

## 實務演講

## Practitioner Speech

日期 / Date: May 24, 2019  
時間 / Time: 14:10-15:40  
地點 / Venue: 2F, Room 203  
主題 / Topic: 機器學習  
Machine Learning  
語言 / Language : 中文 / Chinese

## Speaker:

趙育祥 經理 TEJ 台灣經濟新報 風險管理部

## Topics:

Applying classifier with TEJ API to optimize workflow of event study : a monthly sales announcement case.

以 TEJ API 串接巨量資料搭建分類器優化事件研究流程：以月營收資料公告為例在進行事件研究時，訓練期（即事件日發生前之報酬率樣本）若與其他事件的發生日重疊，亦導致樣本受到干擾，而得出不穩定的檢定結果。為了處理此種問題，以刪除此樣本的方式處理雖然直觀，但將會刪除大量可用的樣本，亦有可能造成進行檢定的樣本有所偏誤，且難以事先確定何類與訓練期樣本重疊的事件，屬於可能干擾樣本的事件，特別是每年以較高頻率發生之事件，更亦與各種事件重疊，進而導致超過半數樣本遭到刪除的問題。另一方面，高頻率發生的事件，將使訓練集報酬率樣本數量膨脹到數千萬筆以上，產生資料前處理的困難，TEJ 透過 API 搭配 python 語言套件進行缺失值補齊與樣本標準化的資料前處理工作。

以月營收公告事件為例，因月營收事件每月發布一次，幾乎不可避免有半數訓練期樣本日期會與季財報發布、股東與董事會召開、宣告配息率等事件重疊，進而導致樣本受到干擾。為了保留更多樣本數，TEJ 嘗試以不同的分類器 (classifier)，將訓練期的報酬率樣本透過此樣本期間的公司財務基本資料、TCRI 等級等財務指標作為變數，輔以是否經歷其他事件作為虛擬變數，以機器學習的方式將樣本進行分類，按照分類結果分組進行事件研究，以更有效率的方式挑選進行事件研究的報酬率樣本，將分類的各組別報酬率資料分別進行 T 檢定與無母數檢定。

為了檢測分類器的效度，將所有樣本隨機切分為測試資料集與訓練資料集，將訓練集樣本分類後不同分組的 t 檢定值所對應的作為分類器的優化目標，以試圖切分出訓練期分屬不同報酬率分配的分類方式，並以測試資料集的 t 檢定值檢測效度。最後根據切分的結果，分別進行月營收公告事件研究，檢測各研究之 t 檢定與無母數檢定之結果。