

摘要

由於網際網路的演進與電子商務的崛起，全球網際網路的傳輸量已經從 TB(1024GB = 1TB)躍升到 PB(1024TB = 1PB)甚至是來到 EB(1024PB=1EB)的級別，在如此龐大的資料量流通的過程中，各式各樣的服務流程已逐漸邁向資訊化，顧客透過搜尋引擎(例如：Google)在網路上搜尋有興趣的產品資訊，然而現在有許多的媒體服務商會藉由顧客的授權並合理地收集顧客的資料，對於企業來說，若能精準地從由顧客行為所產生的數據中預測顧客所產生的價值並投其所好，相信對於獲利的掌控可有較多的彈性。過去的研究對於顧客價值的分析大部分都是使用統計相關(例如：Recency, Frequency, Monetary (RFM)模型)或是經驗管理法則(例如：平衡計分卡)的技術應用在顧客忠誠度、價值或是服務品質等。

本研究採用機器學習的開源數據分析工具，建立三個決策樹相關的模型：XGBoost、LightGBM 與隨機森林。根據顧客操作網站時所留下的行為與流量紀錄以及對應於網路商店的收入來預測顧客之間不同的顧客行為所產生的收入。除了建立模型訓練數據之外，並使用參數調整技術來優化模型訓練的成效。評估模型的方式則使用均方根誤差(RMSE)來檢驗三種模型之間的績效並針對演算風格差異進行討論，最後結果顯示 XGBoost 的預測結果較其他兩個模型誤差最小，其 RMSE 為 49.098 並以此結果作為預測模型之選擇並探討顧客行為的差異，且在實務方面來說則可提供企業在規劃行銷活動、產品規劃或是預算推估等參考。

關鍵字：顧客行為、顧客價值、機器學習、XGBoost、LightGBM