**2019 ECT Project**

**第1- 4題**

** 從data\_new.csv 選取Crossing ~ GKReflexes欄位（共34個屬性）**

** 計算以上所有欄位的平均**

** 加標籤（大於平均：’ Above-average Players’, 小於平均：’ Below-average Players’）**

** 訓練模型，可針對所需的模型進行屬性挑選**

** 切分資料集（test\_size=0.33），並用測試集測試模型**

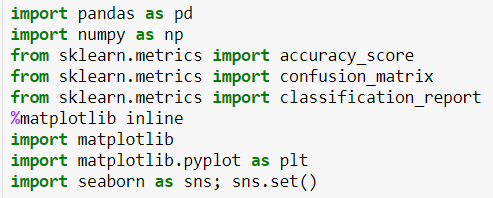
** 分析結果需印出accuracy 、classification report、confusionmatrix**

** 調整模型，讓accuracy達到0.9以上**

** 加分題（每大題至多2.5%）嘗試使用matplotlib等套件將各個演算法結果視覺化**

步驟過程

1. 首先，import所需套件。



1. 依各步驟進行實作。

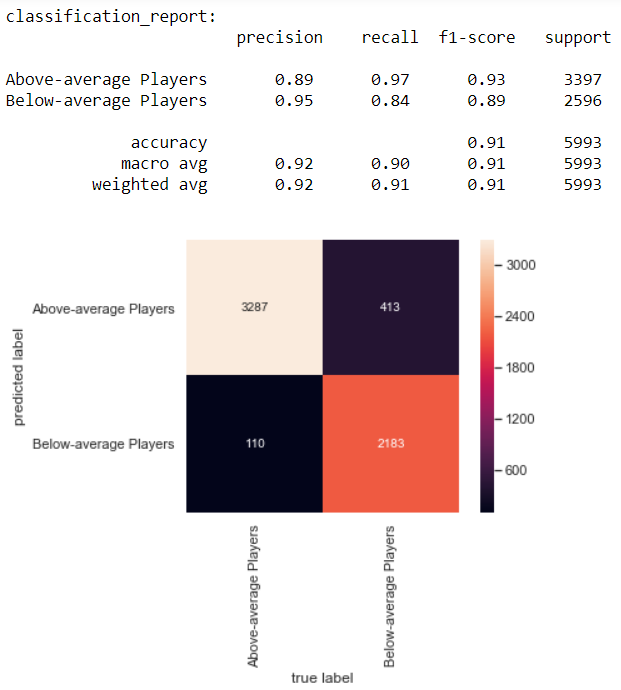


**1. Naive Bayes（20%）**

模型建立與訓練，並進行預測，透過 seaborn 顯示 confusion matrix

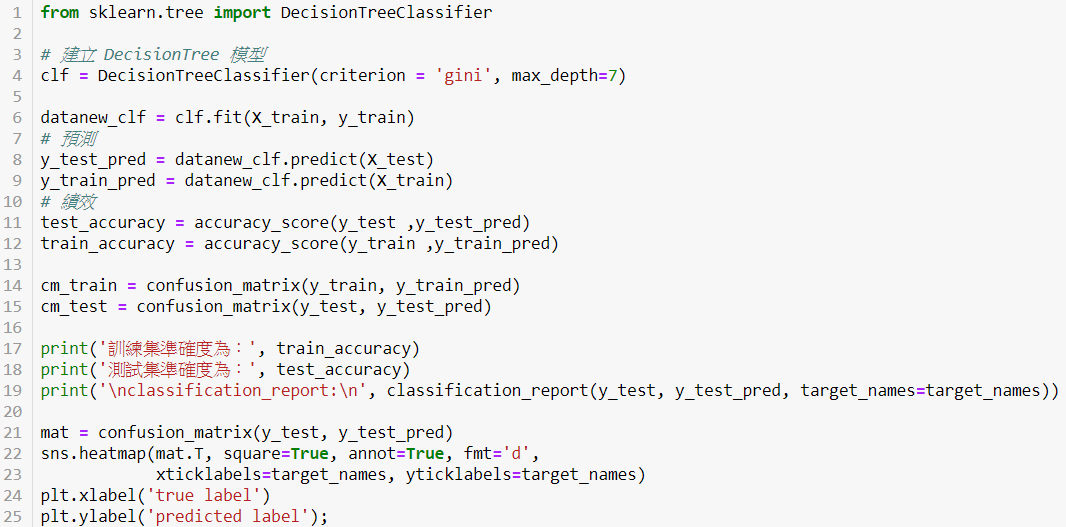


顯示結果如下

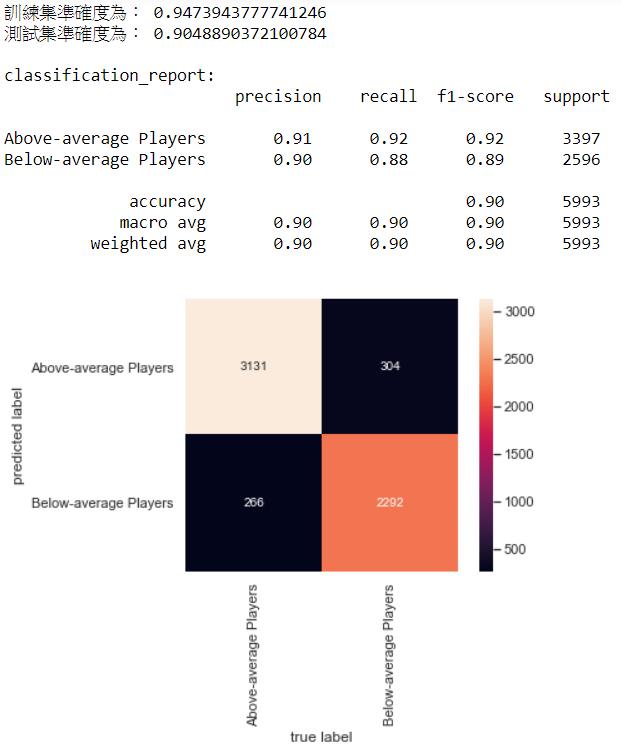


**2. Decision Trees（20%）**

模型建立與訓練，並進行預測，透過 seaborn 顯示 confusion matrix

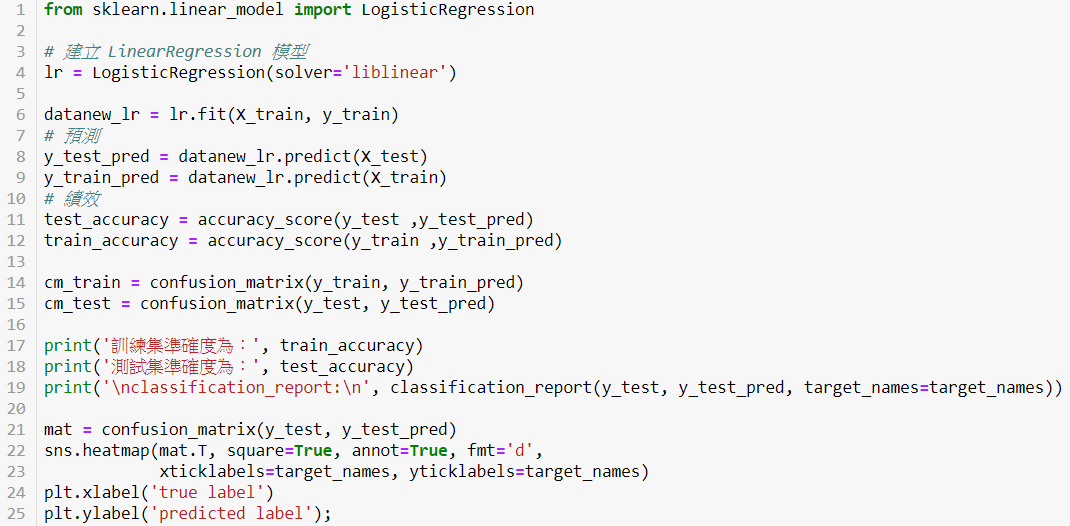


顯示結果如下

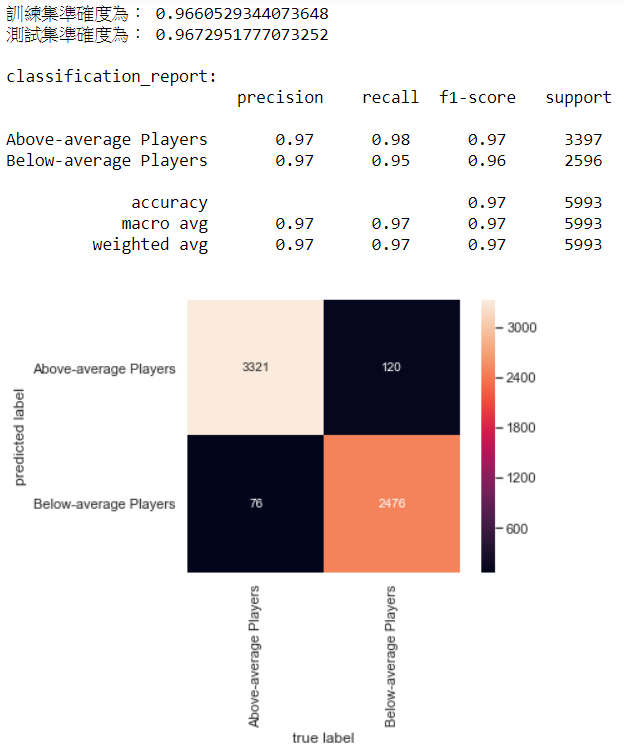


**3. Logistic Regression（20%）**

模型建立與訓練，並進行預測，透過 seaborn 顯示 confusion matrix

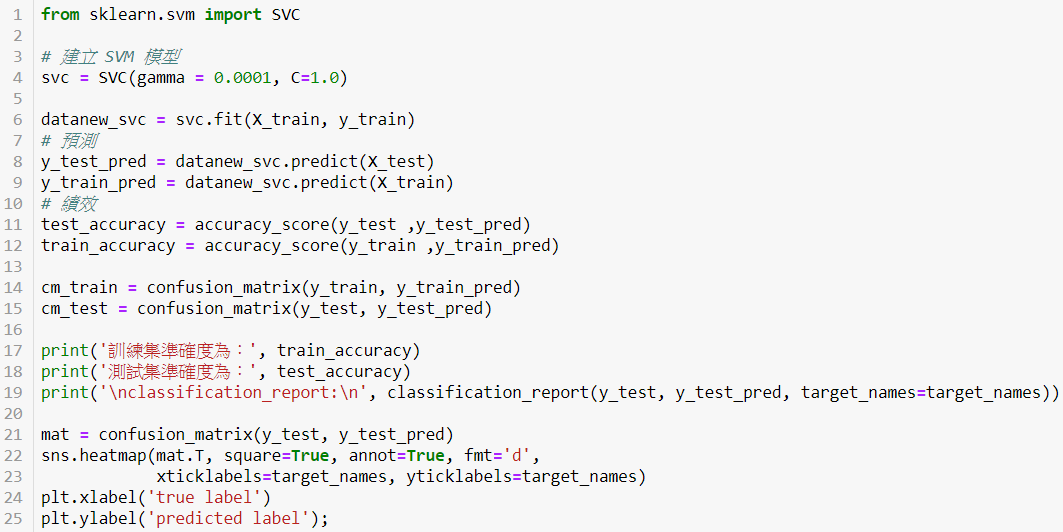


顯示結果如下

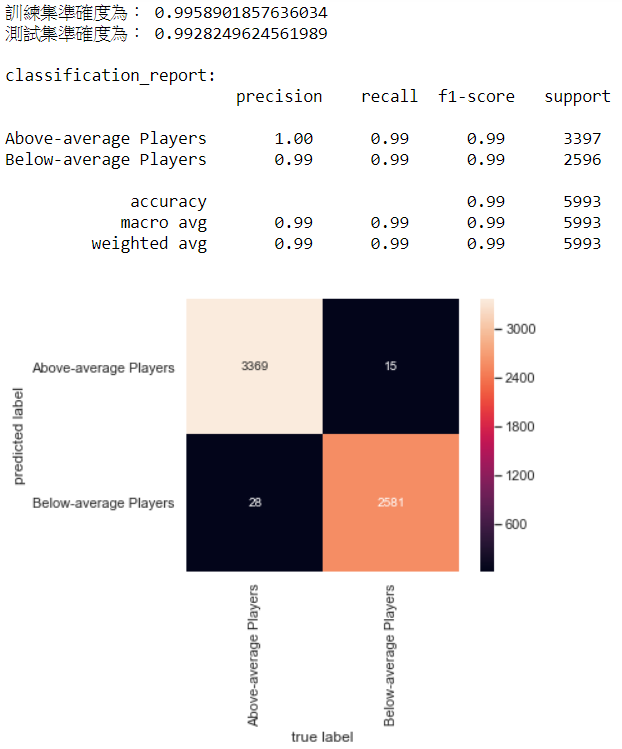


**4. SVM（20%）**

模型建立與訓練，並進行預測，透過 seaborn 顯示 confusion matrix



顯示結果如下



**第5題**

** 取data\_new.csv，進行KNN分析**

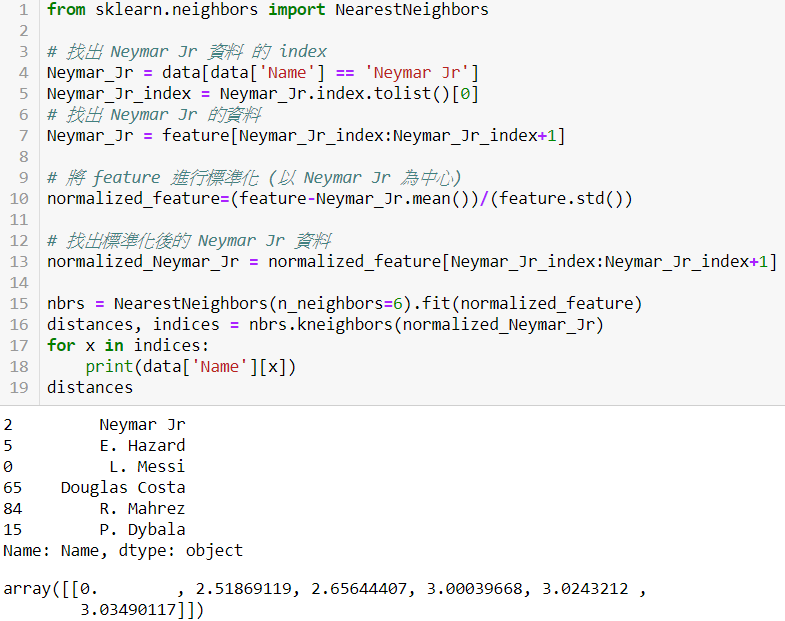
** 可針對所需的模型進行屬性挑選**

選取Crossing ~ GKReflexes欄位，並加上Skill Moves欄位。



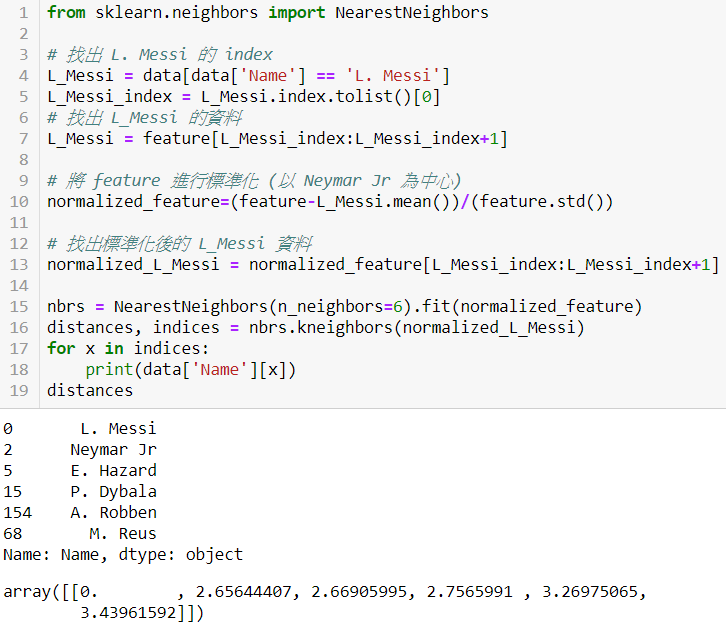
**5. KNN（20%）**

**(a) 推薦與 "Neymar Jr" 相像的前五名足球選手**



1. 先透過原始 data 找出 Neymar Jr 資料 的 index
2. 將feature以Neymar Jr為中心進行標準化，不同於一般標準化的方法，每筆feature中的資料會減去Neymar Jr的資料，再除以feature的標準差。
3. 利用步驟1. 找出的 index 找出 標準化後的feature中Neymar Jr的資料。(此資料將用於後續kneighbors演算法中)
4. 得出E. Hazard、L. Messi、Douglas Costa、R. Mahrez、P. Dybala為與Neymar Jr相像的前五名足球選手

**(b) 推薦與 " L. Messi " 相像的前五名足球選手**



1. 先透過原始 data 找出 L. Messi 資料 的 index
2. 將feature以L. Messi為中心進行標準化，不同於一般標準化的方法，每筆feature中的資料會減去L. Messi的資料，再除以feature的標準差。
3. 利用步驟1. 找出的 index 找出 標準化後的feature中L. Messi的資料。(此資料將用於後續kneighbors演算法中)
4. 得出Neymar Jr、E. Hazard、P. Dybala、A. Robben、M. Reus為與L. Messi相像的前五名足球選手

**第6題**

**6. 加分題（10%）**

** 對資料額外進行有趣的分析**

透過將attribute之間的相關性視覺化，可以更佳了解到屬性之間的相互關係，能做更多加深的應用

