

大數據資料科學家競賽參賽心得

◎文 / 公共衛生研究所 廖翊涵、施佳吟、黃湘閔、黃詠柔



「大數據」議題在近年來成為備受矚目的新機會，透過巨量資料的分析能夠有效率地找出富含價值的資訊，並在市場中奪得競爭優勢。SAS 公司與玉山銀行從 2012 年開始舉辦校園資料採礦競賽，希望透過競賽打造良好的產學合作，培育更多大數據分析與資料科學家的明日之星。SAS 一直是公衛領域常用於資料分析的統計軟體，然而我們對於 SAS 的認識卻一直侷限於醫院資料或是

健保資料的應用上，這次會想要參加比賽主要是希望能夠對於 SAS 在商業領域的運用有更進一步的瞭解，且比賽過程中將提供兩堂有關於 SAS EG 以及 SAS EM 的課程，能在比賽之餘學習到求學過程中較少接觸到的程式，這麼好的機會當然得把握啊！

本次競賽的主題為「FinTech 預測未來挑戰賽」，玉山銀行提供過去一年行內真實的 47 萬筆交易資料，希望



參賽同學能夠利用資料採礦的技巧預測 2016 年第二季最有可能購買外幣的 3 萬名顧客，並且針對其餘未成交的顧客進行特性分析，綜合顧客分群結果以及市場分析後提出完整的數位行銷企劃案。賽制包含初賽、複賽以及決賽三個階段，初賽是個人 EG 能力檢測，比的是大家對於 SAS EG 的操作熟悉度，雖然說是個人能力檢測，卻以小組平均分數作為評量成績，僅有小組平均分數前 40 名的隊伍能夠進入複賽。考試前會先幫大家上一堂 EG 操作的課程，課程結束後到考試前有將近一個月的時間讓同學練習，在這段期間內我們會到比賽

相關的臉書粉絲團下載練習題或是利用 SAS 官方網站上的教學影片複習 EG 的操作。初賽的部分完全就是有練習有分數，怕就怕在有豬隊友的存在，為了避免這種情況的發生，我們每週都會約好一起練習的時間，彼此互相監督、互相學習。

複賽的階段則是此次比賽的關鍵，在上完 EM 的課程後參賽隊伍會拿到銀行顧客的資料，而我們必須在三周內完成分析並繳交完整的行銷企劃案，最後將會有 12 組隊伍進入決賽。在這個過程中我們其實卡關卡了很久，一來因為對於銀行的資料不熟悉，光是瞭解變項



內容就花了將近一星期的時間，再來因為這是我們第一次接觸 data mining 的部分，對於資料探勘一點概念都沒有，在執行的過程中僅能按照課程中所教的步驟操作，一開始所跑出來的模型準確率都不盡理想，這時多虧有亮懿老師的協助，協助我們釐清對於不同資料時間點的處理方式，並提供我們一本參考書〈資料採礦應用 - 以 SAS Enterprise Miner 為工具〉，讓我們對於 data mining 能夠有更深入的認識。在行銷企劃案方面，對我們而言是優勢也是劣勢，非相關科系的我們能夠不被既往的觀念給綁住，提出比其他隊伍更具有創

意的行銷方案，而可行性的部分就另當別論了。

決賽的部分是要將我們的企劃書內容以口頭報告的方式呈現，要在 20 分鐘內說明如何建立預測模型以及介紹行銷方案並不簡單，投影片的編排及內容也需要有邏輯的呈現，需要告訴評審我們利用了哪些關鍵變項預測顧客對於外幣的購買機率，接著說明在未購買外幣的顧客中，我們挑選了哪一個分群的客戶進行行銷方案的撰寫，而選擇的他們的理由又是什麼。而在報告完畢後，SAS 公司與玉山銀行的主管們也會針對資料分析的過程以及行銷計畫案的部分進行提問，在問題答辯的過程中評審可以充分了解我們對於企劃案的熟悉度以及台風の穩健度。這是我們第一次踏出學術界到企業界做報告，整個比賽的過程對我們而言都是很特別的學習經驗，能夠與玉山銀行以及 SAS 資料分析師面對面交流更是難得的機會。

為期三個月的競賽過程對我們幾個研究生來說負荷實在有點大，因為白天需要完成自己的論文進度，與隊員的討論時間幾乎都在深夜進行，且此次競賽主題與我們所學的領域並不相同，對於資料的不熟悉也讓我們在分析的過程



中經歷了好長一段的撞牆期，謝謝隊員們的堅持與決心，也謝謝亮懿老師給予分析上的建議。透過這個比賽我們學習到 SAS Enterprise Guide 以及 SAS Enterprise Miner 的操作，同時也累積了跨領域實務統計分析的經驗，真的收穫滿滿，而對於不是金融相關背景的我們而言，能夠從 150 組的參賽者中脫穎而出獲得佳作的成績，真的很開心。