



逢甲大學學生報告 ePaper

報告題名：臺灣高齡人口長照供需落差分析與需求預測

Analysis of the Long-Term Care Supply–Demand Gap and Demand Forecasting for Taiwan’s
Elderly Population

作者：姚絜歆、林雋勛

系級：人工智慧二

學號：D1345310、D1345337

開課老師：周永振

課程名稱：大數據分析與應用

開課系所：人工智慧技術與應用學士學位學程

開課學年：114 學年度 第 2 學期

中文摘要

人口快速高齡化使長期照顧（以下簡稱長照）資源的區域配置成為重要的公共政策課題。本研究運用政府公開資料，探討臺灣各縣市長照供需落差的空間分布，並預測其未來走向，核心目的在於檢驗「衡量指標的選擇如何左右供需結論」。方法上，先以 Python 自政府開放平臺蒐集人口、老化指數、身心障礙人口、老人福利機構與長照服務涵蓋率等資料，經清理後建立縣市主表與多維度供需指標；再整合面量圖視覺化、關聯式資料庫（SQLite/SQL）查詢、機器學習分類與多方法時間序列回測等技術。需求預測對老化指數同時採線性迴歸、年均複合成長率（CAGR）與移動平均外推至 2030 年，並以遮蔽最後三年之回測（平均絕對誤差，MAE）作為選模依據。研究發現，將供給指標由「每千名老人核定床數」改為以失能需求為分母的「長照服務涵蓋率」後，真實供需缺口由都會集中轉為聚焦於離島與北部都會圈，顯示指標選擇會直接翻轉政策判斷；回測亦證實老化正在加速，線性外推將系統性低估未來需求。最終成果部署為公開互動儀表板，可供各界檢視與延伸應用。

關鍵字：長照供需、涵蓋率、資料視覺化、需求預測、機器學習

Abstract

Rapid population aging makes the regional allocation of long-term care (LTC) resources a pressing public-policy issue. This study uses open government data to examine the spatial distribution of the LTC supply–demand gap across Taiwan’s counties and to forecast its trajectory, with the central aim of testing how the choice of measurement indicator shapes supply–demand conclusions. Methodologically, Python is used to collect population, aging-index, disability, elderly-care-institution, and LTC service-coverage data from government open platforms; after cleaning, a county master table and multidimensional supply–demand indicators are constructed, and choropleth visualization, relational-database (SQLite/SQL) querying, machine-learning classification, and multi-method time-series backtesting are integrated. Demand forecasting extrapolates the aging index to 2030 using linear regression, compound annual growth rate (CAGR), and moving average, with a hold-out backtest (mean absolute error, MAE) guiding model selection. The results show that replacing “licensed beds per thousand elderly” with a demand-based “LTC service coverage rate” shifts the true gaps from metropolitan concentration toward the outlying islands and the northern metropolitan area, demonstrating that indicator choice can reverse policy judgments; backtesting further confirms that aging is accelerating and that linear extrapolation systematically underestimates future demand. The final results are deployed as a public interactive dashboard for review and further application.

Keyword : Data visualization; Demand forecasting; Long-term care supply and demand; Machine learning; Service coverage rate

目 次

一、前言.....	4
二、資料來源.....	4
三、研究方法與流程.....	6
3.1 資料蒐集與前處理.....	6
3.2 供需指標之修正.....	7
3.3 視覺化.....	7
3.4 需求預測方法.....	7
3.5 資料庫與查詢.....	8
3.6 供給不足分類模型.....	8
四、主要發現.....	9
4.1 長照服務涵蓋率之空間分布.....	9
4.2 供給不足分類結果.....	10
4.3 時間序列回測與 2030 年推估.....	11
五、結論與政策建議.....	12
六、方法與限制.....	13
6.1 為何未採用 Hadoop / MapReduce / HDFS	13
6.2 資料庫技術範圍：關聯式與 NoSQL／時序資料庫	13
6.3 半結構化資料與文字處理.....	14
6.4 時間序列分析涵蓋說明.....	14
七、AI 工具使用聲明	14

一、前言

人口高齡化已成為臺灣社會與公共政策的核心挑戰。依內政部戶政統計，2025 年底全國 65 歲以上人口約達 475 萬人，且各縣市在老化速度與長照資源（機構、床位、A/B/C 據點）的布建程度上呈現顯著的空間不均。值得注意的是，長照供需是否失衡，往往取決於衡量指標的選擇：若僅以機構床位密度等供給端指標評估，容易忽略照護人力、需求端失能人口與城鄉基數差異，進而導出偏誤的政策結論。

有鑑於此，本研究以政府公開資料為基礎，嘗試回答三個相互關聯的問題。第一，就可得的公開統計而言，哪些縣市的長照供給相對不足？第二，在既有的老化軌跡下，未來（至 2030 年）需求將如何演變？第三，應採用何種供需指標，方能避免因分母或口徑選擇不當而使結論失真？三個問題分別對應「現況診斷」、「趨勢預測」與「方法論反思」，共同構成本研究的分析主軸。

為回應上述問題，本研究建構一套可重現（reproducible）的資料分析流程，涵蓋資料蒐集、清理、指標建置、視覺化、需求預測，並整合關聯式資料庫與機器學習分類，最終部署為公開互動儀表板。全文後續安排如下：第二節說明資料來源，第三節闡述研究方法與流程，第四節呈現主要發現，第五節提出結論與政策建議，第六節則討論方法之適用範圍與限制。

二、資料來源

本研究所用資料均取自政府公開平臺，並遵循政府資料開放授權（多為 OGDL-Taiwan 1.0 或 CC BY 4.0）。內政部 data.gov.tw 之資料透過開放 API（端點：<https://data.gov.tw/api/v2/rest/dataset/{id}>）取得下載連結，其中人口檔（dataset 77132）每月更新，程式將自動掃描最新月份後下載。各資料集之名稱、提供機關、來源網址與檔案年度彙整如表 1，涵蓋人口結構、老化指數、身心障礙人口、老人福利機構、長照服務涵蓋率與行政區地理界線等面向，構成後續指標建置的基礎。

表 1 各資料集名稱、提供機關、來源與年度

編號	資料名稱	提供機關	來源網址	檔案 / 年度
1	村里單一年齡人口 (dataset 77132)	內政部戶政司	data.gov.tw/dataset/77132	77132_pop_single_age_latest.csv (每月更新, 2025/5)
2	各縣市老化指數／老年比率	內政部戶政司	ris.gov.tw/app/portal/346	ris_aging_index_by_county.xls (2010–2025)
3	身心障礙者人數 (年齡×縣市)	衛福部統計處	dep.mohw.gov.tw/dos/cp-5224-62359-113.html	disability_by_age_county.xls (2025)
4	身心障礙者人數 (類別×縣市)	衛福部統計處	dep.mohw.gov.tw/dos/cp-5224-62359-113.html	disability_by_type_county.xls
5	全國老人福利機構 (dataset 8572)	衛福部社家署	data.gov.tw/dataset/8572	elderly_institutions_8572/ (22 縣市, Big5)
6	長照服務涵蓋率 (表三)	衛福部長照司	1966.gov.tw/LTC/lp-6485-207.html	ltc_coverage_rate.pdf (2025)
7	各縣市長照資源布建 A/B/C (表十六)	衛福部長照司	1966.gov.tw/LTC/lp-6485-207.html	ltc_county_resources.pdf
8	台灣縣市界 geojson	g0v (twgeojson)	github.com/g0v/twgeojson	tw_counties.json
附	人口單一年齡組 (dataset 14226)	內政部	data.gov.tw/dataset/14226	14226_population_single_age.csv (2020)

由於各來源的格式與編碼不一，資料處理須留意數項技術細節。國家發展委員會「人口推計」(dataset 6101)之下載端點會封鎖程式化存取(回傳 HTTP 403)，故未來人口改以戶政司老化指數時間序列自建縣市別預測；機構名冊原始檔為 Big5 (CP950) 編碼，讀取時須明確指定編碼參數；衛福部身障統計之副檔名雖為 .xls，實際上為 OOXML (.xlsx) 格式，須以對應函式庫讀取；此外，縣市界 geojson 之欄位使用「台」與「桃園縣」等舊稱，程式已一併正規化為「臺」與「桃園市」，以確保跨資料集的縣市鍵值一致。

三、研究方法與流程

本研究分析流程如圖 1 所示，由資料蒐集出發，歷經清理與縣市主表建置、供需指標建置，再分流至視覺化、需求預測、關聯式資料庫與機器學習四項分析，最終彙整並部署為公開互動儀表板。此一設計兼顧模組化與可重現性，各階段的產出皆以結構化檔案銜接，以下分節說明。

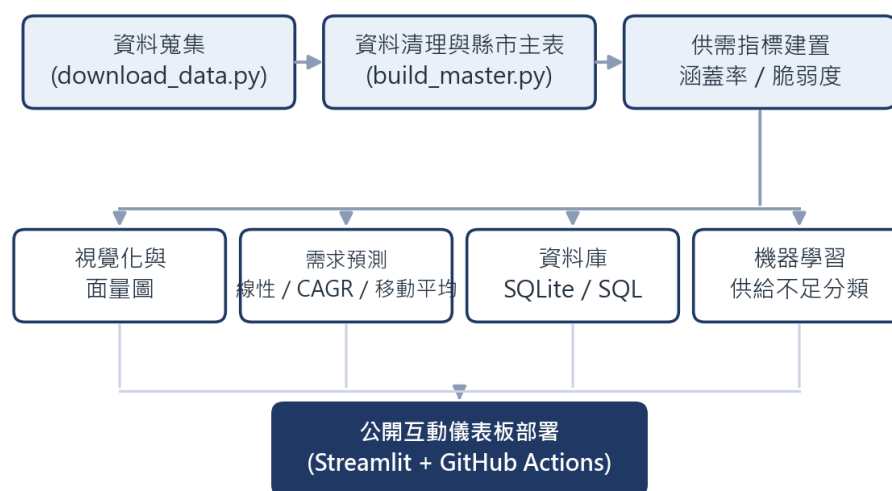


圖 1 研究資料分析流程（自資料蒐集至互動儀表板部署）

3.1 資料蒐集與前處理

資料蒐集以自動化腳本自各機關抓取原始檔。人口檔透過開放 API 取得各月份連結後，採分段請求（僅讀取檔頭部分位元組）並行掃描各檔之期別欄位，以擷取最新月份（民國 114 年 5 月）後再完整下載；其餘來源則以覆蓋方式更新，並以例外處理隔離各來源，確保單一網站失效不致影響整體流程。前處理階段將人口檔依「區域別」彙整至縣市層級，並把逐歲欄位加總為 65 歲以上（65+）與 75 歲以上（75+）人口；機構名冊逐縣市讀取並加總核定床數，身心障礙統計則取最近完整年度分頁定位各縣市數值。上述結果合併為縣市主表，並據以計算核心供給指標「每千名老人核定床數」（核定床數除以 65+ 人口再乘以一千）。為兼顧效率與再現性，重運算僅執行一次並輸出為結構化檔案，後續視覺化與儀表板皆僅讀取結果；整體資料前處理流程如圖 2 所示。

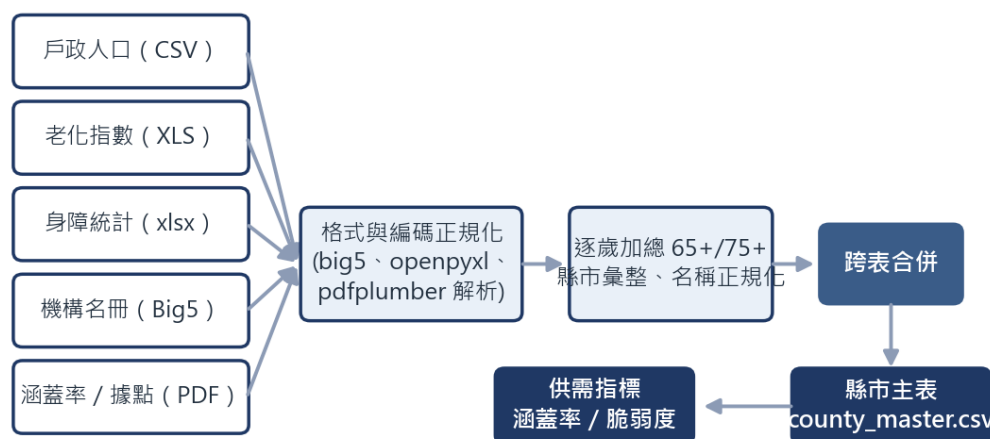


圖 2 資料前處理流程：自異質來源（CSV／XLS／PDF／Big5）經編碼與格式正規化後，彙整為縣市主表與供需指標

3.2 供需指標之修正

研究初期以「每千名老人核定床數」作為供給指標，惟其存在兩項結構性偏誤。其一，該指標僅計床位而未納入照護人力，床位供給若缺乏照服員即形同空床，無法反映實質服務能量；其二，比率型指標易受小基數效應干擾，偏鄉高齡人口絕對數少，少數機構即可使每千人床數大幅偏高，形成「看似充足、實則稀薄」的假象。為矯正上述偏誤，本研究改採衛福部以「失能推估需求人數」為分母之長照服務涵蓋率，並結合 A/B/C 據點達成率，建立兼顧需求端與供給端的多維度脆弱度指標。指標更換對結論的影響，將於第四節與第五節進一步討論。

3.3 視覺化

視覺化同時採靜態與互動兩種形式。靜態圖以程式產出供需缺口排名、老年比例與供需散佈等圖表，並以縣市界地理資料自行繪製臺灣分縣市面量圖（choropleth）；互動版本則於 Streamlit 儀表板以 Plotly 呈現，供使用者依縣市與指標動態檢視。自行以多邊形座標繪圖雖須處理縣市名稱正規化與離島相對位置等細節，但能完整掌握呈現方式，並清楚凸顯區域間的差異，各縣市涵蓋率之空間分布結果詳見第四節。

3.4 需求預測方法

需求預測針對各縣市老化指數與老年比率，以最近十年資料進行外推，並在總人口維持 2025 年水準的保守假設下，以全國中位服務水準推估各縣市 2030 年所需床數

與床位缺口。考量單一模型的局限，本研究並列三種外推方法：線性迴歸假設每年增量固定，適於呈現趨勢方向；CAGR（年均複合成長率）將成長加速納入計算；移動平均則以最近五年平均年增量外推。方法之取捨不以主觀判斷為據，而以回測（backtesting）驗證：遮蔽各縣市最後三年資料，以其前期資料預測該三年並與實際值比較，計算平均絕對誤差（MAE）作為選模準則，其回測概念如圖 3 所示，實證結果則見第四節第三小節。

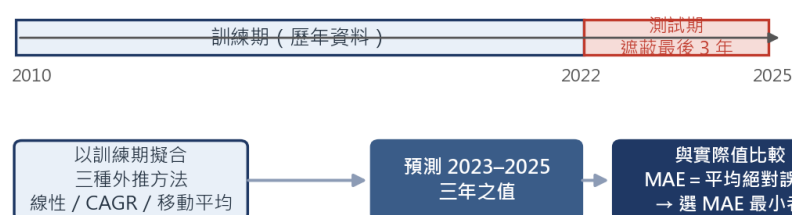


圖 3 時間序列回測（backtesting）示意：遮蔽最後三年為測試期，以訓練期擬合三種外推方法後比較其 MAE

3.5 資料庫與查詢

為支援跨資料表的關聯查詢，本研究將前處理後的多張資料表匯入單一 SQLite 資料庫，並以 SQL 完成聚合、排名與跨表 JOIN 等操作，例如查詢涵蓋率最低之縣市、老年比例最高之縣市，或將縣市主表與 2030 年床位缺口推估相互關聯。儀表板另設「資料庫查詢（SQL）」分頁，於記憶體即時建庫供使用者輸入語法查詢。以「涵蓋率最低前五名」為例，其查詢語法如下所示，展現關聯式模型在多維交叉分析上的簡潔與彈性。

```
SELECT 縣市, 長照服務涵蓋率% FROM county_master ORDER BY 長照服務涵蓋率%
ASC LIMIT 5;
```

3.6 供給不足分類模型

為以資料驅動方式辨識供給不足的結構因子，本研究將問題建構為二元分類：長照服務涵蓋率低於 80% 者標記為「供給不足」（1），其餘為 0，二十二個縣市中有九個被標記為不足。特徵僅採人口與供給結構欄位，包括老年人口比例、高齡人口比例、每千名老人核定床數、每千名老人機構數、A 整合中心達成率與身障老人占比，並刻意排除涵蓋率本身及其衍生欄位，以避免資料洩漏（data leakage）。在僅有二十

二個樣本的限制下，本研究以五折交叉驗證比較邏輯迴歸、決策樹與隨機森林三種模型，兼顧偏誤與變異的評估，完整建模流程如圖 4 所示，分類結果與特徵重要度則詳見第四節第二小節。



圖 4 供給不足分類之建模流程：特徵選取（排除涵蓋率衍生欄位以防資料洩漏）、標記、五折交叉驗證與模型比較

四、主要發現

4.1 長照服務涵蓋率之空間分布

以長照服務涵蓋率衡量，真實供需缺口最大者集中於離島與北部都會圈。如圖 5 所示，連江（33%）、金門（43%）、臺北（58%）、基隆（70%）、新竹市（77%）、桃園（78%）、苗栗（78%）與新北（80%）之涵蓋率明顯偏低，代表其需求未被滿足的程度較高。其中離島情形最為嚴峻：人口稀少使其難以布建 A 整合中心（連江達成率為 0%），亦不易留任照護人力，形成「高齡人口雖少卻乏人照顧」的結構性困境，而傳統床位密度指標對此完全無從反映。相對地，東部與南部縣市涵蓋良好，嘉義縣（115%）、臺東（112%）、花蓮（109%）、屏東（107%）與南投（106%）之涵蓋率均逾百分之百，顯示長照 2.0 的 A/B/C 據點布建在當地已見成效。

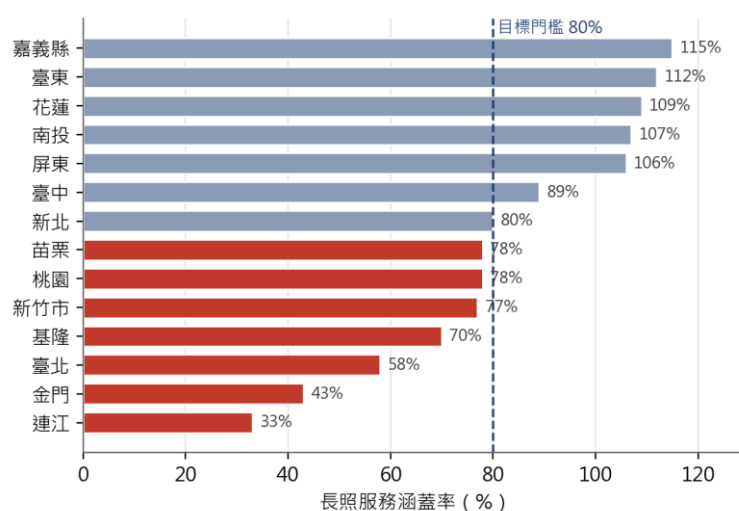


圖 5 代表縣市長照服務涵蓋率（紅色表示低於 80% 門檻，未滿足程度較高）

就人口結構而言，老年人口比例以嘉義縣最高（24.5%）、新竹縣最低（15.3%）；全國 65 歲以上人口約 475 萬人，老人福利機構計 1,107 家、核定床數 62,225 床。尤須強調的是指標校正所帶來的判讀翻轉：若沿用床位密度指標，臺中將被誤判為最危險（實際涵蓋率達 89%），嘉義縣亦會被誤判為供給不足（實際涵蓋率 115%，為全國最佳），此一對比清楚印證了供需指標的選擇會直接改變分析結論，呼應本研究的核心命題。

4.2 供給不足分類結果

在供給不足分類方面，三種模型於五折交叉驗證下的準確率如表 2 與圖 6 所示，以隨機森林表現最佳（ 0.82 ± 0.16 ），優於邏輯迴歸（ 0.78 ± 0.13 ）與決策樹（ 0.73 ± 0.17 ）。由於樣本僅二十二個縣市，準確率的絕對數值宜視為方法示範，交叉驗證與標準差則用以說明模型的穩定度與不確定性。

表 2 三種分類模型之五折交叉驗證準確率

模型	交叉驗證準確率	標準差
邏輯迴歸	0.78	0.13
決策樹	0.73	0.17
隨機森林	0.82	0.16

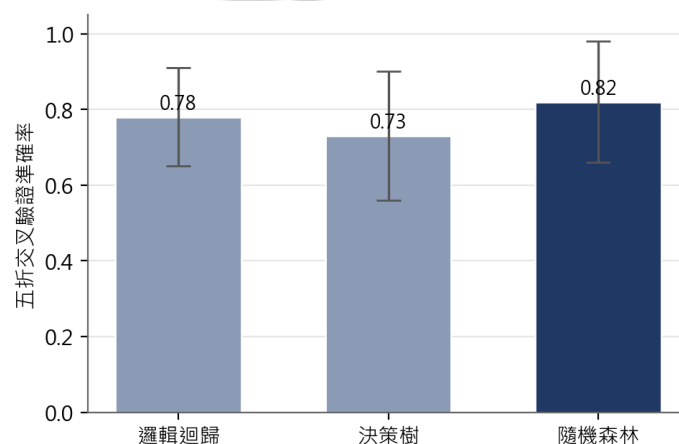


圖 6 三種分類模型之交叉驗證準確率（誤差棒為標準差）

進一步檢視隨機森林之特徵重要度（表 3 與圖 7），可發現決定縣市是否供給不足的關鍵並非供給端變數，而是需求端變數。重要度由高至低依序為身障老人占比、老年人口比例、每千名老人核定床數與高齡人口比例等，其中需求面特徵（身障老人占比、老年人口比例）明顯高於供給面特徵（床位數、機構數）。此一結果與涵蓋率

分析的方向一致，共同指向「需求端的失能壓力」才是供需落差的主導因子，亦說明僅擴充床位未必能有效縮小缺口。

表 3 隨機森林特徵重要度與邏輯迴歸係數

特徵	隨機森林重要度	邏輯迴歸係數
身障老人占比%	0.301	-1.393
老年人口比例%	0.175	-0.046
每千名老人核定床數	0.173	-0.033
高齡人口比例%	0.167	-0.229
每千名老人機構數	0.124	-0.335
A 達成率%	0.060	-0.067

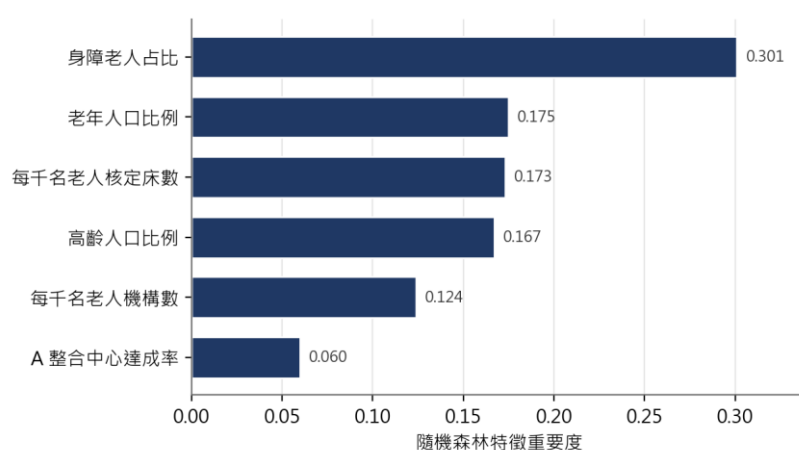


圖 7 隨機森林特徵重要度（需求面特徵之重要度高於供給面）

4.3 時間序列回測與 2030 年推估

時間序列回測結果如表 4 與圖 8 所示：CAGR 的平均絕對誤差最低（3.34），優於移動平均（3.77）與線性（5.25）。回測清楚顯示老化指數呈加速上升（曲線上凸），線性外推因假設固定增量而系統性低估未來值，故其誤差最大；CAGR 因納入複利式成長而最貼近近三年實況。基於此一實證，本研究於推估 2030 年需求時，以線性作為保守下限、CAGR 作為較積極上限，並以「線性～CAGR」區間方式呈現，兼顧解釋性與預測準確度，同時誠實反映預測的不確定性。

表 4 老化指數三種預測方法之回測平均絕對誤差（MAE）

方法	回測 MAE（越小越準）	樣本數
CAGR	3.34	66
移動平均	3.77	66
線性	5.25	66

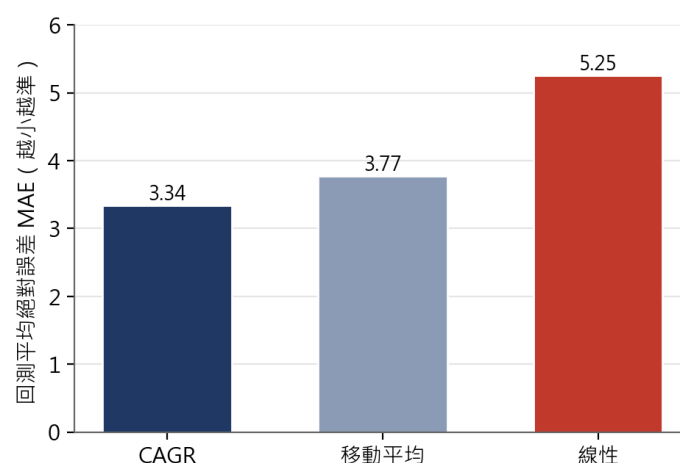


圖 8 三種預測方法之回測 MAE (線性法誤差最大)

五、結論與政策建議

綜合上述分析，以「長照服務涵蓋率」（失能需求相對於實際服務）為主軸衡量，臺灣真實的供需缺口集中於離島與北部都會圈，而非傳統床位密度指標所指向的中部都會。離島為結構上最脆弱的一環，其困境源自人口稀少、據點難以布建與照護人力不易留任；北部都會的缺口則來自龐大需求疊加長照 2.0 給付使用率偏低。相對地，東部與南部因據點布建見效而涵蓋良好，顯示制度設計若能落實於在地，確能縮小缺口。

據此，本研究提出三項政策建議。針對離島地區，宜採在地培力、跨縣市支援與遠距／巡迴照護並行的策略，以補足 A/C 據點與人力缺口；針對北部都會，應著力提升長照 2.0 給付使用率，將外籍看護逐步銜接至正式服務體系，並擴充居家與社區照護量能；就整體資源配置而言，則應以「涵蓋率」與「失能需求成長」作為分配依據，而非僅以床位數多寡衡量。

本研究的核心貢獻，在於以實證方式凸顯「衡量指標的選擇會直接左右政策結論」此一方法論命題：由床位密度改採涵蓋率，使部分縣市的危險程度判定完全翻轉；由線性外推改採多方法回測，則揭示了老化加速情境下線性法的系統性低估。二者共同說明，公共資源的配置決策不僅取決於資料本身，更取決於分析者對指標與方法的審慎選擇與驗證。

六、方法與限制

本研究在資料與方法上仍有若干限制，須於解讀結論時一併考量。在預測層面，未來人口以最近十年迴歸並輔以 CAGR／移動平均回測外推，且 2030 年床位缺口係在總人口維持 2025 年水準的保守假設下推得，主要驅動因子為老年比率的上升，故推估值宜視為情境參考而非精確定值。在指標層面，核定床數僅代表機構住宿式供給，未涵蓋居家與社區式服務量，僅能作為機構式長照供給的代理指標；長照服務涵蓋率之分母為戶籍失能推估、分子則依實際服務地計算，跨縣市服務的流動會使部分縣市涵蓋率超過百分之百；身心障礙人數為「領證」人數，與「失能」在定義上並不完全相同，僅能作為需求面的輔助參考。在模型層面，機器學習僅有二十二個縣市樣本，相關結論宜視為方法示範，不宜過度外推。以下就本研究對課程技術主題的取捨範圍，進一步補充說明。

6.1 為何未採用 Hadoop / MapReduce / HDFS

本研究資料為全國二十二個直轄市與縣市之結構化主表（數百列、數十個欄位，整體規模約數 MB），單機以 Python 之 pandas 即可於記憶體內完成清理、合併、指標計算與建模，毋須分散式儲存與運算。於此資料規模導入 Hadoop／HDFS／MapReduce 不僅無效能效益，反而增加環境複雜度與維運成本，故本研究採輕量、可再現的單機工作流，並以 GitHub Actions 達成定期自動化更新。惟若未來資料擴展至全國村里層級、逐月時間序列或個體層級之長照給付紀錄（達百萬列以上），則有必要重新評估導入 Spark／HDFS 等分散式架構。此為依資料規模所做的工程取捨，而非技術能力之缺漏。

6.2 資料庫技術範圍：關聯式與 NoSQL／時序資料庫

資料庫面向以關聯式資料庫 SQLite 搭配 SQL 完成資料表儲存、跨表 JOIN 與聚合查詢，已充分涵蓋關聯式資料庫與 SQL 的核心要求。至於 NoSQL 與時序資料庫，因本研究資料屬固定結構、欄位明確之小型主表，查詢需求以欄位聚合、排名與 JOIN 為主，關聯式模型即為最適選擇，故未另行導入文件型（NoSQL）或時序型資料庫，以避免不必要的技術堆疊與維運負擔。

6.3 半結構化資料與文字處理

衛福部長照服務涵蓋率與 A/B/C 據點資料原始格式為 PDF，屬半結構化資料；本研究以 pdfplumber 解析其表格並轉為結構化欄位，即為半結構化資料處理的具體實作。至於完整的文字探勘（如中文斷詞、情感分析等自然語言處理技術），因本研究以數值型供需指標為分析核心而非以文字為核心，故未予展開，列為後續可延伸的研究方向。

6.4 時間序列分析涵蓋說明

時間序列分析已於第三節第四小節與第四節第三小節完整涵蓋：對各縣市老化指數同時以線性迴歸、CAGR 與移動平均三種方法外推，並以遮蔽最後三年之回測（MAE）作為選模依據，結果為 CAGR（3.34）優於移動平均（3.77）優於線性（5.25），實證了直線外推會低估加速中的老化趨勢，亦支持本研究以區間方式呈現長期推估的作法。

七、AI 工具使用聲明

本研究之資料蒐集、清理、分析、建模與結論，均由作者親自設計、執行並驗證。開發過程中使用大型語言模型（ChatGPT、Claude）作為輔助工具，其用途限於程式碼撰寫協助與除錯、分析方法之討論與建議、技術文件之理解，以及書面報告之文字潤稿與版面排版協助。所有最終分析結果、數據與結論均非由大型語言模型直接產出，亦未以其代寫完整報告內容；完整的人機協作過程另詳見附錄之互動記錄。

參考文獻

- [1] 內政部戶政司（2025）。村里單一年齡人口（Dataset 77132）〔資料集〕。政府資料開放平臺。<https://data.gov.tw/dataset/77132>
- [2] 內政部戶政司（2010–2025）。各縣市老化指數／老年比率〔統計資料〕。中華民國內政部戶政司。<https://www.ris.gov.tw/app/portal/346>
- [3] 衛生福利部統計處（2025）。身心障礙者人數按年齡及縣市別、按類別及縣市別分〔統計資料〕。<https://dep.mohw.gov.tw/dos/cp-5224-62359-113.html>
- [4] 衛生福利部社會及家庭署（2025）。全國老人福利機構（Dataset 8572）〔資料集〕。政府資料開放平臺。<https://data.gov.tw/dataset/8572>
- [5] 衛生福利部長照司（2025）。長照服務涵蓋率（表三）與各縣市長照資源布建A／B／C（表十六）〔統計資料〕。<https://1966.gov.tw/LTC/lp-6485-207.html>
- [6] 內政部（2020）。人口單一年齡組（Dataset 14226）〔資料集〕。政府資料開放平臺。<https://data.gov.tw/dataset/14226>
- [7] 國家發展委員會（2024）。中華民國人口推估（Dataset 6101）〔資料集〕。政府資料開放平臺。<https://data.gov.tw/dataset/6101>
- [8] g0v 零時政府（2016）。twgeojson：臺灣縣市界地理資料〔GitHub 儲存庫〕。<https://github.com/g0v/twgeojson>