

應用資料科學技術建構以顧客銷售資料為基之機能性布料需求預測決策模式的實務研究

許譽騰、郭軒佑

摘要

台灣許多屬於代工生產模式的中小企業，是以訂單的歷史資料來建立需求預測模型。現階段瓶頸為無法準確瞭解消費者的實際需求，特別在原料儲備、產品庫存和產品研發方面無法有效提前部署。本研究參考C2M(Customer to- Manufactory)製造模式，由消費者角度所取得的外部資訊來建立機能性布料需求決策系統。研究方法為應用網路爬蟲技術從電商平台擷取使用者的消費和反饋數據，然後再比較不同機器學習演算法來建立最適預測模式；並且設計互動式資料儀表板提供決策。透過北部某染整廠進行實務成效的驗證，結果顯示本研究成果確實能夠強化原本公司的需求預測；另外公司高層一致認同，可用於公司未來整體的產能規劃與經營決策。就學術貢獻而言，本研究透過資訊技術與數據分析提出一套具邏輯性的程序。就實務貢獻而言，提供可決策的有效訊息來解決廠商面臨的需求預測問題。

關鍵字：C2M, Python, 染整廠, 機能衣, 網路爬蟲, 機器學習

1. 背景與目的

過去隨著 COVID-19 疫情爆發，在許多不確定因素下，帶動居家辦公與居家健身的風潮。民眾出門機會減少，待在家裡時間變長，牽動服飾業流行趨勢的改變。讓越來越多時尚設計師介入運動機能市場，許多知名運動品牌服飾，如 Nike、Adidas 與 Burberry，均以 Athleisure 為設計，標榜機能與有型的訴求。雖然價格不菲，但仍大受歡迎，甚至在這波疫情下賣到缺貨。目前機能性服飾的市場還是以都會區為主，主要是因為現代都會民眾生活忙碌，自清晨出門工作，常至深夜後才回到家，所以服裝設計除考量美觀外，亦朝向一衣多穿、容易保養以及提高整天活動的舒適度等方向著手。例如，沒有進行瑜珈活動或其他運動時，也會將瑜珈褲或是內搭褲當作日常穿著。機能性紡織品結合時尚及機能性的特性，以男性為例，機能性時尚服飾是機能性地添加防潑水、抗汗、透氣快乾、抑菌除臭、彈性舒適等功能，在時尚設計部分上，則是導入特製裁縫版型設計，除舒適更要美觀，也能在都會生活中

的各種場合裡穿著。根據消費者的需求調查結果，在挑選運動服時，消費者較注重性能，如排汗、透氣程度，以及彈性、拉伸度等，遠大於服飾品牌和造型。即使疫情過後，現代人的生活型態，讓廠商依舊看好運動服的市場。

台灣的紡織業已經是世界知名品牌最重要的供應鏈之一。台灣的紡織廠商生產的機能性布料廣為人知，吸引了 Nike、Adidas、Burberry 等全球品牌直接來台採購優質布料。布料研發需要較高的技術含量，大部分布料生產商的研發能量仍保留在台灣，所以這也是台灣紡織業能夠有機會搶灘智慧製造的發展機遇。台灣紡織業的一大競爭優勢是有能力開發各種功能性布料，如耐熱、涼爽、抗菌、透氣、防臭等。目前，台灣擁有全球市場 70%份額，許多全球品牌甚至跳過服裝業，直接來台灣下訂單購買高品質的布料。相較於下游的成衣業因勞工需求而多半將工廠遷出台灣。機能性布料的生產需要高技術能量，因此中游的織造業是台灣紡織業爭取智慧製造的主要發展軸。事實上，在中游的織造、

染整和下游的服裝生產方面，服裝行業都很關注庫存問題。因此，對於下游的服裝行業來說，若沒有一套有系統的架構來預測需求訂單，將會造成庫存浪費與資金壓力。

過去製造商對於需求的預測，主要是以訂單數據為主來建立模式。2012、2013 年先後有專家與業者提出 Consumer to Manufacturer (C2M)的概念並且付諸實踐，如 2018 年阿里巴巴提出「新製造」的概念，主要內涵同為 Consumer to Manufacturer；同年淘寶與拼多多皆進行相關布局措施，讓 C2M 加速發展，後續發展範圍更為廣闊，也讓 C2M 模式成為新興類型的商業模式。消費數據驅動精準研發製造模式，是顛覆先工廠生產再市場行銷的傳統供銷思維，先由用戶需求驅動生產製造，市場要什麼，工廠就生產什麼，市場要多少，工廠才生產多少。事實上是改良快時尚，盲目上新和低質低價的模式，將消費者需求對接到工廠端。根據消費者實時的需求以及對市場前沿趨勢的把控，進行小批量的快速反向生產，這樣的做法既能快速反應消費者需求，同時不容易導致庫存的累積。

因此，如果我們能夠掌握消費者的消費數據，從多個資料來源所收集的社群網站如臉書的網路輿情、以及網購平台的評論與意見回饋，經過數據分析後，就能夠解析整個機能性運動服飾市場的訊息，找出值得切入的機會，藉此降低進入市場的門檻。同時切合消費者的需求，就可以獲得比傳統的訂單式需求預測得到更多信息，從而提高訂單需求預測的準確性。本研究目的是以消費者的角度來探討需求，並發展一套以數據科學為基的需求預測模式，並且撰寫 Python 以及 Tableau 來實現顧客消費數據的分析。研究成果將可以用來協助台灣中小企業染整廠制定精準生產計畫的決策。

2. 文獻探討

紡織產業面臨服飾品牌快時尚趨勢衝擊，整體紡織供應鏈與全球品牌通路都在推動零庫存、短交

期與少量客製化，生產時間、品質。面對品牌商對 ODM(Original Design Manufacturer)的預測與實際需求常有落差，造成物料管理與大量庫存成本積壓的問題。由於客戶預測需求不準確，常導致備料困難，備料太多會增加積存量、備料太少可能延誤交期。需求計劃取決於能夠開發的操作系統，如 ARIMA 以及與生產計劃相關的工程工具和預測技術[1, 2]。預測技術對紡織業有非常積極的影響，為了選擇合適的預測技術進行分析，必須確定其均方根誤差，以確保該技術給出的銷售結果與實際銷售數據接近[3, 4]。Ku *et al.*, 提出工業 3.5 的架構來開發染色機調度的決策支持系統，以增強智能製造並打破信息孤島。Lorente-Leyva *et al.*, (2019) 以 T 恤系列執行聚合生產計劃為例，提出由裝機容量和工人數量來決定分析制約因素，以及仿生優化技術。結果發現可提供的服務水平提高 96% [5]。

就需求預測而言，Lu (2014)整合 MARS 與 SVR 提出一個兩階段資訊產品銷售預測模式，並成功用於筆記型電腦、液晶螢幕、主機板、硬碟及顯示卡的銷售預測上[6]。Lu *et al.* (2012) 比較 MARS 與多種 ANN 方法於資訊產品批發商的需求預測結果，結果顯示 MARS 能夠建構較有效的資訊批發商銷售預測模式，並找出重要的預測變數[7]。Lu and Wang (2010) 結合 ICA、GHSOM 與支援向量迴歸建構整合式電腦代理商銷售預測模式。實證結果驗證所提的整合式預測模式相較於三種比較的預測技術，有較佳的預測績效[8]。就間斷型的需求預測來說，Prestwich *et al.*(2014)利用 Croston 方法與貝式推論(Bayesian inference)提出雙曲線指數平滑(Hyperbolic-Exponential Smoothing)於處理間斷性需求預測時存在報廢問題。結果顯示，此模式在非間斷性和隨機間斷性需求上與傳統方法無偏差，但當存在報廢問題，所提方法的預測表現良好[9]。Lolli *et al.* (2017)應用與比較倒傳遞神經網路和極限學習機於 24 個汽車產品的間斷性需求預測[10]。Petropoulos *et al.* (2013) 提出啟發式的 Theta 模型作為 SES 估計量的一個替代方法，用於推斷間斷性需求的非零需求量和時間間隔。此研究首先探索

Croston 方法中平滑參數的最佳化並考慮預測過程中的整數舍去影響，進而透過在汽車業的實際間斷需求資料進行實證。結果顯示所提模式的預測表現比一般 Croston 方法有所提高[11]。

由於消費型態改變，電商平台的銷售數據已成為賣家和廠商重要的決策依據。因此，本研究是以網站上機能衣的消費者資料為核心，利用網路爬蟲技術獲取電商平台資料，以及機器學習技術建立預測模型。最後，本研究也將應用警示燈號概念於資訊呈現，以提高預測結果的可解釋性。

3. 研究方法

本研究提出基於電商平台客戶消費數據的預測架構，來分析消費者購買機能性衣物趨勢。研究架構包含網路爬蟲、數據標準化、機器學習、資料調整與視覺化等四個步驟，如圖 1 所示。

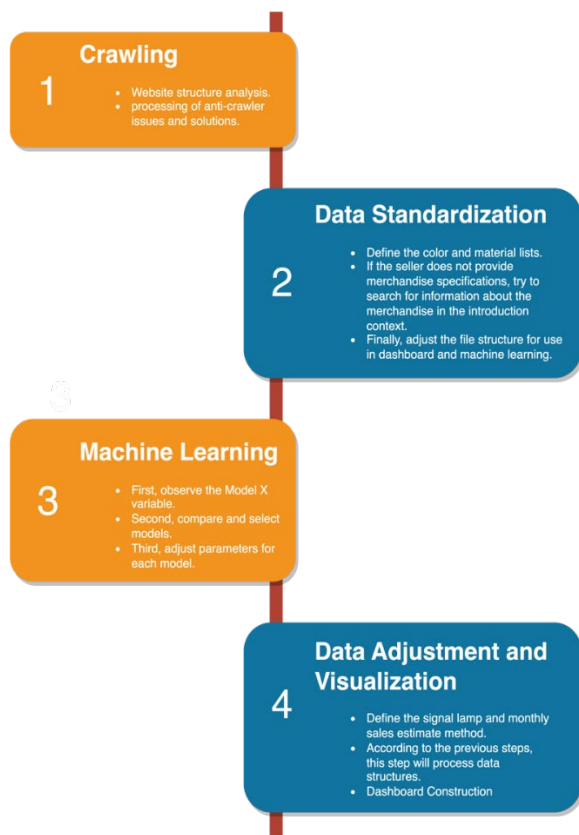


圖 1 研究流程架構

3.1 網路爬蟲

3.1.1 定義關鍵字

當在 Shopee 購物網站搜尋「機能衣」時，可能會受到搜索引擎的限制，因此顯示的數據量會有所不足。Shopee 搜索的限制是每頁顯示 60 個產品，最多顯示 50 頁。此外，也可能會看到一些不相關的產品，例如「舒壓抓抓熱」，如圖 2 所示。

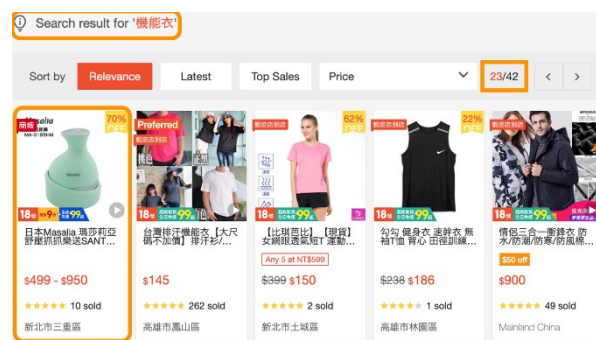


圖 2 機能衣搜尋結果

為避免出現這些問題，本研究根據機能類別作為關鍵字，分別為排汗衣、涼感衣、發熱衣和防風防水外套，以獲取完整的商品資料。

3.1.2 網站架構分析

本研究使用 Chrome DevTools 觀察 Shopee 購物網站的資料載入方式。該網站透過 AJAX 技術 (Asynchronous and XML) 進行網頁載入，透過 JavaScript 與 Server 交換資料。與傳統 HTTP 請求不同，如圖 3 所示，這種技術不僅可以更有效率地提高網站頁面載入速度，也意味著爬蟲程式需要模擬使用者操作，才能載入完整的資料。

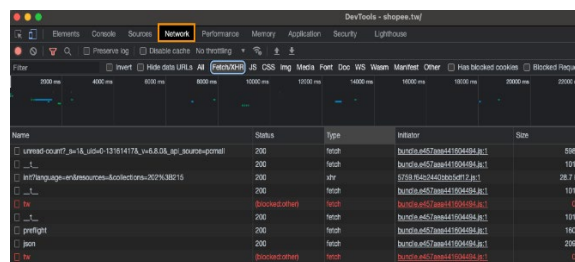


圖 3 Google 瀏覽器開發者工具-網路頁面

使用 Chrome DevTools 找到所需的 API 網址後，可以將 API 分為商品資料、留言資料和搜尋項目 API(請參見使用者可透過 offset 參數拆分請求以取得更多資料，但 offset 上限為 3000 筆。

表 1 需要注意的是，商品資料和搜尋項目 API 是加密的。若未提供正確的標頭，將無法存取。留言資料 API 每次最多回傳 59 筆資料，若超過則無法取得。使用者可透過 offset 參數拆分請求以取得更多資料，但 offset 上限為 3000 筆。

表 1 Shopee API

API	URL
商品資料	https://shopee.tw/api/v4/item/get?itemid={Product ID}&shopid={Store ID}
留言資料	https://shopee.tw/api/v2/item/get_ratings?filter=0&flag=1&itemid={ProductID}&limit={Number of loads}&offset={offsetvalue}&shopid={Store ID}&type=0
搜尋項目	https://shopee.tw/api/v4/search/search_items?by=relevancy&keyword={searchkeywords}&limit=60&newest=0&order=desc

3.1.3 爬蟲程式架構

此研究的爬蟲程式分別為：商品爬取、留言爬取和商品細節資料爬取。商品爬取的程式執行框架，如圖 4 所示，該程序用於從 Shopee 網站上爬取商品信息，如商品名稱、商品 ID 等。

圖 5 為商品留言爬蟲程式的架構。在載入留言資料的過程中，我們攔截請求時的 header 參數，如圖 6 所示。這些 header 參數中有四個，其中一個是加密過的"IF-None-Match"參數，而其餘的參數則是固定的。透過開發者工具分析網頁原始碼發現，"IF-None-Match"參數是使用 MD5 加密技術，因此在本研究中，我們使用 hashlib 套件對加密數值進行解碼，以獲取"IF-None-Match"參數。

商品細節的資料爬取程序執行框架，如圖 7，使用 Selenium 模擬使用者瀏覽網頁，並繞過 Shopee API 加密機制，抓取商品資料 API 封包內容[12]。

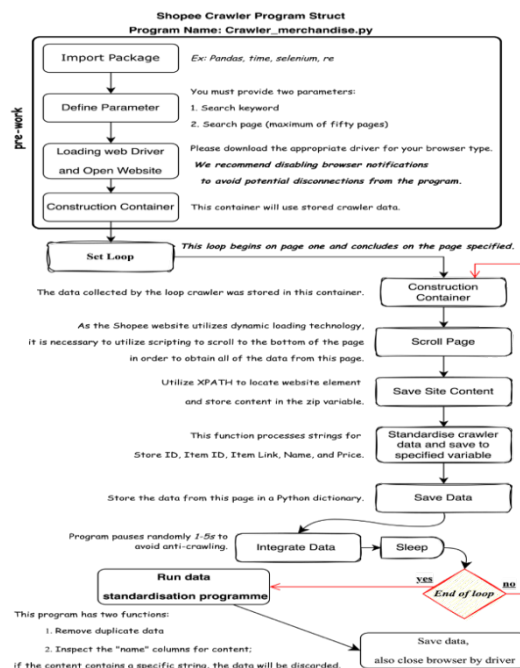


圖 4 商品資料爬蟲程式架構

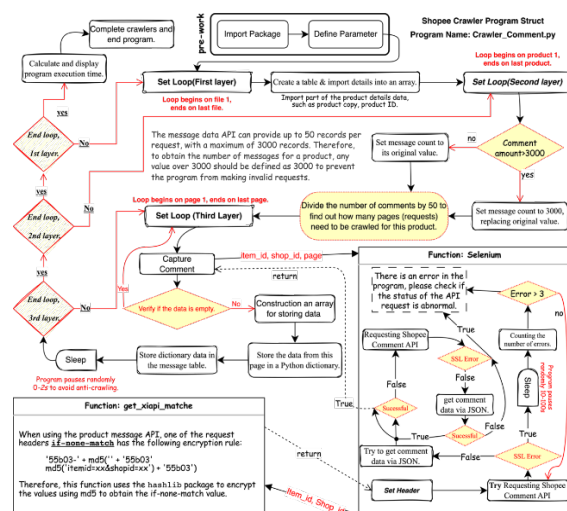


圖 5 商品留言爬蟲程式架構

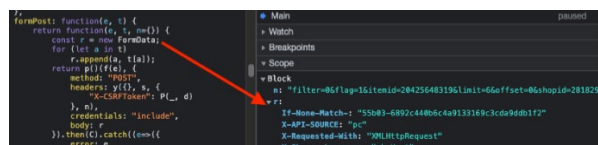


圖 6 留言資料 API-請求標頭

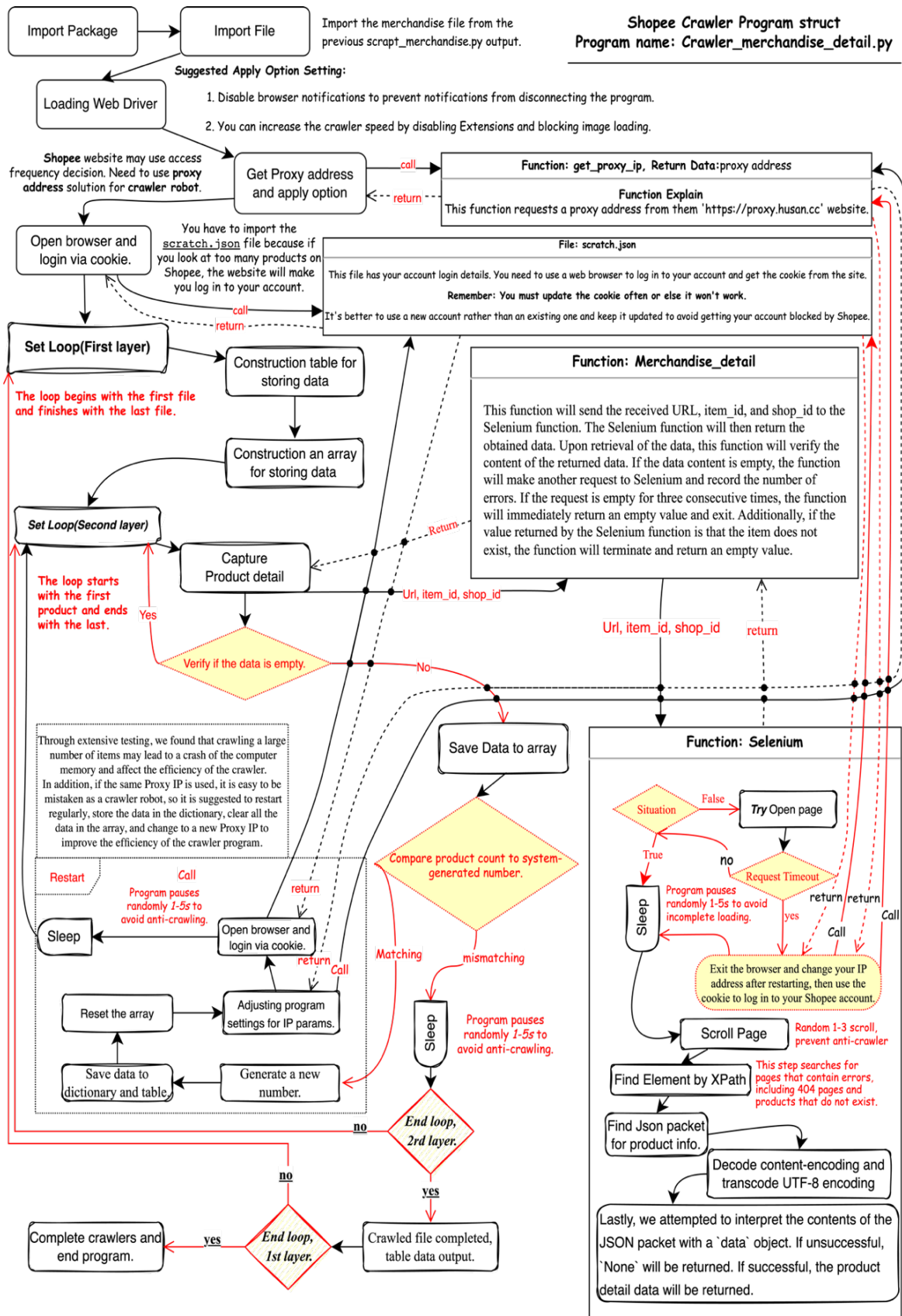


圖 7 商品細節資料爬蟲程式架構

3.2 數據標準化

圖 8 展示了資料標準化的過程，共分為五個步驟。首先，匯入資料並移除特殊字元(例如 emoji、全形字元等)，以確保程式能正確分析字串內容。其次，將評分整數轉換為字串，同時統一顏色和材質數據的定義，例如將聚酰胺條目轉換為尼龍。第三步，程式會分析每筆訂單所購買的項目，並將購買時間、評分等資訊合併到一個清單中。第四步，分析清單數據，例如在類別「排汗衣」中，統計顏色和材質，最後根據資料儀表板和機器學習的需求，分別製作不同的表格。

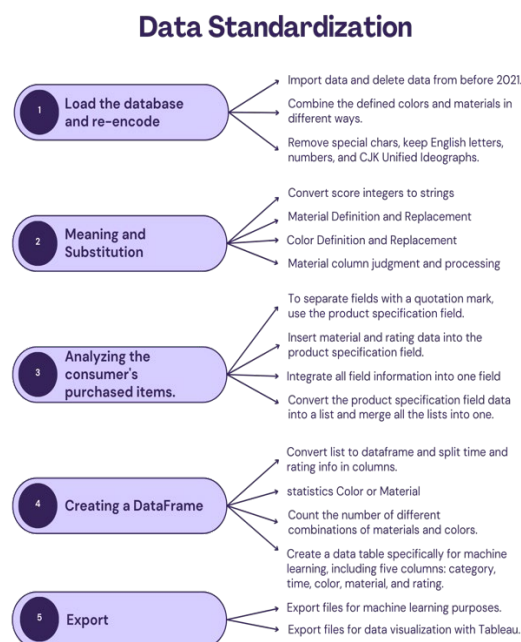


圖 8 資料標準化流程

3.3 機器學習

圖 9 為模型建構流程圖，由於原始數據不足，因此模型將以評論數進行建模。首先將資料進行預處理，並將資料分為訓練集及驗證集，接著尋找最佳學習方法，調整參數，若效果仍不理想，檢查資料架構是否有誤，修正後重新建立模型。完成建模後，我們根據各類產品的 2022 年 12 月銷量，計算出 95% 置信區間。我們使用區間的上限和下限，將模型預測的評論數值除以一个隨機區間值，以估算

產品的真實銷量。最後，我們使用燈號定義預測數值並建立互動資料儀表板。

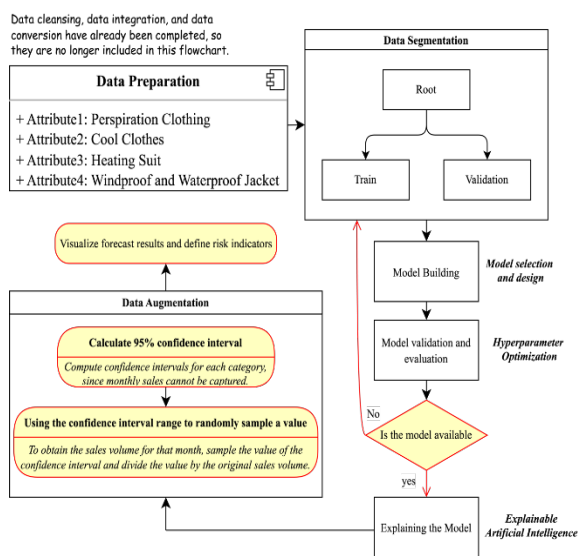


圖 9 模型建構流程圖

3.3.1 數據權重調整與分割

此研究包含 2021 年和 2022 年的四大類別資料。考慮到 2021 年全球經濟下滑和資料量不足的特殊情況，通過權重調整的方式，我們將 2021 年與 2022 年日期欄位中的年份合併，以提高整體模型的精度和預測結果。以排汗衣為例，在權重調整公式 1 中，2021 年的權重約為 0.3，2022 年的權重約為 0.7。調整後，我們得到合併後的日期欄位和重新計算後的總留言數，以此得知 2021 年和 2022 年該組合的留言數，請參見圖 10。最後，我們將資料分為 20% 的驗證資料，80% 的訓練資料[13]。

$$\left(\begin{aligned} \text{Weight of 2021} &= \frac{2021 \text{ Total Sales}}{2021 \text{ Total Sales} + 2022 \text{ Total Sales}} \\ \text{Weight of 2022} &= \frac{2022 \text{ Total Sales}}{2021 \text{ Total Sales} + 2022 \text{ Total Sales}} \end{aligned} \right) \quad (1)$$

	A	B	C	D	E	F	G	H	I	J
1	Type	Color	Material	Month	Rate	Y	Y_2021	Y_2022	2021_Weight	2022_Weight
2	Perspiration Black	Polyester		Jan	5	136.57634	27.414548	109.16179	0.33029576	0.669704239
3	Perspiration Black	Polyester		Jan	4	6.3485212	0.9908873	5.3576339	0.33029576	0.669704239
4	Perspiration Black	Polyester		Jan	3	1.3394085	0	1.3394085	0.33029576	0.669704239
5	Perspiration Black	Wool		Jan	5	2.009127	0	2.009127	0.33029576	0.669704239
6	Perspiration Black	Elastic Fibers		Jan	5	46.961071	1.32183	45.539868	0.33029576	0.669704239
7	Perspiration Black	Elastic Fibers		Jan	4	4.3485212	0.3302959	4.0182254	0.33029576	0.669704239
8	Perspiration Black	Elastic Fibers		Jan	3	1.3394085	0	1.3394085	0.33029576	0.669704239

圖 10 機器學習訓練資料集權重調整

3.3.2 模型篩選

表 2 顯示五種機器學習演算法進行不同服裝進行篩選的結果。並以下列三個指標來評估模型：

- 平均絕對誤差 (MAE)：將預測值與實際值間的差異的絕對值加總，然後除以樣本數，MAE 越小越好。
- 均方誤差 (MSE)：將預測值與實際值間的差異的平方加總，然後除以樣本數，MSE 越小越好。

- R 平方(R^2)：表示模型解釋的變異量佔總變異量的比例，取值在 0 到 1 之間越接近 1 則表示模型越好。

結果發現，使用 Bootstrap Forest 和 Neural Boosted 可以產生良好的訓練和驗證結果，因此我們將對這兩種方法進行後續的模型參數調整。另外，使用 XGBoost 在訓練資料中可以獲得不錯的效果，但在驗證資料中卻非常不理想，因此在下一小節中，我們將探討該方法在訓練上是否過度擬合。

表 2 模型篩選結果

Indicators/ Categories	Perspiration Clothing		Cool Clothes		Heating Suit		Windproof and Waterproof Jacket	
Evaluation Index	R ²	RASE	R ²	RASE	R ²	RASE	R ²	RASE
	Train	Train	Train	Train	Train	Train	Train	Train
	Validation	Validation	Validation	Validation	Validation	Validation	Validation	Validation
XGBoost	0.9657	10.064	0.9603	12.74	0.9305	26.973	0.9446	4.721
	0.6734	28.382	0.1508	49.717	0.2637	81.689	0.3217	19.92
Boosted	0.8975	17.937	0.9154	18.588	0.638	61.486	0.8178	8.563
Tree	0.7676	23.94	0.7253	28.277	0.3281	78.035	0.3772	19.088
Neural	0.8829	18.593	0.8616	23.773	0.761	50.016	0.8391	8.047
Boosted	0.8634	18.357	0.8896	17.926	0.8	42.572	0.6554	14.198
Bootstrap	0.8708	19.527	0.8387	25.668	0.8411	40.786	0.801	8.949
Forest	0.7808	23.251	0.6708	30.953	0.8514	36.701	0.7066	13.1
Decision	0.7632	26.44	0.4846	45.887	0.5605	67.827	0.7546	9.937
Tree	0.5293	34.072	0.3523	43.420	0.6471	56.53	0.4944	17.199

3.3.3 模型參數調整

透過模型篩選，我們選出較佳的模型進行參數調整，並擇優用機器學習。同時，我們檢查使用 XGBoost 方法是否會導致模型過度擬合，評估是否需要增加模型複雜度提高預測精確度。

Neural Boosted 方法中，我們透過 Boosting 技術逐步優化模型，以達到更精確的預測值。為了避免學習率過高導致模型不易收斂，我們將學習率設

定為 0.1，並將模型數量設為 100。在表 3 中可以看到，經過優化後，各類別表現都提升了約 5%。

在 Bootstrap Forest 方法中，我們可以使用預設的參數獲得不錯的成效。然而，有些類別的成效不夠理想。因此，本研究調整了超參數設置，把 "Number Trees"設置為 100 至 500，"Minimum Splits per Tree"設置為 10，"Maximum Splits per Tree"設置為 2000(最大值)。同時，把 "Early Stopping"設置為

True，以防止訓練時間過長。"Early Stopping"是一個功能，可以在模型的驗證性能停止改進時提前停止將樹添加到引導森林，即參數"Maximum Splits per Tree"設置為最大值的目的是。

表 4 為超參數優化的結果。我們測試 100 到 500 顆樹，發現每個類別的驗證效果約提升了 10%。然而，由於參數和模型擬合之間沒有明確的相關性(如圖 11 不同樹木數量對 R²驗證結果的影響-涼感衣所示)，以上結果可能不足以為最佳。

表 3 Neural Boosted 使用 Boosting 方法的優化前後結果

Indicators/ Categories	Perspiration Clothing	Cool Clothes	Heating Suit	Windproof and Waterproof Jacket
R ² (Before)	Train: 0.883 Validation: 0.863	Train: 0.862 Validation: 0.890	Train: 0.761 Validation: 0.800	Train: 0.839 Validation: 0.655
R ² (After)	Train: 0.934 Validation: 0.923	Train: 0.922 Validation: 0.930	Train: 0.814 Validation: 0.864	Train: 0.906 Validation: 0.764
Increase the range	Train: 5.78% Validation: 6.95%	Train: 6.96% Validation: 4.49%	Train: 6.96% Validation: 8%	Train: 7.99% Validation: 16.64%

表 4 Bootstrap Forest 使用超參數設置的優化前後結果

Indicators/ Categories	Perspiration Clothing	Cool Clothes	Heating Suit	Windproof and Waterproof Jacket
R ² (Before)	Train:0.883 Validation:0.781	Train:0.839 Validation:0.671	Train:0.841 Validation:0.851	Train:0.801 Validation:0.707
R ² (After)	Train:0.815 Validation:0.878	Train: 0.843 Validation: 0.849	Train: 0.858 Validation: 0.922	Train:0.777 Validation:0.786
Increase the range	Train: -7.7% Validation: 12.42%	Train: 0.48% Validation: 26.53%	Train: 2.02% Validation: 8.34%	Train: -3.0% Validation: 11.17 %
Parameter	Number of Trees in the Forest: 423	Number of Trees in the Forest: 316	Number of Trees in the Forest: 285	Number of Trees in the Forest: 147

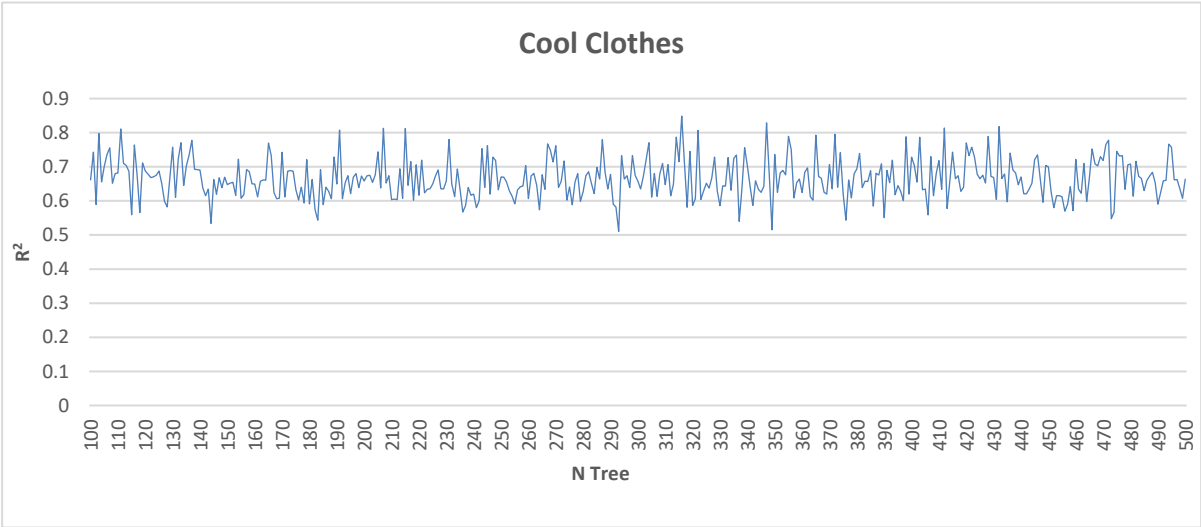


圖 11 不同樹木數量對 R²驗證結果的影響-涼感衣

最後，我們將檢查 XGBoost 方法是否會出現過度擬合的現象。我們使用 80% 的資料進行訓練，20% 的資料進行驗證，並觀察 50 個模型的學習曲線。通過假說檢定，我們驗證了是否存在過度擬合的可能性。

在圖 12 中，我們展示了使用 XGBoost 方法在排汗衣、涼感衣、發熱衣和防風防水外套中的學習曲線。四個模型均出現了嚴重的過度擬合現象。進一步的假設檢定證明，這四種類別均存在過度擬合的情況，其 P 值均小於 0.05。根據這些結果，我們可以得出結論：使用 XGBoost 方法訓練本研究資料會導致過度擬合的問題。

在三種方法中，XGBoost 的過度擬合問題無法透過超參數優化來改善，因此放棄此方法。而在 Boosting 與超參數調整後，Neural Boosted 與 Bootstrap Forest 的模型皆有顯著改善，但整體上 Neural Boosted 的改善效果更好，因此最終決定採用 Neural Boosted 進行模型最佳化。

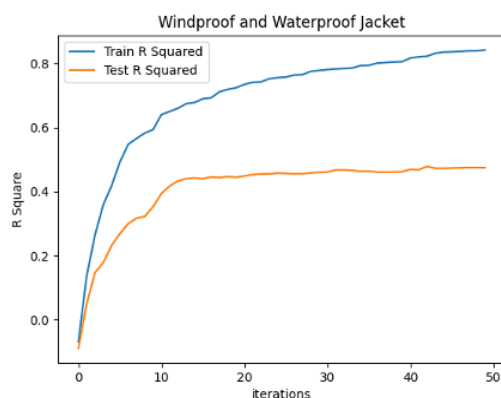
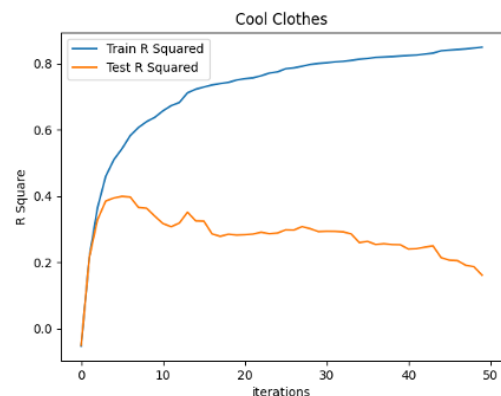


圖 12 學習曲線

3.3.4 模型最佳化

經過對三種機器學習方法進行超參數優化後，發現類神經網路是較佳的機器學習模型。然而，對於"防風防水外套"這個類別，其 R^2 值未達到可應用的水準。為了解決這個問題，本研究採用了 K 重交叉驗證方法，以提高模型的驗證效果、準確性，並使其接近可用水準。

使用類神經網路模型進行 5 折 K 重交叉驗證，每個模型的訓練資料比例為 80%，驗證資料比例為 20%。Boosting 與學習率的設置分別為 100 與 0.1，TanH 為三層。表 5 顯示了各類別使用 K 重交叉驗證的結果，可以看出在防風防水外套的驗證效果已經有所改善。

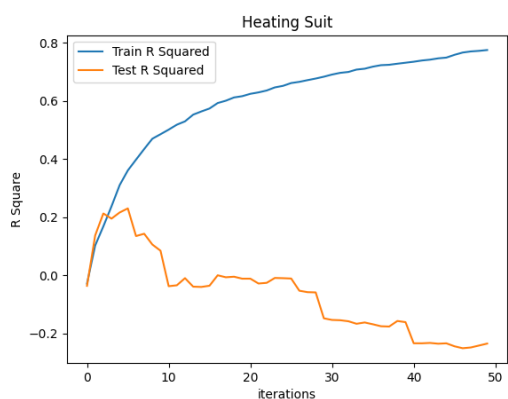
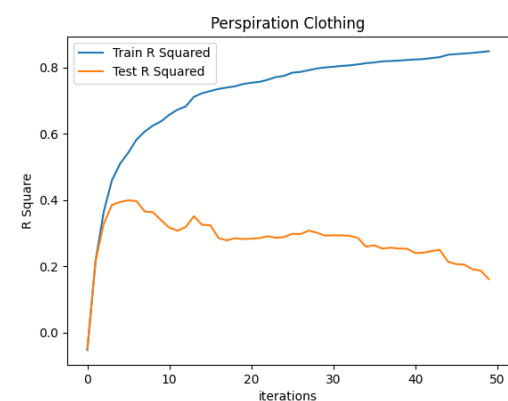


表 5 Neural Boosted（使用 K 重交叉驗證）-各類結果比較

Indicators/ Categories	Perspiration Clothing	Cool Clothes	Heating Suit	Windproof and Waterproof Jacket
R²	Train: 0.9611	Train: 0.9757	Train: 0.9527	Train: 0.8624
	Validation: 0.98	Validation: 0.9858	Validation: 0.8622	Validation: 0.8537
RASE	Train: 10.9515	Train:9.395	Train: 20.188	Train: 7.9416
	Validation: 6.2197	Validation: 8.188	Validation: 47.554	Validation: 7.2609
SSE	Train: 346992.97	Train: 27425.23	Train: 1462040.4	Train: 144936.04
	Validation: 27930.87	Validation: 52697.41	Validation: 2028522.7	Validation: 30262.11
N	Train: 2890	Train: 3143	Train: 3587	Train: 2298
	Validation: 722	Validation: 786	Validation: 897	Validation: 574

3.3.5 模型測試

完成模型建構後，我們使用增量式網路爬蟲獲取 2023 年 1 月份的資料進行模型測試。在評估模型之前，我們需要對預測資料進行預處理，將預測為負值的資料轉換為 0，並移除預測值為空值的資料。表 6 不同類別之評估指標顯示不同功能性衣

物類別的預測準確度。四個類別中，平均絕對誤差相近，排汗衣表現最佳。均方誤差方面，發熱衣誤差較高。R Square 指標顯示發熱衣表現最佳，代表變異量較低，即受快時尚影響較少。涼感衣的數據為負值，顯示其容易受到快時尚影響，但由於一月份為冬季，涼感衣的銷量較低，因此夏季的數據可能會有顯著變化。

表 6 不同類別之評估指標

Indicators/ Categories	Perspiration Clothing	Cool Clothes	Heating Suit	Windproof and Waterproof Jacket
MAE	5.01793722	7.942196532	14.06225166	6.328267477
MSE	174.5695067	432.7976879	1464.377483	234.0790274
R²	0.535230669	-0.697242499	0.900874869	0.596086857

3.4 資料調整與視覺化

本章節旨在利用預測數據集，模擬染整廠商對於機能衣物的顏色、材質、銷量等需求的預測行為，並且利用 Tableau 整合所有的資訊，建立資料儀表板，最終以染整廠商的角度，提供廠商所需的資訊。

● 月銷量推估

由於蝦皮電商平台上的月銷量會隨時間變更，且沒有歷史月銷量資料可供參考。為此，本研究利用爬蟲技術擷取評論資料，其中包含每位消費者購買商品的資訊，例如產品、評價及時間等。藉由商品資料與評論資料建立模型，並將模型推估出來的數據除以其信賴區間隨機值，得到推估月銷量。根據十二月的留言數據，可以計算出推估值。其中 \hat{p} 值為十二月購買數量除以商品資料中的月銷量總和， n 則為十二月購買數量。接下來利用此數據計算 95%信賴區間的上限與下限，其計算公式如下：

$$\left\{ \hat{p} = \frac{\text{評論購買數}}{\text{月銷量}} \right\} \quad (2)$$

$$\left\{ \hat{p} \pm Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p} \times (1 - \hat{p})}{n}} \right\} \quad (3)$$

● 警示燈號方法

警示燈號應用於資訊呈現的方法，輔助染整廠在電商市場中掌握各種衣物顏色、材質組合的競爭力和潛力。本研究定義五種警示燈號，包括綠燈、黃燈、紅燈、上升黃燈和下降黃燈(請參見表 7 警示燈號定義)，並將其應用於資料儀表板上，為染整廠提供相關警示。

● 資料視覺化

為了更準確、高效地傳遞市場趨勢數據，我們使用圖表、圖形和設計元素對其進行可視化。這有助於染整廠更容易地了解數據的意義、趨勢和相關性，以便染整廠發現商業價值的洞見和價值，並提前應對市場變化。圖 13 Tableau 建置流程為 Tableau 建置流程為數據可視化的建置流程。

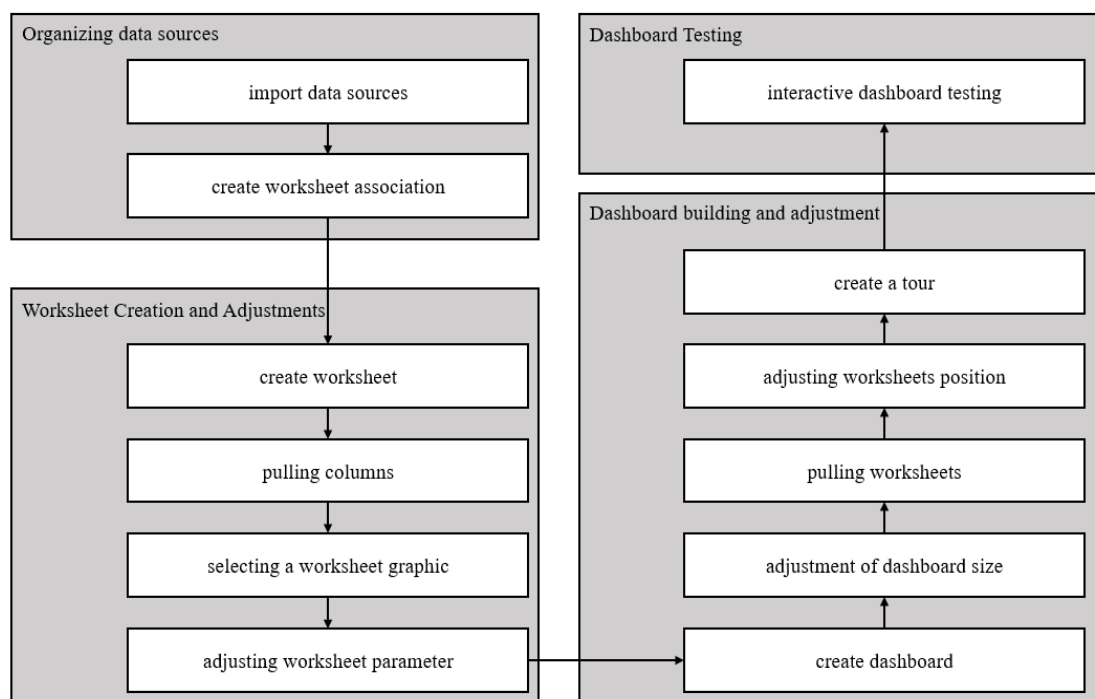


圖 13 Tableau 建置流程

表 7 警示燈號定義

Lights	Red	Yellow-Divide	Yellow	Yellow-Rise	Green
percentile	Below 95%	95% to 99.7% A 20% reduction from last month.	95% to 99.7%	95% to 99.7% 20% more compared to last month	More than 99.7%

在本節中，我們介紹了資料調整與視覺化的方法。首先，我們使用了月銷量推估法，以得知不同顏色和材質的組合的銷售數量。接著，我們使用

警示燈號法，將銷售數值轉換為燈號。請參照圖 14 利用月銷量推估並結合燈號警示方法調整後的數查看調整後的數據。最後，我們使用調整後的數據在 Tableau 上建立了一個互動式資料儀表板。

Type	Month	Material, Color	Rate	Y	Validation	Predicted Y	Expend_Predict_Y	P-Value	Light
1 Perspiration Clothing	4	Polyester; Black	5	629.71542863	Training	626.52267794	2245.349608	0.2790312367	●
2 Perspiration Clothing	5	Polyester; Black	5	734.85211938	Validation	717.31415633	2571.6935167	0.2789267662	●
3 Perspiration Clothing	7	Polyester; Blue	5	729.83853071	Training	596.20837355	2149.5263256	0.2773673281	●
4 Heating Suit	12	Synthetic Material and cotton; Black	5	1337.0392182	Training	1368.7004639	9288.2886386	0.1473576584	●
5 Heating Suit	12	Polyester and Polartec; Gray	5	1255.8957989	Training	1006.3552398	6829.3378915	0.1473576584	●
6 Heating Suit	12	Polyester and Polartec; Dark Gray	5	1368.1001395	Training	912.14691928	6190.0204541	0.1473576584	●
7 Windproof and Waterproof Jacket	11	Synthetic Material; Black	5	259.17676914	Training	191.89632159	1243.5718796	0.154310599	●
8 Windproof and Waterproof Jacket	11	cotton; Black	5	276.89374871	Training	223.63876963	1449.2767907	0.154310599	●
9 Windproof and Waterproof Jacket	12	Polyester; Black	5	221.02632556	Training	176.7490017	1157.4354607	0.1527074361	●
10 Windproof and Waterproof Jacket	12	cotton; Red	5	24.21955849	Training	30.658290618	200.76488346	0.1527074361	●
11 Windproof and Waterproof Jacket	12	Polyester; Dark Gray	5	19.811471013	Training	37.62110996	246.3606941	0.1527074361	●
12 Windproof and Waterproof Jacket	12	Synthetic Material; Dark Gray	5	162.58485661	Training	112.17608855	734.58170336	0.1527074361	●
13 Windproof and Waterproof Jacket	12	Polyester; Multi-color(over two color)	5	211.09919538	Training	112.4205893	736.18280908	0.1527074361	●
14 Windproof and Waterproof Jacket	12	Synthetic Material; Multi-color(over tw...	5	19.89139674	Validation	40.597595974	265.85212222	0.1527074361	●
15 Windproof and Waterproof Jacket	12	cotton; Multi-color(over two color)	5	0.6563234991	Training	45.507279632	298.00303629	0.1527074361	●
16 Perspiration Clothing	3	Polyester; Black	5	414.28713097	Training	460.42045581	1721.4736771	0.2674571572	●
17 Perspiration Clothing	3	Polyester; Blue	5	196.81326972	Validation	209.66533456	783.92119613	0.2674571572	●
18 Cool Clothes	2	Polyester; Black	5	151.55747111	Training	164.66824962	640.9717475	0.2569040683	●
19 Cool Clothes	3	Polyester; Black	5	467.93660297	Training	475.24081662	1842.4570876	0.2579386081	●
20 Heating Suit	9	Polyester; Black	5	53.284535743	Training	69.555034713	473.21814319	0.1469830261	●
21 Heating Suit	10	Polyester; Black	5	350.68023497	Training	428.76130217	2926.0172928	0.1465340971	●
22 Heating Suit	10	Synthetic Material; Black	5	300.29677442	Training	284.50022135	1941.5291522	0.1465340971	●
23 Windproof and Waterproof Jacket	8	Polyester; Black	5	9.6252939963	Training	29.704829926	194.32484689	0.1528617179	●
24 Windproof and Waterproof Jacket	10	cotton; Dark Blue	5	41	Training	47.173162722	307.5642354	0.1533766196	●
25 Perspiration Clothing	1	Polyester; Black	5	136.5763391	Training	169.99205787	643.25447339	0.2642687535	●
26 Perspiration Clothing	1	Polyester; Gray	5	53.687929671	Validation	56.804084216	214.94816721	0.2642687535	●
27 Perspiration Clothing	1	Polyester; Blue	5	59.779056834	Training	68.796998496	260.32967418	0.2642687535	●
28 Perspiration Clothing	2	Polyester; Black	5	128.24604334	Validation	155.7969661	595.1224745	0.2617897538	●
29 Perspiration Clothing	3	Polyester; Gray	5	152.54900095	Training	165.17761657	617.58532963	0.2674571572	●
30 Cool Clothes	1	Polyester; Black	5	135.53154222	Validation	139.12662188	524.99583655	0.2650051909	●
31 Heating Suit	1	Elastic Fibers; Black	5	106.50936485	Training	99.183142227	668.51194239	0.1483640545	●
32 Heating Suit	1	Synthetic Material; Black	4	39.380215039	Training	65.780674266	443.37339328	0.1483640545	●
33 Windproof and Waterproof Jacket	1	Synthetic Material; Black	5	28.124118011	Training	93.559608403	614.46516918	0.1522618581	●
34 Windproof and Waterproof Jacket	1	Nylon; Black	5	74.483309263	Training	87.897825921	577.28065986	0.1522618581	●
35 Perspiration Clothing	1	Polyester; Black	4	6.3485211938	Validation	8.5595760892	32.389663839	0.2642687535	●
36 Perspiration Clothing	1	Polyester; Black	3	1.3394084775	Training	1.8481455953	6.9934321436	0.2642687535	●
37 Perspiration Clothing	1	Wool; Black	5	2.0091127163	Training	1.2365580891	4.6791687352	0.2642687535	●
38 Perspiration Clothing	1	Elastic Fibers; Black	5	46.86107128	Training	38.697068567	146.43073785	0.2642687535	●
39 Perspiration Clothing	1	Elastic Fibers; Black	4	4.3485211938	Training	1.4933321032	5.6508084418	0.2642687535	●
40 Perspiration Clothing	1	Elastic Fibers; Black	3	1.3394084775	Training	1.357842465	5.1381120431	0.2642687535	●

圖 14 利用月銷量推估並結合燈號警示方法調整後的數

4. 分析與討論

本研究使用 Tableau Public 建立互動式儀表板，展示機能衣物銷售的模型預測。總共建立了四張儀表板，包括主頁面、指標資訊、燈號警告和組合趨勢。主頁面顯示四大類別中最佳組合的月銷量預測折線圖。點擊右上角的巡覽按鈕，可轉移至其他儀表板，如圖 15。指標訊息儀表板分為 2023 年度預測總結和燈號重點警示兩個區塊。2023 總結主要展示四大類別的整體銷售量、月銷量和趨勢。警示燈號提醒銷量強勢和可能提高的組合。透過這些資訊，廠商可了解各類別機能衣物的可能變化和提高整體效益的生產組合，進而提前進行原料的儲備、研發、生產，提早佈局市場，如圖 16。

透過燈號的分類歸屬，廠商可以了解當前市場的偏好和變化，例如黃燈上升燈號，可以得知可能成為熱門商品的組合；綠燈燈號，則表明該組合為普遍受歡迎的組合，圖 17。

警告儀表板雖然擁有所有類別組合的燈號，但無法呈現所有組合，必須透過條件的保留或排除，才能呈現所需資料。顯示組合趨勢儀表板所有組合每月銷量表現的折線圖。如需較完整的資料數據，則須以條件排除的方式進行篩選，如圖 18 所示。

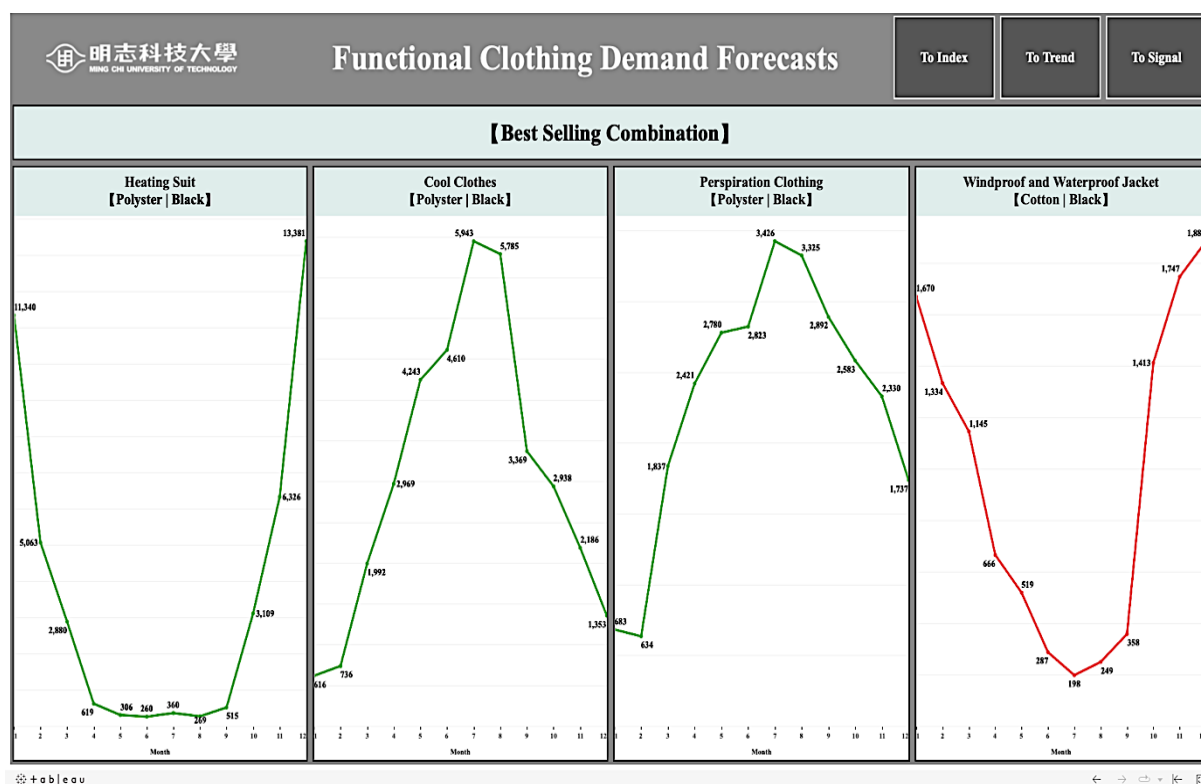


圖 15 互動式儀表板-首頁

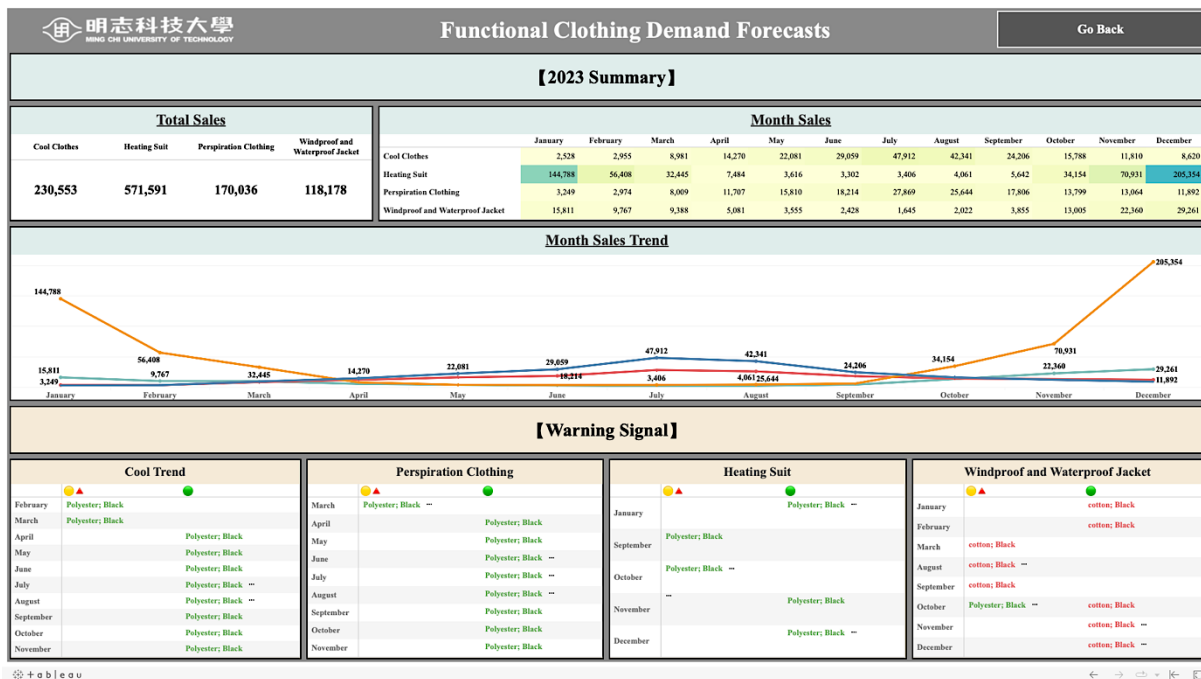


圖 16 互動式儀表板-指標資訊



圖 17 互動式儀表板-燈號警告

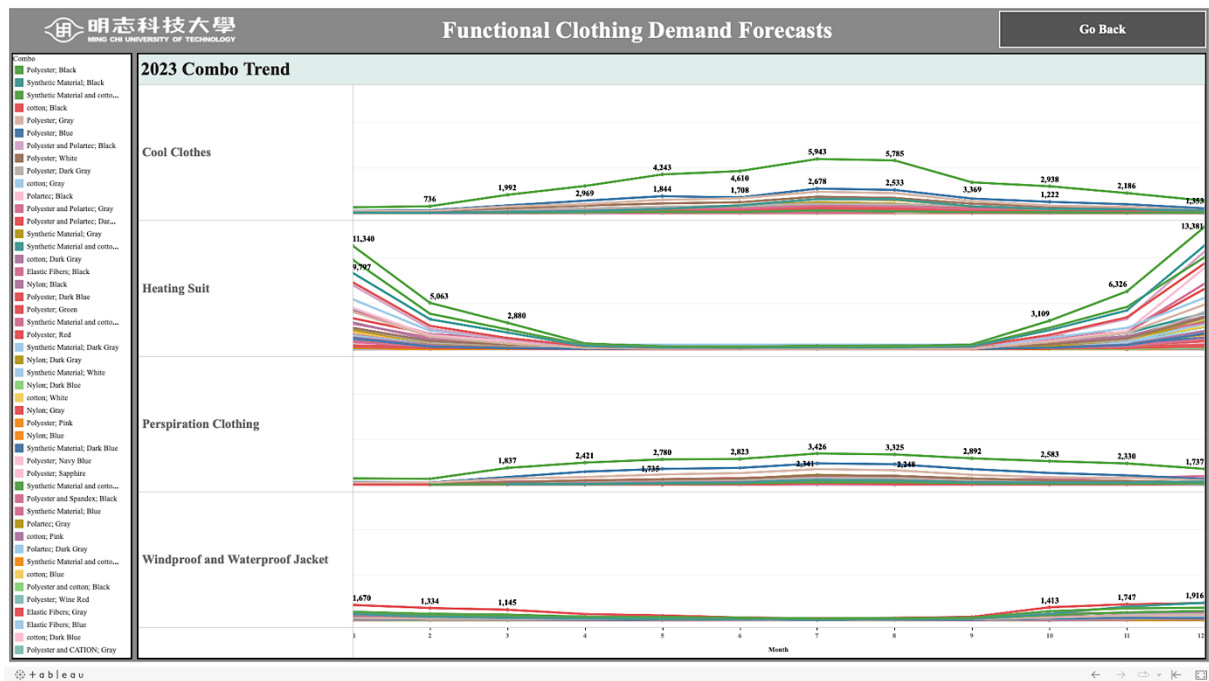


圖 18 動式儀表板-組合趨勢

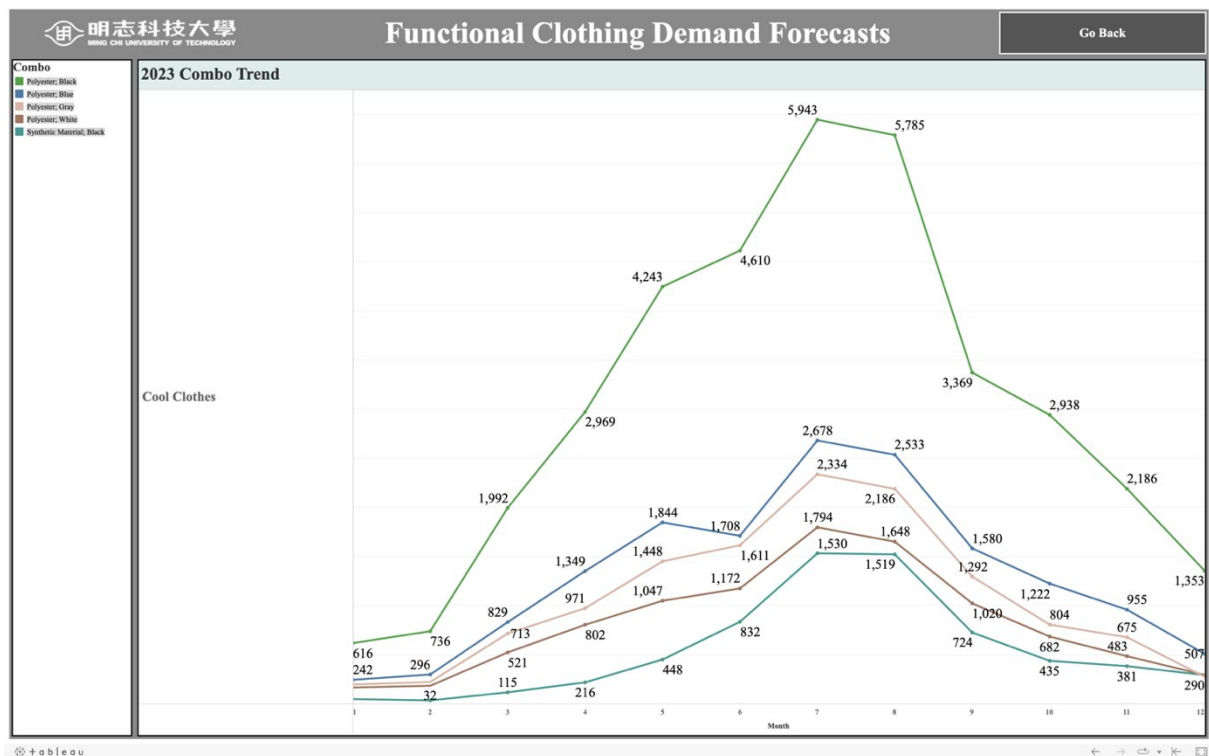


圖 19 互動式儀表板-組合趨勢條件保留

藉由折線圖的呈現，廠商可以了解不同類別組合的每月銷售表現，並探討月份對銷量的影響。舉

例來說，夏季發熱衣的銷售量肯定會比冬季低，但不知道有多少，廠商無從參考，但透過折線圖的呈

現，廠商可以根據資料來擬定生產計畫，提前一步進入市場。考慮到本研究中的組合數量較多，因此儀表板的設計和數據的呈現相對複雜。但是，在染整產業中，這些資訊尤其重要且關鍵，因為從消費者的角度進行分析，廠商能夠瞭解在各種機能類別中什麼顏色配上什麼材質，能夠實現最佳銷售量，這是染整產業必須了解的資訊。

5. 結論

本研究，由消費者角度所取得的外部資訊來建立機能性布料需求決策系統。並透過北部某染整廠來進行實務成效的成果效應的驗證。結果顯示本研究成果確實能夠強化原本公司的需求預測；另外公司高層一致認同，可用於公司未來整體的產能規劃與經營決策。

在研究過程中，本研究使用網路爬蟲擷取資料，評估各家電商平台的資料以及爬取難度後，發現蝦皮電商平台在使用者、資訊量和爬蟲效率方面表現優異，因此成為研究對象。然而，僅有一個平台的資料量不足以推估全台灣電商機能衣物的銷量，因此只能得知蝦皮電商平台上的機能衣物銷量。在模型建立方面，蝦皮電商平台的熱門綜合排名及高銷量商品置頂等因素，導致消費者偏好較高銷量或排名較佳的商品，進而使原始資料的極端值較大，因此必須先處理資料，才能建立模型。

此外，在學術與實務的貢獻方面，就學術而言，本研究透過資訊技術與數據分析提出解決爬蟲問題與機器學習資料處理的方法的一套具邏輯性的程序；就實務而言，則提供可決策的有效訊息來解決廠商面臨的需求預測問題。

在導入實際場域的過程中發現，染整廠雖然可以透過資料儀表板快速得知未來趨勢，但資料內容並未能定時更新。因此，本研究未來將持續所開發的程式置放於伺服器中，全天候執行爬蟲與資料處理，使儀表板的數據更加準確且能顯示更多趨勢。

參考文獻

1. Ha, C., Seok, H., & Ok, C. (2018). Evaluation of forecasting methods in aggregate production planning: A Cumulative Absolute Forecast Error (CAFE). *Computers & Industrial Engineering*, 118, 329-339.
2. KALAOGLU, Ö. İ., Akyuz, E. S., Ecemiş, S., Eryuruk, S. H., SÜMEN, H., & Kalaoglu, F. (2015). Retail demand forecasting in clothing industry. *Textile and Apparel*, 25(2), 172-178.
3. Rendon-Sanchez, J. F., & de Menezes, L. M. (2019). Structural combination of seasonal exponential smoothing forecasts applied to load forecasting. *European Journal of Operational Research*, 275(3), 916-924.
4. Chuang, C. H., & Zhao, Y. (2019). Demand stimulation in finished-goods inventory management: Empirical evidence from General Motors dealerships. *International Journal of Production Economics*, 208, 208-220.
5. Lorente-Leyva, L. L., Murillo-Valle, J. R., Montero-Santos, Y., Herrera-Granda, I. D., Herrera-Granda, E. P., Rosero-Montalvo, P. D., ... & Blanco-Valencia, X. P. (2019, September). Optimization of the master production scheduling in a textile industry using genetic algorithm. In *International Conference on Hybrid Artificial Intelligence Systems* (pp. 674-685). Springer, Cham.
6. Lu, C. J. (2014). Sales forecasting of computer products based on variable selection scheme and support vector regression. *Neurocomputing*, 128, 491-499.
7. Lu, C. J., Lee, T. S., & Lian, C. M. (2012). Sales forecasting for computer wholesalers: a comparison of multivariate adaptive regression splines and artificial neural networks. *Decision Support Systems*, 54(1), 584-596.

8. Lu, C. J., & Wang, Y. W. (2010). Combining independent component analysis and growing hierarchical self-organizing maps with support vector regression in product demand forecasting. *International Journal of Production Economics*, 128(2), 603-613.
9. Prestwich, S. D., Tarim, S. A., Rossi, R., & Hnich, B. (2014). Forecasting intermittent demand by hyperbolic-exponential smoothing. *International Journal of Forecasting*, 30(4), 928-933.
10. Lolli, F., Gamberini, R., Regattieri, A., Balugani, E., Gatos, T., & Gucci, S. (2017). Single-hidden layer neural networks for forecasting intermittent demand. *International Journal of Production Economics*, 183, 116-128.
11. Petropoulos, F., Nikolopoulos, K., Spithourakis, G. P., & Assimakopoulos, V. (2013). Empirical heuristics for improving intermittent demand forecasting. *Industrial Management & Data Systems*, 113(5), 683-696.
12. Raghavendra, S. (2021). *Python Testing with Selenium: Learn to Implement Different Testing Techniques Using the Selenium WebDriver*. Apress Media LLC.
13. 李家岩, & 洪佑鑫. (2022)。製造數據科學：邁向智慧製造與數位決策。前程文化。