

# 資料探勘講座——心得報告

407402382 陳嫻淇

為拓寬學生的知識面，加強學術與實務知識的結合，呂奇傑老師利用星期三下午的選修課，幫我們邀請到大數據中心的研究員張廷任老師，給我們發表以資料探勘為主題的學術講座。以基礎理論為基底，輔以多年的從業經驗，張廷任老師向我們分享他身為輔大歷史系系友“斜杠”數據分析師的求職經驗，熱情又詼諧的向我們講解資料探勘的基礎理論。長達三小時的講座乾貨滿滿，令人受益匪淺。因此，本心得作為講座回顧，希望能總結一下自己的收穫和體會，以便更好的指導以後在資料探勘領域的學習。

從最基本的問題「何為資料探勘？」為出發點，老師向我們說明了資料探勘的跨領域性質，以及應用領域的大致分類。資料探勘是一門結合機器學習、分類學、管理科學、資訊系統、資料庫管理、特徵分析等多學科的復合性學科，其應用領域涵蓋自然語言處理、圖像識別、電信服務等。

老師接著講解資料探勘的過程，從數據庫、清洗數據（以得到有價值的數據）、資料倉儲、進行資料探勘、特徵分析，到最後的獲取知識以輔助決策，老師在每個環節都仔細講解該環節的重要性，最後以實例讓我們了解各環節的執行細節。老師多次強調兩個重點，一是建立數據庫，二是資料探勘的目的在於輔助制定決策。

在建立數據庫時，必須依據需求選擇資料庫類別，並且在最初設計架構時就必須明白資料庫「牽一髮而動全身」的特性。針對資料庫架構，老師向我們推薦了 Hadoop，並大致的介紹了架構師誘人的年薪和工作內容。有鑒於我近半年來所接觸到的網頁架站、資料庫以及小專題，我對大型專案的複雜度望而生畏，也非常佩服經驗豐富的工程師能夠在執行專案前在腦袋中完成整個專案的藍圖，更羨慕的是它有別於基礎打工人的百萬薪資。

因此，對於金錢的渴望驅使我搜了一下架構師的工作內容和職位要求。心得大致可以總結為「資山有路勤為徑，學海無涯苦作舟」。雖然算是完全的門外漢，但我個人對架構師這份職位的理解是，具有一定工作經驗且優秀的開發人員、具有良好抽象思維、了解實務需求和執行邏輯、重視商業價值的這麼一份具挑戰性的工作。至於這條路的學習路徑，以及「什麼樣的架構是好架構？」這一困擾我許久的問題，我仍需繼續尋找答案。

回到講座內容，老師接著講解了資料正規化的方法。最近學習統計學也接觸了正規化的計算方式，我很幸運能夠同時學習基礎統計的計算以及應用統計上的邏輯。正規化是為了讓不同維度的數據能夠橫向對比，因此必須建立一個同樣的標準以供比較，常用的手段是將數字放大倍數，或者找到數據間的刻度。此外，老師以自己的研究為例，講解為何相同數據有分群和沒有分群，得出的分析結果會有精準度上的差異。一旦將數據利用固定刻度分群之後，分析結果會較沒有分群的準確。

此次演講讓我受益匪淺，讓我再次深刻領悟「學海無涯，回頭是岸（誤）」，我也得以藉此描繪往後三年的學習路徑，希望未來能以加強基礎學科知識為主，將架站基本功訓練透徹，在程式撰寫上不斷精進。