**心得報告**

B10609047 李政翰

今天的課程中我對random forest等機器學習模型有了更多的認識，也對於機器學習目前的使用趨勢有所了解，誠如老師所說:「我們需要站在巨人的肩膀上」，如果所有錯誤都要自己嘗試，人類始終只會一直鑽木取火，無法堆疊。

另外，老師利用實際案例說明，讓我更加了解為什麼許多公司願意花大錢請資料科學家做數據分析。

以往要執行一項行銷研究，受限於技術，時間和成本都很高。更重要的是，在執行行銷研究前，還有統計抽樣的問題要克服。如何有效地接觸到能充分代表母體的樣本，是行銷研究的難題之一。而隨著行銷資料科學的出現，解決行銷問題的時間與成本有望降低。同時，蒐集資料的來源，雖然未必等於母體，但已經比行銷研究的範圍廣大許多。

透過爬蟲技術，從社群媒體、部落格、論壇、應用程式介面等，擷取消費者的上網行為。也可以透過網站監控軟體，記錄消費者瀏覽網站的路徑，還可透過影像辨識技術，搜集消費者在賣場中的真實消費過程（例如：在哪個時間、哪個位置、看了哪件衣服、看了多久…等），而這一部分還可延伸至「物聯網」的各種應用，或是透過次級資料來源，像是「政府資料開放平台」等，蒐集巨量的外部資料。

而在分析資料的過程中，過去「行銷研究」透過「統計分析」、「多變量分析」等工具，協助找到問題的解答。行銷資料科學則還可透過機器學習和文字探勘等工具，給予資料更多的解釋並創造更高的價值，例如：發展出準確度高的預測模型。

最後，感謝老師無私分享這些學習經驗以及程式碼，讓同學們可以有更多機會精進現在最夯的大數據科學。