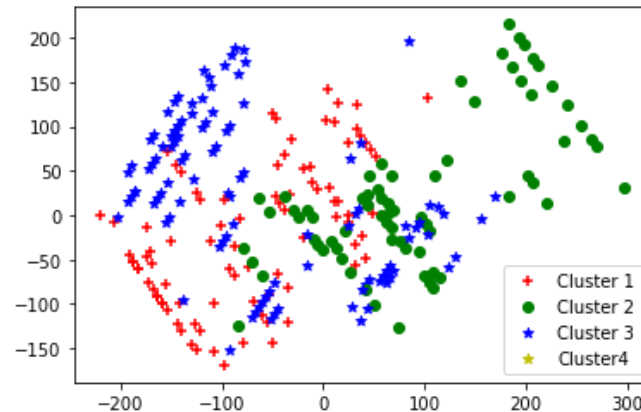
少取 wall 當feature，各模型的accuracy皆下降了5%左右，圖形也與原本的圖形差非常大，因此 wall 是極具有代表性的特徵。

問題二：所選的特徵數量多寡、是否具有代表性，也會影響到模型學習的成效。

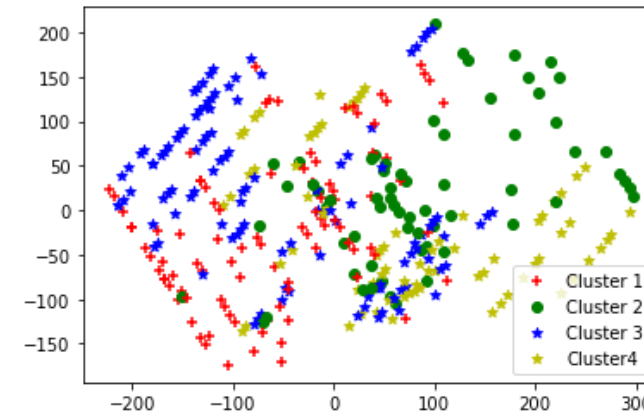
(1)以 頭距離牆壁的x向量距離 和 頭距離牆壁的y向量距離 為例：



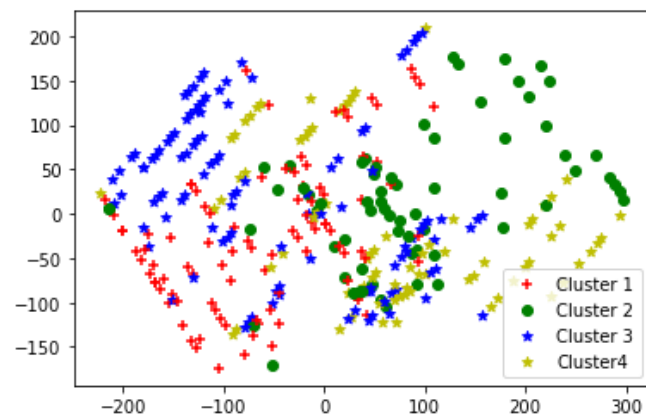
MLP



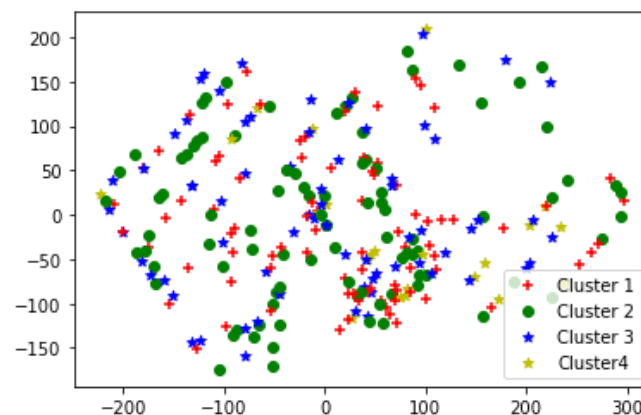
KNN



Extra Tree



Decision Tree

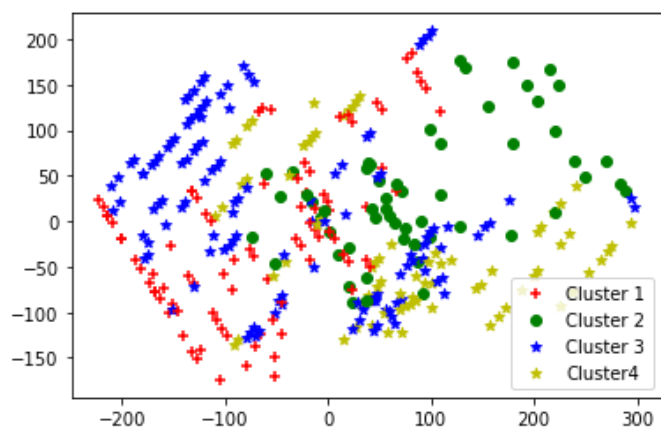


Actual Label

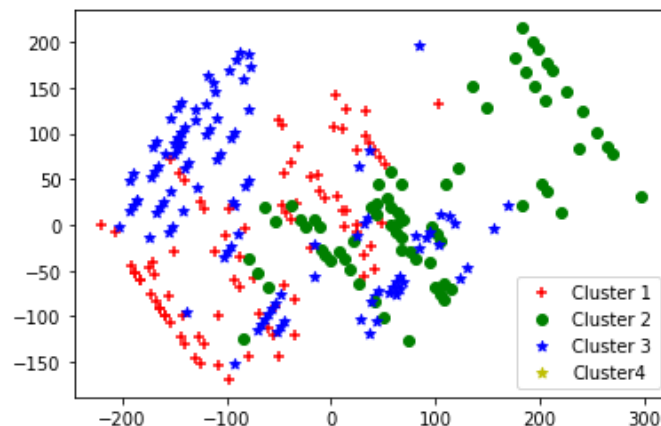
少取 頭距離牆壁的x向量距離 和 頭距離牆壁的y向量距離當 feature，各模型的accuracy皆下降了5%左右，圖形也與原本的圖形差非常大，因此頭距離牆壁的x向量距離 和 頭距離牆壁的y向量距離是極具有代表性的特徵。

問題二：所選的特徵數量多寡、是否具有代表性，也會影響到模型學習的成效。

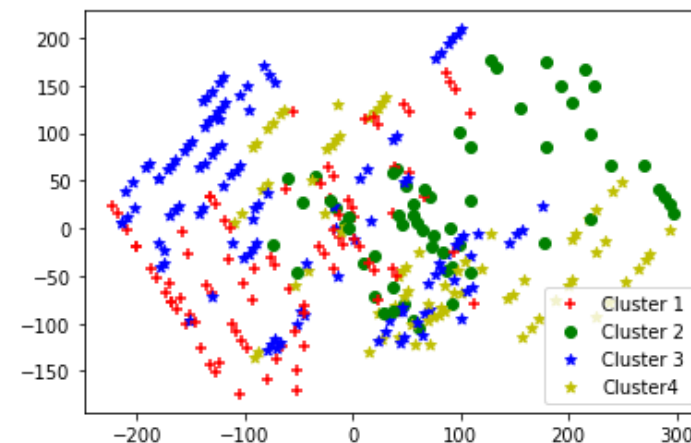
(2)以 body 為例：



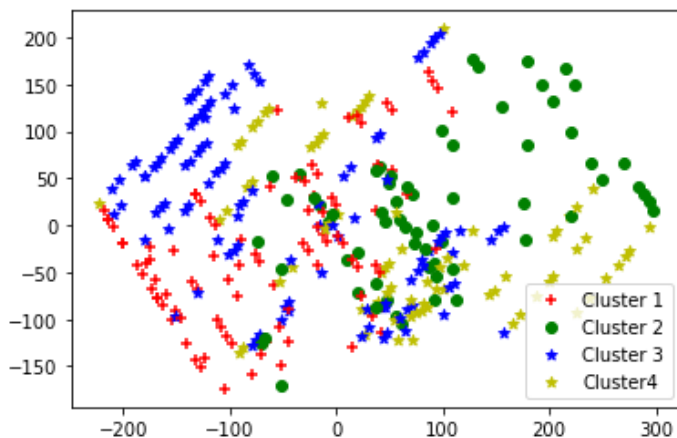
MLP



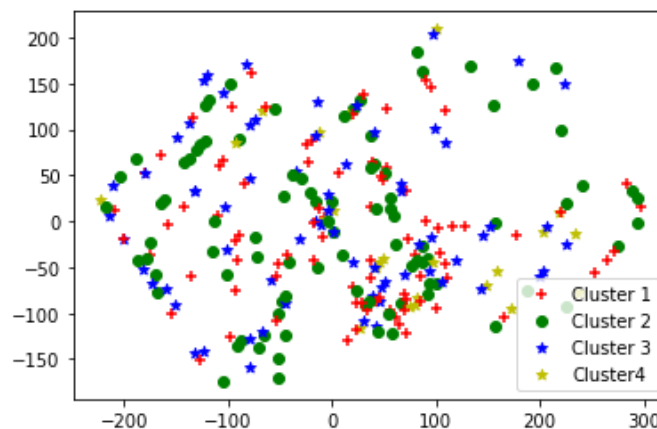
KNN



Extra Tree



Decision Tree



Actual Label

少取 body 當feature，  
各模型的accuracy皆下  
降了6%左右，圖形也  
與原本的圖形差非常大，  
因此 **body** 是極具有代  
表性的特徵。

### 問題三:使用不同機器學習的模型，分數的比較

MLP

Score: 68  
Score: 43  
Score: 35  
Score: 43  
Score: 37

KNN

Score: 61  
Score: 45  
Score: 47  
Score: 15  
Score: 14

Extra Tree

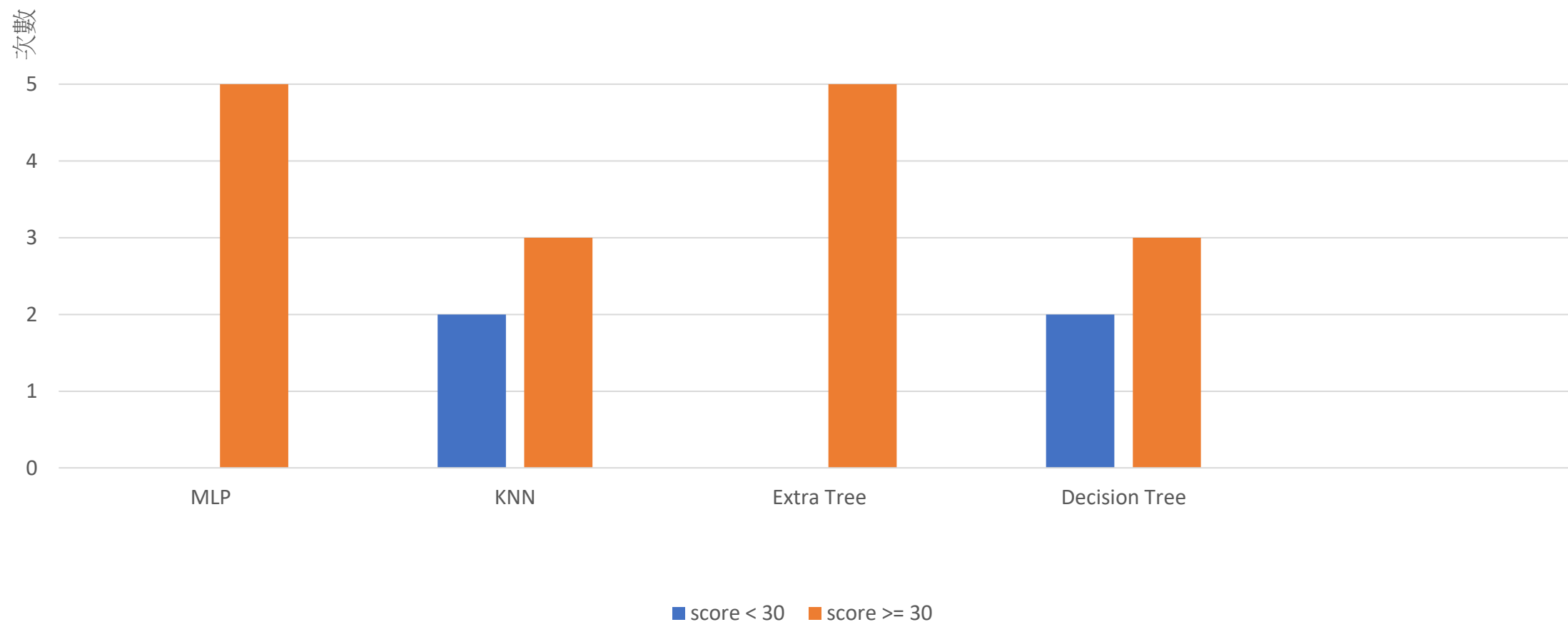
Score: 59  
Score: 49  
Score: 59  
Score: 36  
Score: 41

Decision Tree

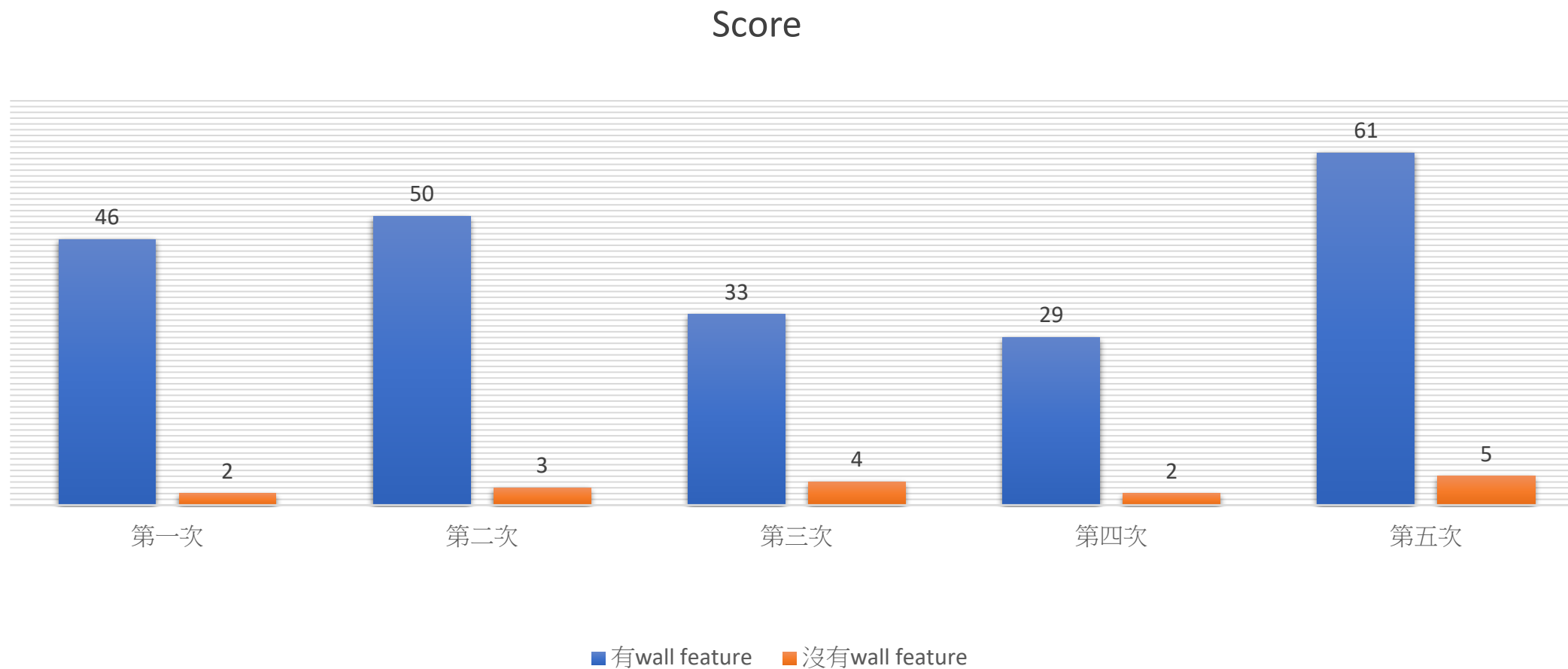
Score: 36  
Score: 65  
Score: 25  
Score: 30  
Score: 26

### 問題三：使用不同機器學習的模型，分數的比較(每個模型跑5次)

MLP vs. KNN vs. Extra Tree vs. Decision Tree

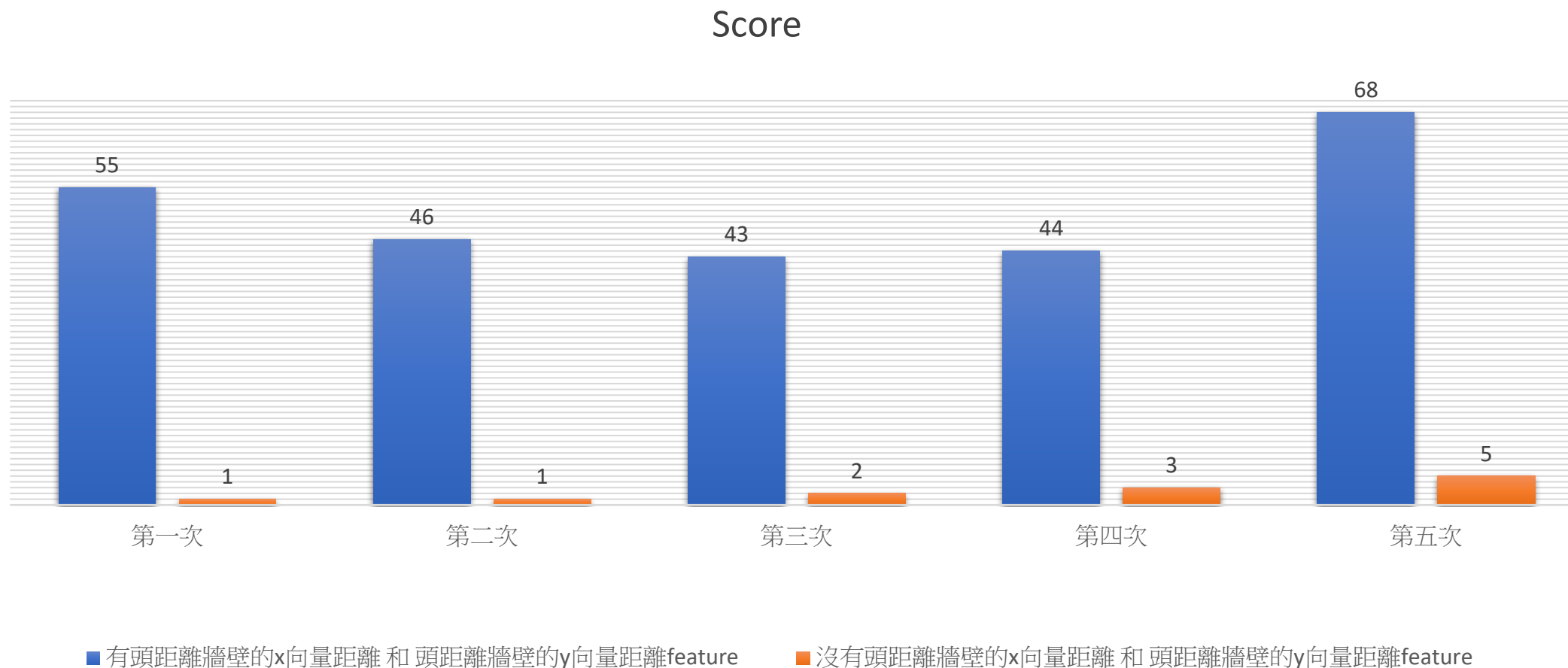


問題四:提取特徵多了 wall 與少了 wall ，分數比較  
以我選用的MLP演算法為例(五次):

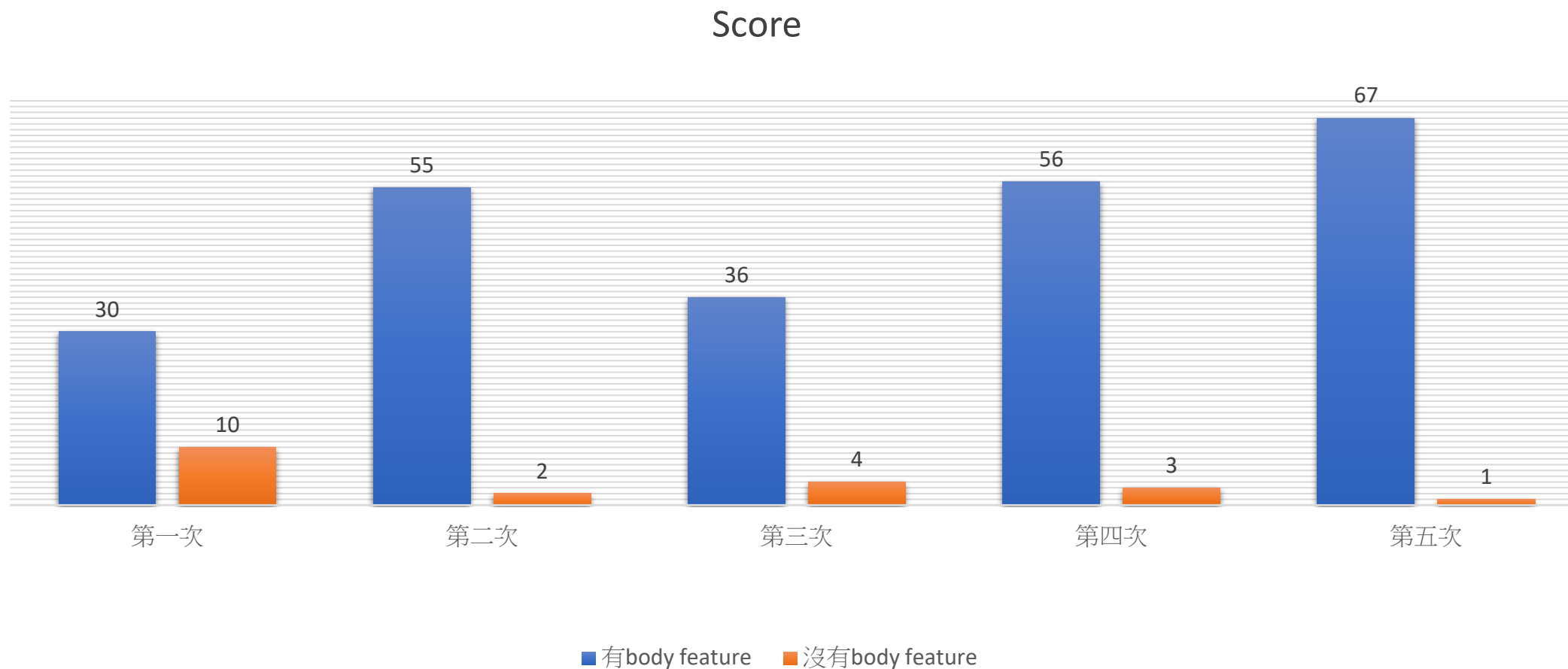




問題四:提取特徵多了 頭距離牆壁的x向量距離 和 頭距離牆壁的y向量距離  
與少了頭距離牆壁的x向量距離 和 頭距離牆壁的y向量距離 ，分數比較。  
以我選用的MLP演算法為例(五次):



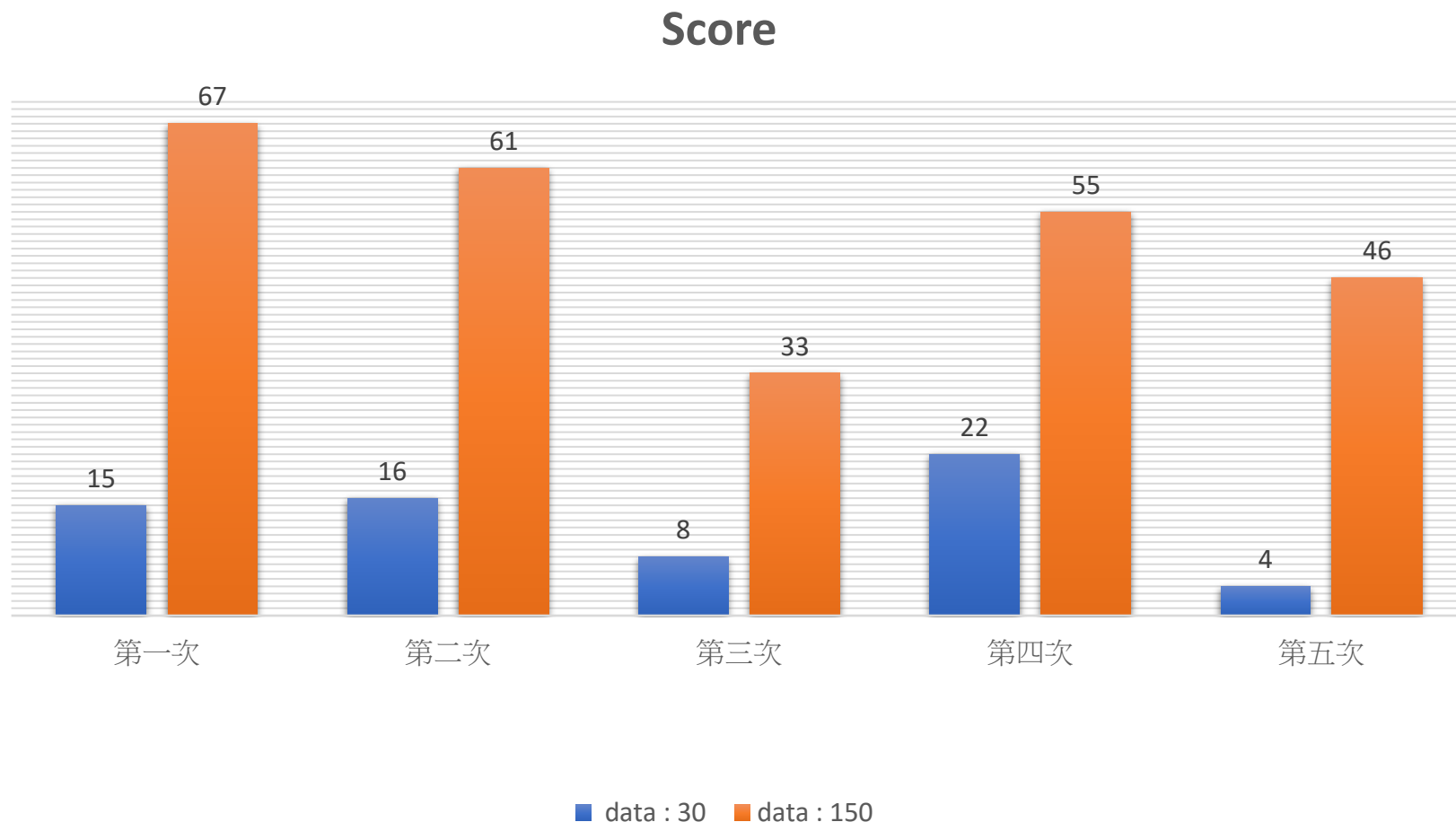
問題四:提取特徵多了 body 與少了 body ，分數比較  
以我選用的MLP演算法為例(五次):



問題五:data的多寡，是否影響訓練模型的成效?

30個pickle檔 vs. 150個pickle檔

以我選用的MLP演算法為例(五次):



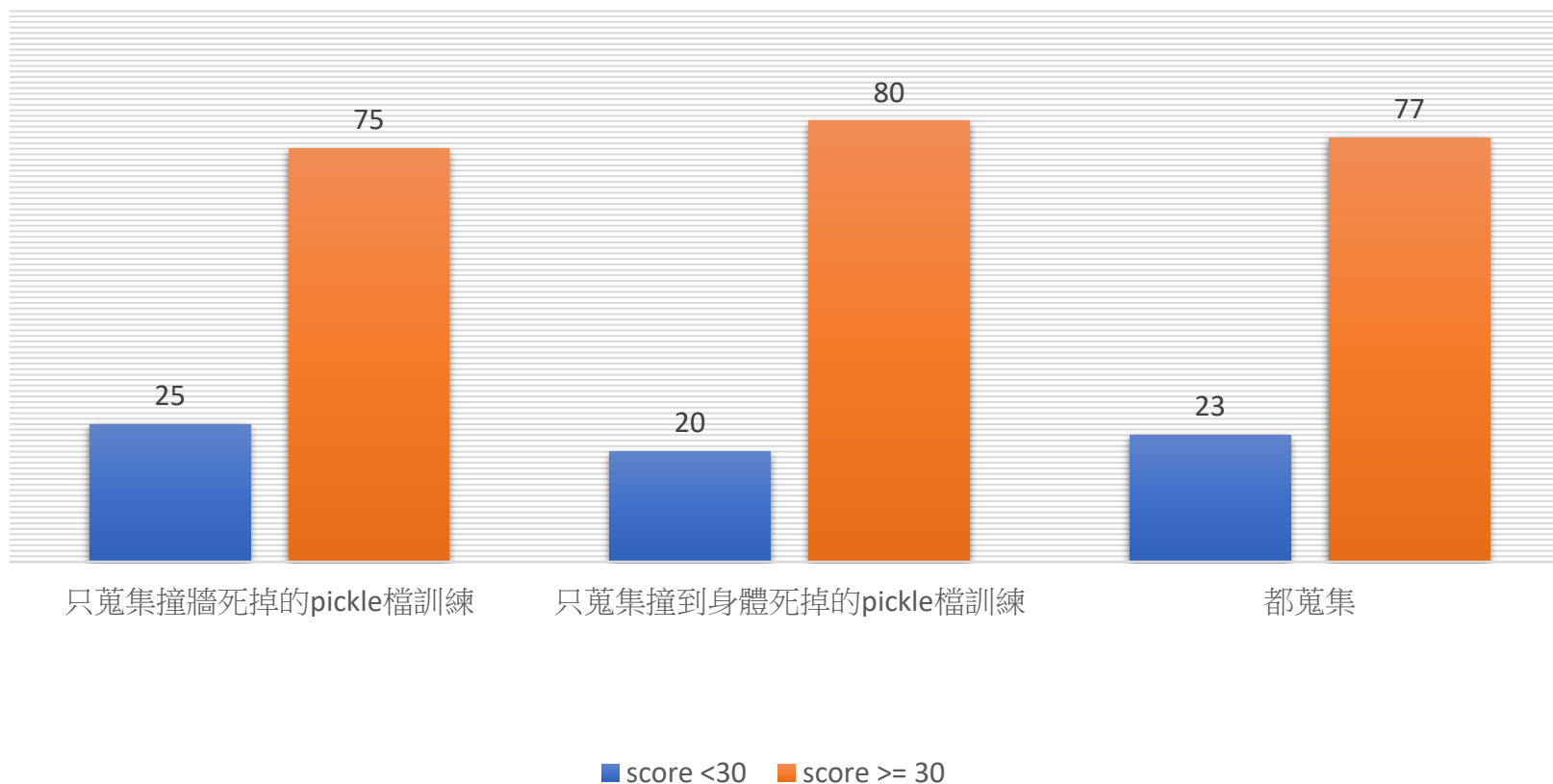
資料量越多，特徵也就蒐集的越多，因此score會比資料量少的高。

問題六: data多元性是否足夠，是否影響訓練模型的成效?

發想：如果都蒐集因為 撞牆死掉的pickle檔，會不會訓練出來的model分數  
會比都蒐集因為 撞到身體死掉的pickle檔高?

以我選用的MLP Classifier為例(跑100次):

次數



原本想說都蒐集不是因為 撞到身體死掉的pickle檔訓練，可能訓練出來的model就不會撞到身體，可以活比較久，結果其實跟只蒐集撞到身體和都蒐集的pickle檔跑出來結果差不多，因此 只蒐集撞牆而死的pickle檔並無法提升訓練模型的成效。

# CONCLUSION :

一看到貪食蛇遊戲竟然能用AI技術實現，覺得很不可思議，因為覺得光是用rule就不可能可以讓蛇遵循著rule去吃食物。一開始寫rule的時候，考慮完撞牆，我花了很多時間思考如何寫一個rule可以不撞到蛇的身體。當我終於寫出一個能破30的rule後，又花了很多時間到scikit-learn看有哪些演算法可以train。這次作業最大收穫就是我接觸並了解我從來沒碰過的一項技術---神經網路。神經網路是一種自適應系統，通俗地講就是具備學習功能，因此可以運用在貪食蛇遊戲，讓貪食蛇學會去吃食物。