

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 classifier，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 classifier 的參數設定，還蠻投入的

李焱晶: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