作業五

404410039 李維哲

一、方法與流程

主要方法:

使用RandomForest

流程:

1.同時讀入.csv資料庫(使用pandas)

同時讀入的目的是為了等下encode時統一一種格式

把train資料的click數切開，分成了fullData跟fullTarget

2.同時對train跟test作labelencoder

目的是因為RandomForest的工具只支援int->需要encode

3.建立RandomForest的classfier，並利用train data作fit

(使用sklearn工具)

為了檢查正確度，有從fullData中切出testdata(來自traindata的)，以此來計算在trainData中的準確度

4.利用fit出來的RandomForest預測test data的結果(click)

5.將結果輸出成.csv檔(使用pandas)

二、組員評分

林佑荃:2分

人有出現，但沒貢獻，稍微有說話，不過都是說自己寫不出來

林政賢:9分

有嘗試過很多classfier的參數設定，還蠻投入的

李淼晶:1分

人有出現，但沒貢獻

廖翊凱:8分

有幫助隊員debug，還蠻投入的