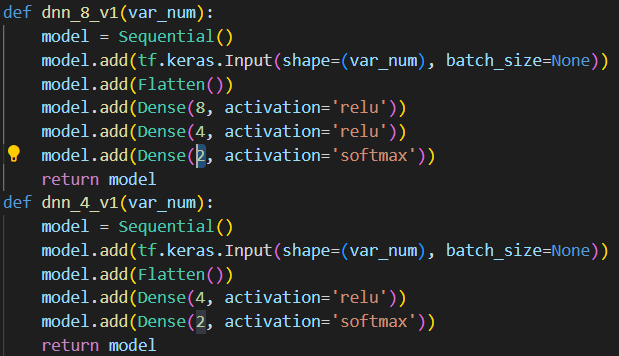
**AI目標**：給定一組變數(經度、緯度、一天的第幾小時、溫度、相對溼度、風速、降雨量、雷達迴波)，預測後一小時是否會降雨

**資料**：24個全台局屬測站(資料較穩定不易出錯)、雷達合成的dBZ強度，取用2021/1/1~2022/6/30資料，時間解析度一小時，共24(測站)\*24(24小時)\*546(天)= 314,496筆資料

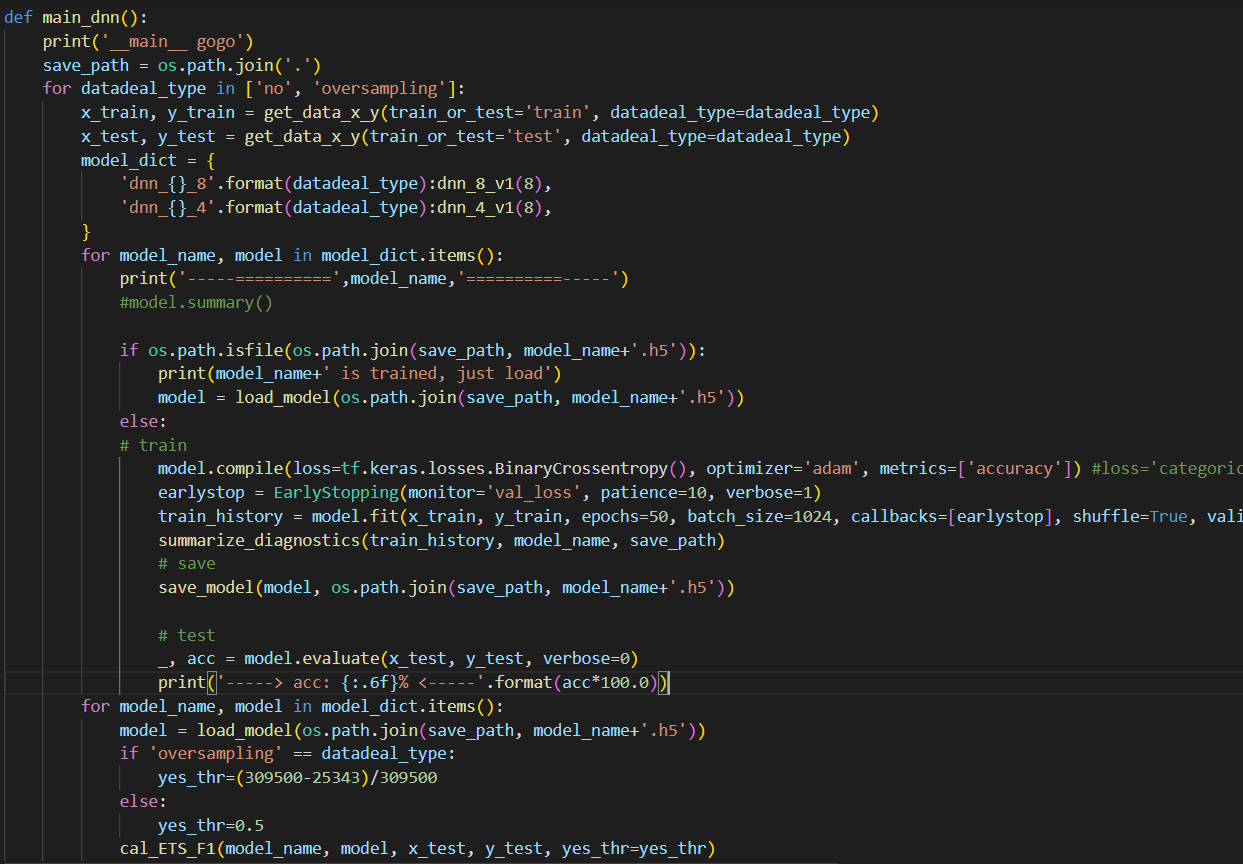
**資料處理**：

* 降雨量-9998代表雨跡，以降雨量0.1mm代替；回波上dbZ<0.0dB就=0.0
* 測站觀測若有-9991 -9996 -9997 -9999等觀測問題，以及雷達若有nan無觀測問題，就去除該筆資料
* 分別將各項變數根據其最大最小值縮放到0~1之間
* Label則是一小時後降雨量>0即視為有降雨，反之則無
* 以上原始Label的”有/無”為25343/284157，因此另外準備一個oversampling資料集，將有降雨的資料作複製擴增，最終預測時提高輸出有降雨的閾值從0.5->(無降雨樣本數/全樣本數)

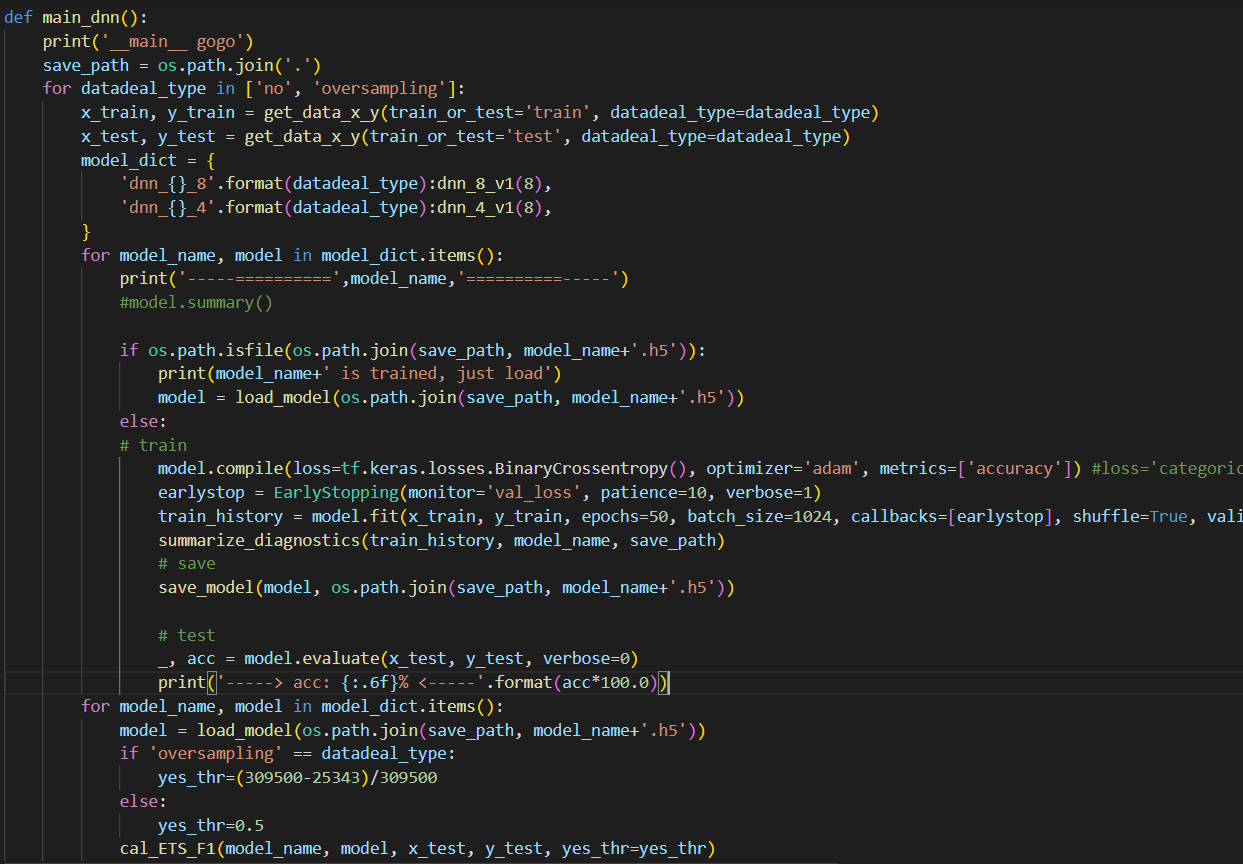
**AI架構**：DNN，堆疊Dense全連階層，以變數數量為準每層減半神經元數量(變數7，有7-8-4-2跟7-4-2兩種架構)，中間非線性函數採用relu，最終神經元剩兩個後softmax並輸出



**訓練**：如圖，對所有資料處理法跟架構作訓練，資料訓練集為總資料量的0.7



**測試**：如圖，對所有資料處理法跟架構作測試，資料訓練集為總資料量的0.3



結果上採用F1 score跟ETS公平預兆得分，分別取得0.8與0.48的成績，具有相當的預測能力

**應用預測**：在實際應用上，使用者輸入後經由後端運算，結合觀測與預報取得最貼近使用者的一組天氣情況，並且將這些天氣變數做剛剛資料處裡的最大最小值縮放(訓練的最大最小值有被記錄)，並將其送入模型得到預測

**SHAP分析**：

除了一般的訓練、驗證跟實際應用預測外，另外還根據合作賽局理論提出的Shapley Values (Lloyd Stowell Shapley)進行SHAP(SHapley Additive exPlanations)分析，得到以下變數重要性分布：

