

資料科學與開放資料

高嘉宏教授

國立臺東大學應用數學系

摘要

在 2012 年的哈佛商業評論中，資料科學家被譽為是 21 世紀最為誘人的職業。對於企業來說，如何從海嘯般的資料中去蕪存菁、抽絲剝繭，藉此找到重要商業問題的答案，且做成決策，是資料科學家對於企業的魅力。而對於資料科學家來說，能夠了解各個利害關係人的領域問題、整合且應用數學、統計與資訊等跨領域知識及工具，協助識別出隱藏於資料背後的蛛絲馬跡，並產生價值與貢獻，則是資料科學家這個職缺的有趣之處與挑戰。眾多文章皆詳細列舉出成為資料科學家的學習主題，從基礎的程式設計、統計學、資料蒐集匯入、資料整理、資料視覺化，到進階的機器學習、巨量資料、文字探勘，以及自然語言處理等等。在眾多學習主題中，如何逐步建立起相關的知識與技術，並累積實作經驗呢？在此次的分享中，除了資料科學的簡介外，更將介紹國內豐富的開放資料與開放資料平台，以及數種常見的開放資料格式，並分享在資料蒐集匯入、資料整理，與資料視覺化等階段的經驗。希冀透過這一次的分享，同學即可針對自己有興趣的主題，跨出成為資料科學家的第一步。