

醫療可及性是否左右房價？機器學習之迴歸樹及隨機森林模型的應用

彭蒂菁 **

摘要

台灣邁入高齡社會，面對醫療資源區域不均，住居周遭有無充足醫療資源，對長者尤其重要，可能因而左右住居選擇。本研究透過 Luo and Qi (2009) 提出的進階兩階段流動搜尋法，衡量長者之三類醫療資源可及性（醫療需求者與供給者的空間隔絕程度）：醫師、病床與救護車。再透過機器學習之決策樹以及隨機森林迴歸，在考量住宅與其他因素下，探討長者醫療可及性對大台北地區房價的影響。

隨機森林在整體表現上優於決策樹迴歸。決策樹與隨機森林皆發現：長者之三類醫療可及性因素對大台北地區房價有顯著的影響力。由於取代台北市與新北市的平均壽命長期居各縣市之首，長者主要需要的是慢性疾病的醫療資源，相較於緊急醫療所需的救護車，醫生以及病床的可及性更為重要。這些實證結果應能提供相關單位參考，讓醫療資源達到地理區位分配最適，極大化居民生活環境品質。

關鍵詞：醫療可及性、年長者、決策樹、隨機森林

JEL 分類代號：I11、O30、R21、R30

** 國立臺北大學不動產與城鄉環境學系副教授，電話：(02)8674-1111#67309，Email：tcpeng@mail.ntpu.edu.tw。作者由衷感謝匿名評審們的指正與建議。

DOI：10.3966/054696002021060109004

醫療可及性是否左右房價？機器學習之迴歸樹及隨機森林模型的應用

彭蒂菁

壹、前言

人口老化是現今各國需嚴肅面對的社會議題。根據聯合國的統計資料，全球的老年人口（定義為 65 歲以上）在 2009 年為 5.1 億人數，於 2019 年就增至 7 億，預計在 2050 年全球老年人口將攀升至 15 億老年人口（約佔全世界人口的 16%）。根據世界衛生組織的定義，一個國家的老年人口佔總人口數比例達 7%，定義為高齡化社會（aging society）、達 14% 稱為高齡社會（aged society）、達 20% 則稱為超高齡社會（hyper-aged society）。以亞洲地區而言，深受老化趨勢影響的日本，依序在 1970 年、1995 年與 2006 年，邁入高齡化社會、高齡社會以及超高齡社會。

近二十年來台灣人口老化迅速。早在 1993 年，台灣已經邁入高齡化社會。根據行政院主計總處的資料，台灣於 2012 年的老年人人口比例高達 11.15%，距離高齡社會已經相差不遠。就各縣市而言，老化狀況之差異就更加明顯。以圖 1 為例，台北市老年人口於 2012 年就高達 13.04%，遠高於鄰近的新北市（8.97%）。

人口老化速度加快，帶動銀髮族各方面的需求，尤其是醫療方面。根據行政院衛生署中央健保局「100 年全民健康保險費率精算報告」之附表 6-3¹，歷年資料顯示，固然各年齡層仍有醫療需求（有些年齡層還有需求下滑趨勢），但高齡人口平均每年每人門診次

¹ 行政院衛生署中央健康保險局的「民國 100 年核定全民健康保險費率精算報告」。取自 <https://reurl.cc/V3dpdY>，檢索日期：2013/04/14。

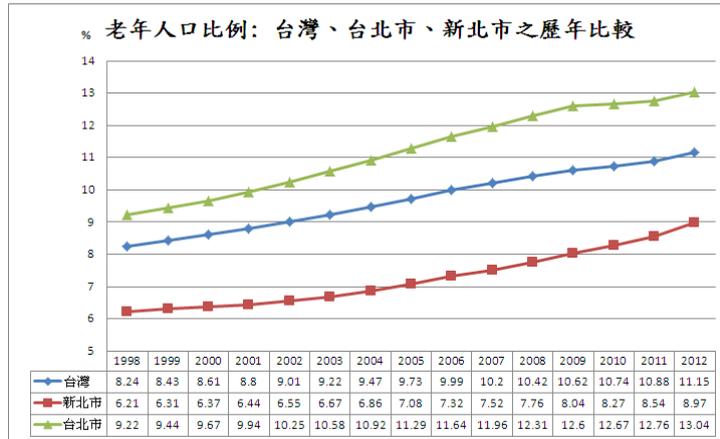


圖 1 老年人口比例：台灣、台北市、新北市之歷年比較

數約為總體人口的兩倍，而且逐年增長；在 2010 年，長者每年門診次數已高達 32.1 次，於十年間成長幅度約 11.75%。由於考量人口老化因素，估計未來 25 年西醫住診成長率甚將高達 2.04%。

隨著逐年醫療需求之增加，加上大型醫院周遭商圈往往有成熟生活機能以及便捷交通，爲了就醫方便，尤其是退休銀髮族，還有家中有長者的家庭，近年都開始傾向選擇醫院周邊的住處，因而帶動醫院附近的住宅房價。早在 2011 年，住商不動產企研室就比較過台北、新北市主要醫院附近之住宅行情；當時台大醫院、台北長庚、以及三軍總醫院附近商圈之住宅在短短一年之內的漲幅就高達 38.5%、36.7% 以及 31.0%；大台北多數醫院商圈房價都比行政區的平均房價來的高，像是台北榮總房價高於區域平均價格 26%。

就醫療供給面而言，根據中華民國醫師公會統計，台灣 368 個鄉鎮，有 6 成沒有急診醫師，3 成沒有婦產科醫師²。根台北市是台灣 22 個縣市中醫生供應最豐富的地區（每萬人執業醫師人口數爲 168），而人口數是全台之冠的新北市，病床供應卻很少（每萬人

² 立法院公報，2012，第 101 卷第 52 期上冊 (3993)，取自 https://lci.ly.gov.tw/LyLCEW/communique1/final/pdf//101/52/LCIDC01_1015201_00004.pdf，檢索日期：2015/01/16。

中僅 44 張床位)，甚至比台東 (68 張床位)、澎湖要低 (53 張)³。對於主要醫療使用者但行動不便的長者們來說，重要的並不是醫院所在位置，而是能否在地理區位上就近取得醫療資源。當地醫療資源供應的差異，將改變居民獲得其所需醫療資源的程度。不同類別的醫療資源，對有醫療需求的長者，也存在不同程度的重要性。醫師的多寡，可能決定日常生活慢性病處方籤或輕症病痛診斷需求的便利性；病床的多寡，可能決定就醫或較長期醫療的可能性；救護車的多寡，就能左右緊急病況就醫的救急性。

面對醫療需求增加，再加上近年醫療資源供給有區域分配不均，人們對於居住地區是否能取得醫療資源應更為重視。醫療可及性著重分析醫療資源提供者以及需求者之間的空間隔絕程度；例如，若是一個住戶所在地區有一家大型醫院，但是這家醫院卻嚴重缺乏醫護人員、病床甚至於救護車，如同有醫療資源可供應性卻沒有資源可取得性，其醫療空間可及性是相對低的。這類議題，在醫療文獻雖廣被討論，但在住宅領域中卻鮮少探討醫療可及性對當地房價的影響。在住宅價格文獻中，目前僅有 Peng and Chiang (2015) 將住宅至醫院之距離，納入住宅鄰里特徵，並透過分量迴歸探討醫院距離對大台北地區房價的影響。然而，台灣醫療在供給與需求上的區域差異，本研究採用 Luo and Qi (2009) 進階式兩階段流動搜尋法 (enhanced two-step floating catchment area method, E2SFCA)，在同時考量醫療供給及需求之下，根據各村里人口以及各鄉鎮的醫療資源，估計長者對三種醫療資源的可及性：醫師、病床以及救護車。根據 Rosen (1974) 的特徵價格理論，除了考量住宅特徵以及鄰里特徵之外，亦將醫療可及性視為住宅的鄰里特徵之一。或許有些讀者會認為，一般民眾對於所居地區醫療資源的確切數字可能不清楚。但對於日常生活中有醫療需求的長者而言，就能切身感受到自己所在地是否有足夠的醫療資源，像是掛門診看醫生是否容易，或是臨時叫救護車，能否儘快抵達等。以花蓮來說，主要的醫療資源都集中在花蓮市，以花蓮富里鄉而言 (該地以六十石山聞名)，在當地若要叫救護車送病患到醫院，有可能到等待約兩個小時。此醫療可及性的鄰里特徵，

³ 聯合新聞網，2018，「相同健保費，看病命運大不同：地圖看台灣醫療資源」，取自 <https://reurl.cc/kVAgvq>，檢索日期：2019/11/11。

自然反應到房價上。

有些讀者或許認為各年齡層皆重視醫療資源，因此「所有人口」（而非僅限「長者」）的醫療保健可及性，可能是評估住宅價格的更好指標。然而，在台灣和其他面臨老化議題的國家，相較其他年齡層，高齡人口增加幅度大而且醫療需求比例快速增加。以台灣而言，在 2001 至 2010 年間，當一般人（全部人口）每年平均門診次數持平約在 15 次左右，長者門診次數從 28.7 次增至 32.1 次，但有些年齡層的醫療需求反而有下降的趨勢（請參閱註腳 1）。青壯年人口本身較少有就醫需求，而且也對於自身健康的認知較為正向⁴，所以在購屋時，醫療可及性比較不會是他們主要的考量。再加上長者住宅自有率是所有年齡層最高⁵，因此高齡人口能否有適當的醫療可及性，理應對住房價值有一定的影響⁶。

長者醫療可及性對住宅價格的影響，是本研究的核心。住宅領域在研究方法上，主要採用計量迴歸分析，但難免受限於相關的資料與模型假設，經濟學家 Varian (2014a) 指出，大數據領域的機器學習模型在這方面較為彈性，並沒有預先假設應變數（例如：房價）與自變數（例如：住宅屬性）之間的關係，讓研究者得以探討更為複雜多元的數據，而且不受到統計問題的影響，因此有可能發掘在住宅議題上尚未觀察到的現象。Fan et al. (2006)、Grömping (2009) 皆應用樹狀結構的機器學習演算法（亦即決策樹與隨機森林演算法）於房價的相關分析，因為這類演算法能夠處理高維度數據，並且透過估計各變數對模型架構的貢獻度來決定變數的重要性。

⁴ 根據衛生福利部國民健康署於 2012 年的調查，相較於 18-54 歲人口自評健康狀態良好的比率均超過 80%，僅有約 57% 的高齡人口有此認知。取自 <https://reurl.cc/E298xR>，檢索日期：2016/02/21。

⁵ 美國和英國在 2015 年時，高齡人口房屋自有率大約為 80%，日本在 2014 年為 87%。根據衛生福利部的統計，在台灣高達 95% 的長者居住在自有或家人的住宅。

⁶ 作者亦有採用機器學習模型分析「總人口」之醫療可及性對住宅價格的影響；該結果與本文的長者之醫療可及性的結果類似。對此有興趣的讀者，請聯絡作者。

根據大台北地區的實價登錄資料，本研究根據 Rosen (1974) 特徵價格理論，考慮住宅本身特徵、鄰里屬性以及三類醫療可及性程度，透過機器學習的迴歸樹以及隨機森林模型，探討醫療可及性對大台北地區 (台北市及新北市) 住宅價格的影響。此實證結果應能為後續政策上對醫療資源的空間安排，提供一定程度的參考。

貳、文獻回顧

一、房價之影響因素

住宅文獻主要根據 Rosen (1974) 的特徵價格法 (hedonic price model)，分析各類住宅特徵對房價的影響。此理論主張：消費者消費的並不是商品本身 (good itself)，而是商品具備的特徵 (characteristics)；商品是多項特徵的組合，因此商品本身價格也就由這些特徵共同決定。特徵價格理論廣泛的應用在各個領域，像是汽車安全效能 (Andersson, 2005)、酒類的品質像是高粱酒 (陳信字與楊育誠, 2019)、消費者線上購物傾向 (Overby and Lee, 2006) 等。

特徵價格理論更是廣為應用在探討住宅價格，主要是因為房屋是高度異質性商品，並具有耐久性、不可分割性以及不可移動性；因此，住宅所具備的特徵相對於其他商品來的複雜。Sirmans et al. (2005) 回顧了約 125 篇應用特徵價格法探討住宅價格的文獻，住宅價格的特徵價格模型主要以下式 (1) 呈現：

$$\text{房價} = f(\text{住宅特徵, 其他變數}) \quad (1)$$

住宅價格是該住宅本身特徵與其他因素的函數 (f 代表函數，function)。住宅本身特

徵，也就是住宅本身屬性，像是屋齡、建坪、樓層、衛浴設備數目等等，在住宅文獻中皆有討論 (Clapp and Giaccotto, 1998; Leishman, 2001; Sirmans et al., 2005)。由於房屋的不可移動性，因而影響住宅價格的其他因素，主要包含：住宅所在的地理位置，周遭的環境品質及鄰里設施。過往文獻發現，位於市中心的住宅價格較高 (相較於郊區)，主要是因為市中心提供工作便利性以及方便的生活機能 (Archer et al., 1996)；環境品質因素，則包含像是住宅周遭的綠意 (Panduro and Veie, 2013)，空氣品質 (Chay and Greenstone, 2005) 等。

文獻主要將鄰里設施分為兩類。迎毗設施 (yes-in-my-backyard, YIMBY) 的存在能提升住宅本身的使用價值以及交換價值，也為當地居住品質帶來正向外部性，因而吸引居民在選擇住宅時，希望在地理位置上與它更接近以提升房價；像是學區 (Machin, 2011)、公園 (Crompton, 2001)、教堂 (Carroll et al., 1996)、交通運輸設施 (洪得洋與林祖嘉, 1999) 等。鄰避設施 (not-in-my-backyard, NIMBY) 最初設置是為了提升當地居民福利，然其負外部性反而讓當地居民望之卻步，因此往往對當地房價產生負面影響，像是垃圾處理場 (Hite et al., 2001)、汙水處理廠 (Batalhone et al., 2002) 等。由於設施對房價的影響取決於其地理位置，因此文獻往往透過地理資訊軟體計算各住宅到此設施的直線距離，然後基於特徵價格理論，將此距離視為住宅特徵之一 (Ready and Abdalla, 2005)，或將距離區段化並用虛擬變數來估計不同距離範圍的影響 (彭建文等, 2009)，亦或是用虛擬變數來定義在一定範圍內有無此設施 (Landis et al., 1994)。

然而，面對現今快速高齡化的趨勢，住宅特徵價格文獻鮮少討論醫院這個最接近生活的設施。李泳龍等 (2009) 認為，醫院屬於半嫌惡設施 (semi-obnoxious facility)，也就是兼具迎毗 (例如：就醫便利性) 與鄰避特徵 (例如：人潮或救護車警笛聲)；該研究透過問卷探討成大醫院與鄰近住宅環境的關係，發現住戶對於半嫌惡設施的接受程度，受到自身居住地與半嫌惡設施距離遠近的影響。就實證分析而言，目前僅有 Peng and Chiang (2015) 透過分量迴歸探討醫院截段 (spline) 距離對大台北地區房價的影響。該研究發現，在醫院近距離範圍內 (0-500 公尺)，距離醫院愈遠，房價愈高；在此近距離範圍內，

醫院鄰避的影響，像是救護車警笛聲，遠大於迎毗的優勢。因此，相較於醫師或病床數這兩類醫療資源，救護車可能兼具迎毗（例如：緊急醫療救護）以及鄰避（例如：警笛聲噪音）特徵。

本研究並沒有採用常規距離測量來衡量醫療需求（例如：長者）與醫療供給（例如：醫療資源）之間的空間分散程度，而是選擇採用 E2SFCA 方法來測量長者對三類醫療資源（醫師、病床與救護車）的空間可及性。主要是因為這方法具有技術上的優勢（請見後續內容）；再者，醫療需求者不一定會去距離最近的醫療資源，也有可能選擇附近甚至於跨區的醫療資源；第三，醫療資源分配在各醫院之間往往存在不均的現象（例如，某些醫院可能擁有更多的病床或醫生）；第四，不同醫療資源的可及性程度對當地居民有不同程度的意義，因此對該地房價應有不同程度的影響。例如：醫生的可及性對各年齡層居民都至關重要，但病床或救護車的可及性可能對長者更為重要。

如何衡量這些住宅特徵或鄰里特徵對房價的影響，則是另一個議題。不同的房價文獻，所考量的住宅特徵往往相異；像是有的文獻有考量住宅是否有冷氣，或考量住宅是否有湖景，亦或是納入鄰里的人口結構等等；雖然納入更多元的特徵變數更能解釋房價的變化，但實證分析往往受限於能取得的二手資料，因此過往房價文獻才會有在特徵價格模型中採用不同變數的現象（Sirmans et al., 2005）。在實證研究上，主要採用不同計量迴歸模型，探討上述各項特徵對房價的邊際影響。傳統計量迴歸架構在許多模型假設條件之上（像是變數的常態分配等），但住宅價格本身往往具有非線性特色，而且影響因素眾多。誠如 Varian (2014b) 指出，資工領域的機器學習演算法，主要是無母數分析，對於模型架構或是資料假設較為彈性，因此能更有效率地處理資料之間複雜的關係，在處理高維度以及非線性的資料上有較強的計算優勢。本研究希望採用決策樹與隨機森林演算法建構模型，以探討三類長者醫療可及性、住宅特徵與大台北地區房價之間的關係。

二、醫療可及性：進階式二階段流動搜尋法

醫療可及性在醫療地理領域 (healthcare geography) 是一項很重要的議題：如何確保有限醫療資源，在一定空間範圍內，能有效且均等讓醫療所需者受益。醫療可及性若僅用空間距離衡量 (例如：各住宅至最近醫院的距離)，容易失真。在老化趨勢下，長者醫療需求日增，但醫療供給區域分配不均，因此地方醫療供給資源的多寡 (像是醫生、病床數目等等)，也會是居民 (尤其是銀髮族) 選擇購屋的重要考量。即便住宅位於醫院僅 5 分鐘的距離，但若醫院本身醫療資源嚴重不足，此住宅位置的醫療可及性相對低，對房價的影響自然減弱。換言之，醫療可及性除了考量醫療供給面之外，也須考量醫療需求面。如果供需不均 (供過於求或是供不應求)，醫療可及程度自然較低。由於 GIS 系統的逐漸演進，醫療可及性的估計方法也不斷更新。本研究採用 Luo and Qi (2009) 提出的進階式二階段流動搜尋法 (E2SFCA)，分析大台北地區長者的三類醫療可及性。首先簡介過往空間可及性的衡量方法，再說明進階式二階流動搜尋法。

(一)空間可及性衡量方法

在地理區域分析的相關領域，可及性 (accessibility) 泛指從任一定點能到達服務據點 (像是工作地點、公共設施等) 的難易程度。因此，可及性需要同時考量有限資源的供給與需求，兩者如何在地理空間中取得平衡。醫療資源可及性更是廣泛討論的議題 (Love and Lindquist, 1995; Guagliardo, 2004; Wang and Luo, 2005)。可及性主要分成兩大領域：潛在 (potential) 與實際 (revealed) 可及性以及空間 (spatial) 與非空間 (aspatial) 可及性 (Wang, 2006)。以醫療為例，實際可及性 (revealed accessibility) 著重分析需求者能獲得的實際醫療資源；潛在可及性 (potential accessibility) 分析需求者能獲得的潛在醫療資源。由於需求者實際上用了多少醫療資源之調查資料往往難以取得，所以從政策計畫的角度，主要是分析現有醫療系統的潛能服務多少地方群眾。空間可及性 (spatial accessibility) 強調地理空間位置對醫療供給與需求所造成的隔離程度；非空間可及性

(aspatial accessibility) 探討非空間因素 (像是種族、性別等社會經濟因素) 造成某類群眾無法取得醫療資源 (Joseph and Phillips, 1984)。換言之，醫療資源可及性需同時考量醫療供給面及需求面在空間上的隔離程度，而不是單獨考慮其中一方。本研究將著重分析潛在空間可及性 (potential spatial accessibility)，透過 Luo and Qi (2009) 提出的進階式二階段流動搜尋法 (E2SFCA) 以衡量大台北地區各村里的長者 (醫療需求面) 對於三類醫療資源 (供給面) 之空間可及性，並將其視為實價登錄各住宅樣本的醫療可及性特徵，以探究醫療可及性以及其他住宅特徵是否會左右大台北地區的房價。

在實務上該如何計算醫療空間可及性，則是另一項議題。最直覺的方式，就是採用距離概念來計算醫療可及性，像是 Lovett et al. (2002) 採用開車移動距離所花的時間，衡量英國在地的醫療可及性。另一個方式就是以各行政區為單位，衡量該區的醫療提供者與人口比例，這又稱為供給率 (supply ratio) (Campbell et al., 2003)，也就是以各行政區內醫療提供者 (例如：醫師人數) 為分子除以醫療使用者 (例如：長者) 為分母，此數值表示該地理範圍內的人口所共享的醫療資源多寡程度 (例如：每一萬人中有 10 位醫生)。但這忽略了空間距離在現實中所造成的隔閡，畢竟人們就醫意願可能隨著與醫療資源距離的增加而減少，而且此方式也忽視了人們跨區就醫的可能性 (Guagliardo, 2004)。

解決跨區就醫的進階方法，就是流動搜尋法 (floating catchment area, FCA)，以及後續延伸的二階流動搜尋法 (two-step floating catchment area) 與進階式二階流動搜尋法 (enhanced two-step floating catchment area)；這些方法改進了空間隔離性 (spatial dispersion) 的衡量方式。流動搜尋法 (FCA) 最初應用在職缺空間可及性 (Peng, 1997)，主要著重從單一角度 (供給面或需求面) 來衡量，但已不再侷限在行政區域。例如：以醫療資源為中心，向外延伸 20 公里，以此空間範圍內的人口資料，計算供給率。但是，此方法僅關注特定地理區域內的醫療保健供需。

為了處理這個限制，Radke and Mu (2000) 提出了一種空間分解方法 (spatial decomposition)，也就是將欲分析的空間區域重疊切割；根據此概念，Luo and Wang (2003) 認為應兼顧供給與需求以衡量空間可及性，因而發展出兩階段流動搜尋法 (2SFCA) 應用

於芝加哥地區醫療保健的空間可及性。簡言之，此方法分兩階段計算醫療需求者的醫療空間可及性；第一階段從醫療供給面的角度，搜尋其合理的服務範圍 (catchment area：此範圍可用距離或移動時間衡量) 內所涵蓋的醫療需求人數，進而計算出醫療資源數與需求者的比例 (例如：醫師與長者的比例)；第二階段從醫療需求面的角度，以醫療需求者的人口點為中心，搜尋在合理空間範圍內所涵蓋的醫療資源提供者，並將這些服務提供者在第一階段算出的比例加總；最終數值就是代表該區域的醫療可及性。此方法廣泛應用於許多研究，像是 Wang and Luo (2005) 探討美國伊利諾州的初級醫療的空間可及性或是 Fransen et al. (2015) 分析比利時日照中心空間可及性等等。

(二)進階式二階流動搜尋法

原本 Luo and Wang (2003) 提出的二階流動搜尋 (2SFCA)，主要是假設在各醫院可及範圍內，醫療可及性皆是相同，並不會隨著距離而遞減。為了考量可及範圍內距離遞減 (distance decay) 的因素，Luo and Qi (2009) 提出進階版的二階流動搜尋法 (E2SFCA)：在各個醫療可及區域內，根據移動的距離 (或時間) 進一步分隔出數個醫療可及性範圍；在不同的移動範圍內給予權重 (weights)，以反應醫療可及性隨著距離遞減的現象。與 2SFCA 方法類似，E2SFCA 也是採用兩階段計算，但各階段的計算皆須考量距離權重。根據此方法，Kanuganti et al. (2016) 探討印度地區初級醫療的空間可及性，Kim et al. (2018) 分析南韓首爾人民的醫師資源可及性，Wan et al. (2012) 則分析美國德州大腸癌治療服務中心的可及性，或是廖興中 (2013) 估計台灣小兒科醫療資源在鄉鎮層級的空間可及性。此方法的相關理論與實務操作，請見第參章節詳細說明。

三、機器學習：決策樹迴歸與隨機森林迴歸

機器學習的目的是設計可自動學習的演算法；換句話說，主要是讓機器 (例如：電腦) 根據演算法從訓練資料中學習規則並根據該規則建構模型，再透過測試資料集進行

預測 (Kelleher et al., 2015)。簡言之，機器學習式建立在「資料輸入透過函數模型得到預期輸出」這個基礎上。

機器學習的相關演算法，在許多領域廣為應用，像是資工、工業生產、醫療等領域；在商管領域亦是廣泛應用於各項議題，像是會計 (Li, 2010)、金融市場相關商品或價格之變動 (Wang et al., 2014) 等。值得注意的是，在以計量分析為主的經濟領域，也開始採用各種機器學習演算法探討不同議題，像是經濟成長 (Basuchoudhary et al., 2017)、能源價格預測 (Ghoddusi et al., 2019)、農業管理 (Storm et al., 2020) 或是與人力資源相關的教育議題 (Sansone, 2019) 等。Varian (2014a) 指出機器學習演算法與傳統迴歸的差異，尤其強調機器學習演算法的運算彈性，應用於經濟領域有其優勢。Varian (2014b)、Basuchoudhary et al. (2017) 或是 Mullainathan and Spiess (2017) 對於機器學習與經濟議題有相當完整的文獻回顧。

在住宅價格領域，機器學習演算法也有一些相關應用。Park and Bae (2015) 應用天真貝氏演算法 (Naïve Bayesian) 分析美國維多利亞州的獨棟住宅價格；Mimis et al. (2013) 則採用人工神經網路演算法 (artificial neural network) 探討希臘雅典的住宅價格；Plakandaras et al. (2015) 應用支援向量機器演算法 (support vector machine) 預測美國房價變動。Pérez-Rave et al. (2019) 指出，機器學習與特徵價格理論的結合應用讓房價相關分析更為多元與彈性。

另外兩個常見的機器學習演算法則是決策樹 (decision tree) 與隨機森林 (random forest)。Fan et al. (2006) 採用決策樹演算法探討新加坡的公共住宅價格變化，Pérez-Rave et al. (2019) 透過決策樹估計美國哥倫比亞地區的房價趨勢；Antipov and Pokryshevskaya (2012) 則採用隨機森林演算法分析公寓價格，Yoo et al. (2012) 透過隨機森林迴歸，了解影響美國紐約地區房價的因素。Varian (2014a, 2014b) 指出，相較於傳統計量模型，決策樹與隨機森林演算法有較強的運算優勢，能處理較高維度、非線性或是遺漏值的資料。

決策樹與隨機森林演算法屬於監督式學習，可用於資料分類 (目標變數為類別資料) 或是迴歸預測 (目標變數為連續資料)。決策樹在操作上，會將取得之資料分為用以建構

決策樹的訓練資料以及判定決策樹預測能力優劣的測試資料。首先決策樹根據取得資料的屬性來區分類別，建立從上至下的樹狀式分層結構，從中歸納、擷取規則，了解不同層級的資料屬性對目標變數的影響；再根據決策樹推導出的規則，進行對未知樣本的預測，以衡量此決策樹分類的表現以及穩健性 (Berry and Linoff, 2011)。

James et al. (2013) 指出，決策樹類似人們做決策的流程，結果以樹狀結構圖示呈現，解釋性高；決策樹模型能處理非線性或複雜度較高的資料，具有變數挑選與遺失值填補的機制，是一個廣泛應用的模型。然而單一決策樹模型有較高的不穩健性，因為決策樹架構很容易受到抽樣資料的影響，也就是樣本有些許變化，估計結果也就隨之變動。集合式學習 (ensemble algorithm) 結合數個基本模型的預測結果，更能提高模型穩健性。隨機森林就是集合演算法之一種；相較於單獨一棵決策樹的表現，隨機森林是眾多決策樹的集合，合併所有樹的預測結果，降低單一決策樹的變異度或不穩定性，避免過度配適 (overfitting) 問題，取得更精確並穩定的結論。Varian (2014a) 指出，隨機森林的優勢，讓它成爲是現今最爲廣泛使用的演算法之一。因此，本研究以決策樹迴歸爲基準模型 (benchmark model)，進一步採用隨機森林迴歸，以比較兩個模型對於醫療可及性與房價之間的預測，是否存在差異。若是兩者能呈現一致性結果，則能佐證醫療可及性對房價的影響力。

參、研究方法

一、進階式二階流動搜尋法

(一)資料來源與基本設定

本研究採用進階式二階流動搜尋法 (E2SFCA)，計算大台北地區中，各村里長者 (醫療需求者) 對各鄉鎮市區⁷三類醫療資源供給 – 醫師、病床與救護車 – 的空間可及性。在台灣的相關研究中，章殷超等 (2011) 採用二階流動搜尋法 (2SFCA) 衡量台灣各鄉鎮每萬人口西醫師數之差異，廖興中 (2013) 則採用進階式二階流動搜尋法 (E2SFCA) 衡量台灣小兒科醫療資源在鄉鎮層級空間可及性，也就是以鄉鎮為醫療供給單位，以村里為醫療需求單位；本研究亦是如此設定。

就資料來說，台北市有 12 個市區以及 456 個村里，新北市有 29 個市區以及 1032 個村里。本研究採用的醫療供給面資料取自於不同來源；醫師數資料來自於中華民國醫師公會全國聯合會公布的「2012 年台灣地區執業醫師、醫療機構統計」中的「各鄉鎮市區執業醫師數」，這筆資料並非以醫院為單位，而是統整列出各鄉鎮市區有多少執業醫師⁸；病床與救護車資料則來自於衛生福利部公布之臺北市、新北市的「醫療院所數及各類病床數、救護車輛數按地區別分」⁹，也就是列出各鄉鎮市區各別的病床數與救護車輛數。醫療需求面則是採用國土資訊系統於網路上的公開資料，台北市與新北市各村里五歲年齡組人口統計，進而計算各村里的 65 歲以上人口。

在說明研究方法前，需先說明本研究的基本設定。本研究在 E2SFCA 的計算過程中，醫療需求者為長者，醫療供給來源為醫師、病床與救護車數。由於需要界定各醫療可及區域內 (catchment area)，根據移動的距離所隔出的數個醫療可及性範圍。本研究參照章

⁷ 鄉鎮市區為行政區位階相同或相似的行政區劃名詞。新北市與台北市是直轄市，在市之後的單位為區；市區等同於其他縣的鄉鎮。鄉鎮市區之後的行政單位為村里。

⁸ 資料來源：中華民國醫師公會全國聯合會，2012，取自於 https://www.tma.tw/stats/index_AllPDF.asp，檢索日期：2013/04/22。

⁹ 台北市資料來自「表四十五臺北市醫療院所數及各類病床數、救護車輛數按地區別分」，取自 <https://reurl.cc/Q7vznM>；新北市資料來自「表五十二新北市醫療院所數及各類病床數、救護車輛數按地區別分」，取自 <https://reurl.cc/ra7dpr>，檢索日期：2013/04/28。

殷超等 (2011) 與廖興中 (2013) 這兩篇在台灣相關空間可及性衡量的設定值如下：區分三個移動範圍 (0-5 公里、5-10 公里以及 10-15 公里) 以及三個權重 (1、0.5、0.25)。

本研究在 E2SFCA 的實務操作流程上，主要是參考廖興中 (2013) 的作法，採用 ArcGIS 地理資訊軟體，進行資料前置作業。簡言之，先用軟體中的「資料設置」(data management) 設定中的「屬性至點位」選項 (feature to point) 建立各村里幾何中心點，代表醫療資源需求者 (長者) 的位置；再用「空間統計工具」(spatial statistics tool) 中的「中心點平均」(mean center) 根據各市區的村里地理幾何中心點，計算村里人口密度加權以定位各市區的人口中心點，此市區人口中心點代表醫療資源提供者的位置。後續再以「起迄分布矩陣」(origin-destination (OD) matrix) 計算各村里中心點 (需求面) 至各市區中心點 (供給面) 之間的最短行車距離；再採用「屬性選擇」(select by attributes) 將前一步驟計算出的距離結果中，先挑出小於 15 公里的組合，然後再將這些組合區分為三個範圍：0-5 公里、5-10 公里以及 10-15 公里。後續進階式二階流動搜尋法的相關說明與操作，主要參照 Luo and Qi (2009) 及廖興中 (2013)。

(二)理論說明與實務操作

1.第一階段：供給面

(1)理論說明：

根據每個醫療資源「供給」據點 j (例：醫師人數)，決定 j 的移動距離門檻值 (例如：15 公里) 以定義其資源提供範圍。在各範圍內，計算三段移動範圍區 (zones)，例如：0-5 公里、5-10 公里、10-15 公里 (zones 1-3)。在 j 的各三個時段移動範圍內，搜尋各個時段醫療資源的「需求」據點 k ，然後計算此醫療據點之加權供給需求比 R_j (weighted supply-to-demand ratio)：

$$R_j = \frac{S_j}{\sum_{k \in \{d_{kj} \leq D_r\}} P^k W_r} = \frac{S_j}{\sum_{k \in \{d_{kj} \leq D_1\}} P^k W_1 + \sum_{k \in \{d_{kj} \leq D_2\}} P^k W_2 + \sum_{k \in \{d_{kj} \leq D_3\}} P^k W_3} \quad (2)$$

S_j : 醫療供給據點 j 的醫療供給程度 (capacity)

d_{kj} : k 與 j 之間的移動距離

D_r : 第 r 個移動範圍區 (zone), $r = 1 - 3$

P_k : 位於醫療供給據點 j 的資源提供範圍內的 k 個醫療「需求」據點

W_r : 各移動範圍區的權重。用以衡量醫療可及性隨距離遞減。

以圖 2 為例，其中有 15 個鄉鎮，但只有鄉鎮編號 7 有家醫院 A；以醫院 A 為中心（假設只有一位醫師）向外劃分三個距離移動範圍，第一個距離移動範圍 D1 (0-3 分) 涵蓋一個居民 7 (鄉鎮 7 的居民聚集點)，第二個時間移動範圍 D2 (3-5 分) 涵蓋 2 個居民 (2、3)，第三個時間移動範圍 D3 (5-10 分) 涵蓋 4 個居民 (1、4、6、10)。假設距離權重隨著區間遞減 1/2，A 醫院的醫療供給資源 = $1(1) + 2(1/2) + 4(1/4) = 3$ 。A 醫院雖然位於鄉鎮編號 7，但其服務不侷限於該行政區範圍；A 醫院的醫師跨區服務到三個居民。

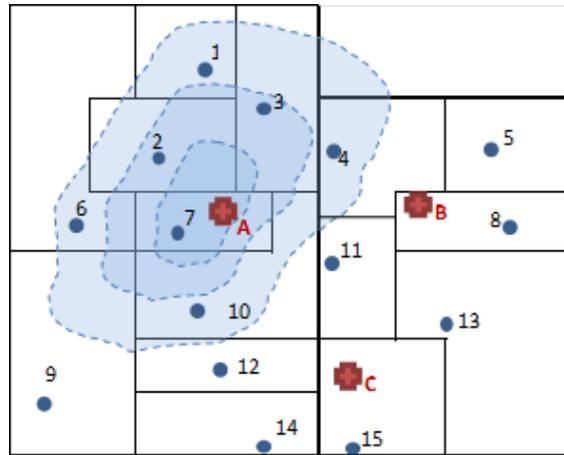


圖 2 進階式二階段流動搜尋法-第一階段

(2)實務操作簡說

在此列出新北市淡水區的資料為例，見表 1。在前置作業階段已經將該市區的地理中心點（以此為醫療資源中心點）向外區分為三個醫療服務範圍（沒有局限於行政區範圍）：0-5 公里涵蓋 10,018 位長者，5-10 公里涵蓋 11,013 位，10-15 公里涵蓋 78,483 位；由於各範圍權重依序是 1, 0.5 與 0.25，因此村里人口權重（65 歲以上）為 35145.25（也就是(10018*1) + (11013*0.5) + (78483*0.25)）。由於淡水區的醫生數為 367 位，因此淡水區的鄉鎮醫生與人口比（65 歲以上）為 0.01044（也就是 367/35145.25）；鄉鎮病床與人口比與鄉鎮救護車與人口比也是如此計算。

表 1 進階式二階段流動搜尋法 第一階段（實際資料範例）

市區	CITY	距離範圍	村里人口 _65 歲以上	村里人口權 重_65	市區醫 生數	市區病 床數	市區救 護車數	市區醫生 與人口比 _65	市區病床 數與人口 比_65	市區救護 車數與人 口比_65
		0-5km	10018							
淡水區	新北市	5-10km	11013	35145.25	367	1108	7	0.01044	0.03153	0.0002
		10-15km	78483							

2.第二階段：需求面

(1)理論說明

從每個居民聚集點 i ，在特定移動範圍尋找此範圍內的所有醫療供給據點 j 。然後加總這些供給據點的供給需求比。

$$A_i^F = \sum_{j \in \{d_{ij} \leq D_r\}} R_j W_r = \sum_{j \in \{d_{ij} \leq D_1\}} R_j W_1 + \sum_{j \in \{d_{ij} \leq D_2\}} R_j W_2 + \sum_{j \in \{d_{ij} \leq D_3\}} R_j W_3 \quad (3)$$

A_i^F ：代表需求據點 i 的醫療可及性

R_j ：需求據點 i 之時間移動範圍內的各供給據點之供給需求比（即：第一階段計算

的結果)

D_r ：第 r 個移動範圍區 (zone)， $r = 1 - 3$ (依序為 0-5 公里、5-10 公里、10-15 公里)。

d_{ij} ：需求據點 i 與供給據點 j 之間的距離

W_r ：各個時段移動範圍區的權重。用以衡量醫療可及性隨距離遞減。亦是依序設定為 1、0.5 與 0.25。

以圖 3 為例。在上述第一階段算出 A 醫院 (一位醫生) 能服務 3 個居民 (據點)；意即，每個居民的醫療可及性是 1/3 (在這範圍內，每個居民能獲得 1/3 的醫院服務)。假設在第一階段算出 B 醫院以及 C 醫院也各別能服務 3 個居民據點，而且距離範圍權重亦是 (1、0.5、0.25)，居民據點 7 的醫療可及性 = $((1/3)*1) + ((1/3)*0.5) + ((1/3)*0.25) = 7/12$ 。換言之，在居民據點 7 向外擴展 15 公里的範圍內，每 12 位居民能獲得 7 位醫師的醫療服務 (或是需求據點 7 在 15 公里範圍內平均每人能獲得約 0.58 的醫療資源)。

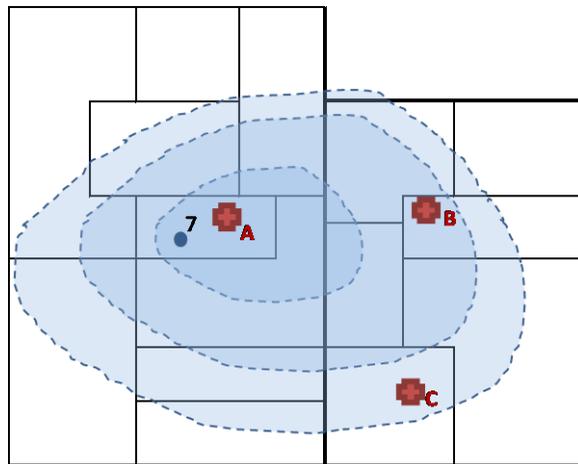


圖 3 進階式二階段流動搜尋法 - 第二階段

(2)實務操作簡說

由於新北市淡水區涵蓋多個村里，在此以民權里為例，請見表 2。以民權里為中心，向外延伸三個區域範圍；以醫生資源來說，其中 0-5 公里沒有涵蓋到任何市區的醫生資源（市區醫生與人口比），但 5-10 公里與 10-15 公里各涵蓋市區醫療資源數值為 0.0266 與 0.0061；因此，民權里的村里醫生與人口比為 0.0148（也就是： $(0*1) + (0.0266*0.5) + (0.0061*0.25)$ ）。換言之，在新北市淡水區民權里，平均每位長者能享受到 0.0148 醫生的醫療資源（或是平均每 100 位長者能享受到 1.48 醫生的醫療資源）。

本研究根據需求面的各村里老年人口數以及供給面的各市區三類醫療資源（醫師人數、病床數、救護車數），採用進階式二階段流動搜尋法計算大台北地區各村里的三類醫療可及性：長者的醫師可及性、長者的病床數可及性以及長者的救護車可及性。再透過地理資訊系統軟體（ArcGIS）圖層，確認各住宅所在地點的醫療可及性，作為該住宅的醫療特徵。由上述說明可知，位於同一空間（村里）的住宅，會具有相同的長者醫療可及性。但請注意，目前醫療資源統計的最小單位是到鄉鎮市區，本研究已採用最精細的資料衡量長者之醫療可及性。

表 2 進階式二階段流動搜尋法 – 第二階段（實際資料範例）

村里	市區	CITY	距離範圍	市區醫 生與人 口比 _65	市區病 床與人 口比_65	市區救 護車數 與人口 比_65	村里醫 生與人 口比_65	村里病 床與人 口比_65	村里救護 車數與人 口比_65
			0-5km						
民權里	淡水區	新北市	5-10km	0.0266	0.0701	0.0003	0.0148	0.0393	0.0002
			10-15km	0.0061	0.0168	0.0001			

二、決策樹與隨機森林迴歸

(一)理論概說

決策樹的演算法有許多種，彼此的差別在於分枝準則以及適用的資料類別。本研究希望探討住宅特性及醫療可及性（屬性變數）對房價（連續數值之目標變數）的影響，因此採用 Breiman et al. (1984) 提出的分類與迴歸樹（classification and regression tree, CART），因為此演算法能處理類別以及連續資料。

統計計量的迴歸（regression）與機器學習的決策樹迴歸（regression tree），皆以預測為目的。根據應變數（目標變數）的尺度，計量迴歸可簡單區分為線性迴歸（應變數為連續資料），以及羅吉特迴歸（應變數為類別資料）；決策樹亦可區分為分類樹（classification tree），也就是目標變數為類別資料（例如：考試成績為及格、不及格），以及迴歸樹（決策樹迴歸；regression tree）：也就是目標變數為連續資料（例如：住宅售價）。分類樹用來尋找「能區分目標資料的類別」的屬性變數，迴歸樹則是用來尋找「能區分目標資料之連續數值的相近度」的屬性變數。因此本研究採用決策樹迴歸，以預測住宅特徵以及醫療可及性（屬性變數）對大台北地區房價（目標變數）的影響。

統計中的迴歸，可說是全域性模型（global model），將迴歸模型是套用在所有資料空間上；相對的，Breiman et al. (1984) 提出以樹狀結構遞迴分割樣本的方法，將原本複雜的資料空間，簡化成較為單純的資料子集合（partition），然後在每個子集合中，利用迴歸方法建立決策樹迴歸模型。根據 James et al. (2013)，線性迴歸以下式呈現（ x 代表屬性變數， β 代表迴歸係數）：

$$f(X) = \beta_0 + \sum_{j=1}^p x_j \beta_j \quad (4)$$

決策樹迴歸以式 (5) 呈現，其中 R_1, \dots, R_M 代表各屬性空間的分支：

$$f(X) = \sum_{m=1}^M c_m \cdot I(x \in R_m) \quad (5)$$

從式 (4) 及 (5) 可見，線性迴歸能解釋目標函數與屬性函數之間的線性關係；對於兩者間存在高度非線性或相對複雜的關係，決策樹迴歸會有更佳的解釋力。

根據 James et al. (2013) 的說明，基於遞迴二元分割 (recursive binary partitioning) 的原則，決策樹是從上至下的倒置樹狀架構 (inverted top-down tree-like structure)，所有觀察點從根結點 (root node) 開始，然後根據分支法則 (rules) 切割出二元空間的內結點 (internal node)，持續分割，直到所有資料都進入最終的葉節點 (leaf node)，就完成決策樹。當決策樹建立好後，就可以將根部到葉節點的每一套獨特路徑，做為資料分類規則的決策。決策樹是個很貪心 (greedy) 的演算法，因為在一開始演算過程中，會考量所有屬性變數以及各屬性切割出的二元空間切割點 (cut point)，從大量可能性中選擇出能讓決策樹有最小殘差平方和 (residual sum of squares, RSS) 的屬性變數以及切割點，作為分支法則 (如下式，其中 R 代表切割空間，共有 1 至 J 個空間； \hat{y}_R 則是在第 j 個空間中的觀察數值的平均值)。分支過程會反覆在各切分後的次群聚中執行，直到達到停止切割的條件。

$$\text{minimize RSS} = \text{minimize} \sum_{j=1}^J \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2 \quad (6)$$

上述過程會建構出分支非常複雜的決策樹，固然對訓練資料有很好的預測表現，但極有可能對資料過度配適，也就是對測試資料的預測能力大幅降低。因應這個問題的重要策略就是剪枝 (pruning)。換言之，建構決策樹的過程如下：先建構一顆完整且最複雜的決策樹，並加入樹的複雜度作為目標函數的懲罰因子 (複雜度參數；complexity parameter，式 (7) 中的 α)，讓加入懲罰因子的最小殘差平方和的數值最小，以此修剪出最適的決策樹模型，如下式 (7) (T 表示決策樹葉節點的數量)。主要是透過交叉驗證法計算每個數值的交叉驗證誤差 (cross-validated error, X-val error)，以確認能使交叉驗證誤差最小的 α 最適值，進而建立最適的決策樹結構。

$$\text{minimize}(\text{RSS} + \alpha|T|) \quad (7)$$

決策樹的分枝準則決定樹的寬度及深度；Breiman et al. (1984) 的 CART 演算法，根據各個屬性在分枝上的不純度，也就是各屬性降低迴歸樹殘差的能力，決定各屬性在建構決策樹的重要性；不純度愈小表示該屬性愈適合作為分枝屬性；愈是重要的屬性，距離根結點愈近。當決策樹建立好後，就可以將根部到葉節點的每一套獨特路徑，做為資料分類規則的決策。迴歸樹的詳細理論解釋，請參考 James et al. (2013)；相關分枝準則的細節請見 Bühlmann et al. (2016)。在實務應用上，決策樹的優點在於結果容易解釋、將結果視覺化呈現，而且能處理高維度或非線性的資料；但是此演算法較易有過度配適的狀況，而且決策樹的結果較不穩健，很容易受到抽樣資料的影響；換言之，只要樣本資料些許改變，就有可能大幅更動決策樹的結果 (James et al., 2013)。

基於決策樹的優點，若能結合眾多決策樹的表現，就能突破單一決策樹的限制，這類的演算法統稱為集合式學習，隨機森林就是此類演算法。隨機森林可視為多棵決策樹組成的森林，然後根據所有決策樹預測結果以確定整個隨機森林的結果。隨機森林演算法主要是由 Breiman (2001) 提出，隨機森林之所以隨機，在於其建構森林中各顆樹的抽樣過程，不僅隨機抽樣觀察點，也隨機抽樣屬性。隨機森林採用 Breiman (1996) 提出的裝袋法 (bagging) 抽樣觀察點以建構眾多決策樹；裝袋法以拔靴法 (bootstrapping) 為基礎，從一組訓練資料中重複隨機抽取多組不同樣本 (抽出放回)，採用不一樣的資料建構各棵樹並取得各輸出預期結果；在迴歸之下，隨機森林的輸出結果為眾多決策樹的預測值平均 (averaging)；若是分類，隨機森林的輸出結果則是採用多數投票法 (voting)，所以裝袋法又稱為引導聚集算法 (bootstrap aggregating)。隨機森林亦採用裝袋法抽取不同屬性 (features) 以建構各棵決策樹的節點。隨機森林中，每棵樹的建立，平均用到三分之二的觀察點數以及採用三分之一的屬性 (Hastie et al., 2009; James et al., 2013)。

隨機森林的抽樣過程，能避免決策樹之間存在相關性，並透過結合眾多決策樹的預測表現，增加資料分析的多元性，大幅提升預測準確度以及穩定性，對於非線性資料或

遺漏資料有計算上的優勢。Breiman (2001) 透過定理證明，在大數法則下，隨機森林中的決策樹數量夠多，就能避免過度配適¹⁰。更多的細節，請見 Breiman et al. (1984)。

(二)實證模型設定說明

本研究採用 R 軟體中的 rpart 套件建構決策樹迴歸以及 randomForest 套件建構隨機森林迴歸，以探討各類住宅特徵與三類長者醫療可及性對房價的影響；也就是以房價為目標變數，以各類住宅特徵及醫療可及性為屬性變數。決策數以及隨機森林屬於無母數演算法，因此並不受到傳統計量模型的共線性或是離群值問題所影響。

由於決策樹演算法較易有過度擬合的傾向，因此本研究參考 Fan et al. (2006) 建構決策樹的流程，一開始以隨機的方式，以 75% 與 25% 的比例，將總觀察點 (14,817 筆) 區分為訓練資料 (總計 11,113 筆) 及測試資料 (總計 3,704 筆)。以訓練資料建立模型，再以測試資料檢驗模型的表現。在訓練資料建構決策樹模型過程中，需要決定幾個參數數值：minisplit 表示每個節點至少所涵蓋的觀察點數 (預設值為 20)，maxdepth 代表樹的深度 (預設不超過 30)；較為關鍵的是前述用以修剪決策樹結構的複雜度參數 (complexity parameter, cp 值)，代表每一個規則所能改善模型配適度的程度，主要是用來修剪樹的分

¹⁰ Leo Breiman 於 1984 年提出 CART，2001 年的文章讓 random forest 演算法廣為人知；Breiman (2001) 透過定理證明，明確指出：“Random forests are an effective tool in prediction. Because of the Law of Large Numbers they do not overfit”。

Leo Breiman 與 Adele Cutler 在 University of California (Berkeley) 的 Random Forests 教學網頁有以下說明：“In random forests, there is no need for cross-validation or a separate test set to get an unbiased estimate of the test set error. It is estimated internally, during the run, as follows: Each tree is constructed using a different bootstrap sample from the original data. About one-third of the cases are left out of the bootstrap sample and not used in the construction of the kth tree” (取自 https://www.stat.berkeley.edu/~breiman/forests/cc_home.htm，檢索日期：2019/12/24)。

支，找到最適決策樹結構。在 rpart 套件中，根據訓練樣本，透過交叉驗證法 (10-fold cross validation) 計算各個複雜度參數的交叉驗證誤差 (x-val error)，以得到交叉驗證誤差最小的最適複雜度參數 (James et al., 2013)。從圖 4 可見，能讓交叉驗證誤差數值最小的複雜度參數為 0.002。

本研究再採用 R 軟體中的 caret 以及 e1071 套件，進行交叉驗證 (10-fold cross validation)，根據列出各複雜度參數值的均方根誤差 (RMSE) 以及平均絕對誤差(MAE) 數值 (數值愈小愈好)，確認此參數的最適數值為 0.002。如表 3 所示 (由於數值可無限切割，因此僅列出部分資訊供參考)。

單一決策樹的結構容易受到抽樣樣本的影響，因此本研究進一步採用隨機森林演算法，呈現眾多決策樹的平均表現。本研究是根據 randomForest 套件建構隨機森林迴歸，主要有兩個參數：ntree (隨機森林中樹的個數) 以及 mtry (樹節點的屬性變數個數)。隨機森林中該設定多少顆樹，是根據樹的數量相對應的均方誤差 (mean squared error, MSE) 來決定。本研究選擇 ntree 為 1,000，因相較於其他樹數量，綜合 1,000 顆樹的均方誤差總值最小 (1.663417×10^9)。接著採用 caret 以及 e1071 套件進行交叉驗證，根據各 mtry 值相對應的 RMSE 與 MAE 數值，採用 mtry 為 10 (RMSE = 40781.45；MAE = 26572.93)。

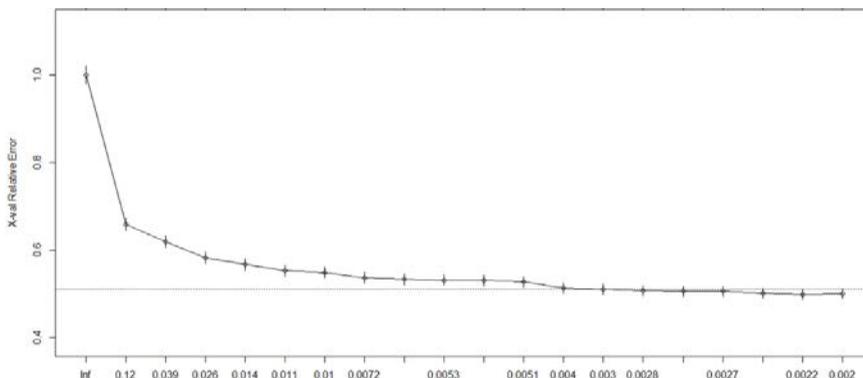


圖 4 複雜度參數的交叉驗證誤差值

表3 複雜度參數對應之均方根誤差 (RMSE) 與平均絕對誤差 (MAE)

cp	RMSE	MAE
0.002	44563.8	30713.1
0.003	44960.5	31276.3
0.004	45196.1	31362.3
0.005	45661.1	31555.0
0.006	46053.0	31829.3
0.007	46023.9	31785.4
0.008	46023.9	31785.4
0.009	46355.2	31995.1
0.010	46526.8	32150.4

本研究是下列三項指標評估、比較這兩類機器學習演算法整體表現：均方根誤差 (RMSE)、平均絕對誤差 (MAE) 及平均絕對百分比誤差 (MAPE)，這三項指標愈小，表示誤差愈小，模型預測的表現愈好。公式如下 (其中 \hat{y}_i 代表估計值， y_i 代表觀察值)：

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (8)$$

$$MAE = \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (9)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (10)$$

至於各別住宅特徵 (屬性) 對房價的影響力，決策樹迴歸主要是根據各個屬性在建構樹的過程中分枝的貢獻，也就是在分支過程中各變數降低誤差的程度。隨機森林迴歸則採用均方誤差增加率 (increase in MSE, %IncMSE) 以及結點純淨度增加幅度 (increase in

node purity, InNodePurity) 這兩項衡量值，以衡量各住宅特徵的對房價的解釋力；各變數的均方誤差增加率 (%IncMSE) 數值愈大，表示若刪除該變數，均方誤差 (MSE) 增加幅度愈大，也就反應該變數對隨機森林迴歸預測的重要性；同樣地，結點純淨度增加幅度 (IncNodePurity) 表示若一變數能增加隨機森林中的結點純淨度 (主要是以 MSE 數值減少幅度來衡量)，該變數就愈重要。

本研究採用這兩個演算法以了解哪些住宅屬性以及醫療可及性能左右大台北地區的房價；值得注意的是，若這兩模型挑選出的重要變數具有一致性，亦可佐證這些變數對房價之影響具有一定程度的穩定性。

肆、資料與變數假設

此研究採用內政部實價登錄中，民國 102 年至 103 年大台北地區 (台北市及新北市) 總計 14,817 筆住宅交易資料，包含各個住宅大樓交易總價、住址、交易年月、建築完成年月、總樓層、所在樓層、建物型態、建物面積 (平方公尺)¹¹、建物現況格局 (幾間房、廳、衛浴)、有無車位¹²等等詳細資訊。本研究主要是分析醫療資源對居民的影響，所以只考慮住家用途之不動產，採用交易單價作為目標變數 (總價除以建物面積)。交易年與建築完成年的差異，即是各建物的屋齡。實價登錄的住宅門號都已被區段化 (例如：台北市萬華區洛陽街 51-100 號)，在實證分析過程中，將取各門號區段的中位數用以定位座標，建立大台北地區住宅房價座標底圖。

¹¹ 地政事務所對建物/土地面積的計算單位以平方公尺為主，實價登錄亦是。1 平方公尺 = 0.3025 坪。

¹² 本研究沒有納入車位變數，因實務上房屋交易主要以總價附加車位價格計算，但在實價登錄的車位相關欄位易有登載錯誤或缺漏等情形。請參考「新北市受理實價登錄申報常見錯誤彙整」，取自：<https://reurl.cc/mqmd6M>，檢索日期：2019/12/12。

由於鄰里設施會影響房價，因此，根據交通部數值路網圖的相關圖資以及地理資訊軟體 (ArcGIS) 的「路網分析」(network analyst) 功能，進一步計算各住宅到各設施的路網距離 (也就是實際沿著道路的移動距離)，本研究考慮兩類設施，迎毗設施包含小學、國中、高中、大學、郵局、公園及捷運，鄰避設施包含變電站及焚化爐。由於迎毗設施在生活中較為廣泛，因此本研究參照 Peng and Chiang (2015) 的作法，以 300 公尺定義這類設施可及性的虛擬變數 (300 公尺內定義為 1，反之為 0)；例如：某住宅距離最近的捷運站是 277 公尺，因此該住宅的捷運虛擬變數定義為 1。誠如楊宗憲與蘇倬慧 (2011) 指出，迎毗或鄰避設施由於性質不同，因此設施空間分布狀況大不相同，對房價的影響必有差異；由於鄰避設施，像是焚化爐或變電站，主要設置在距離民眾居住較遠的距離，因此以 1 公里定義這類設施可及性的虛擬變數 (若超過 1 公里，定義為 0)。

或許有讀者認為，亦可採用進階二階段流動搜尋法 (E2SFCA) 精準計算這些設施的可及性；然而，本研究的重點在於醫療可及性，並非其他設施。而且從第三章研究方法的詳細說明可見，採用此方法需要非常細膩的資料。若要以 E2SFCA 計算各村里的小學可及性，就需要各市區的小學老師人數，但教育部的教師人數是以學校為單位 (而不是市區)。過往研究中 Langford et al. (2012) 透過二階流動搜尋法，根據數位公車時間表以及公車站位置作為供給面的資料，以估計公共運輸之可及性；Dony et al. (2015) 採用可變寬度流動搜尋法 (variable-width floating catchment area)，根據公共公園 (public park) 內九項設施的狀況 (包含棒球場、網球場、籃球場等)，以定義公園可及性。若要參照上述文獻方式，取得大台北地區所有捷運站時刻表相關資料或是公園內部設施相關資料，實在有一定難度。但這些方法是值得後續研究參考的方向。

再者，誠如在文獻回顧中提及，如何定義設施的可及性，各研究採用不同的方式。許多台北地區房價的文獻皆有提及捷運這項設施，洪得洋與林祖嘉 (1999) 計算各住宅至捷運的直線距離，彭建文等 (2009) 則採用捷運距離區段化的虛擬變數，亦或是 Peng and Chiang (2015) 採用一定範圍內有無捷運的虛擬變數。換言之，衡量設施對房價的影響力，存在多元的計算可能；該採用哪個，主要仍是取決於各篇文章研究的主軸。例如，

洪得洋與林祖嘉 (1999) 這篇文章的重點在於捷運可及性，因此該研究詳細計算捷運距離，其他設施則採用虛擬變數的方式 (1 公里內有嫌惡設施定義為 1，反之為 0)；Peng and Chiang (2015) 的重點在於採用截段 (spline) 方式詳細計算各住宅至醫院之距離，但針對其他設施，像是捷運，亦僅考慮一定範圍內的虛擬變數。

綜上所述，本研究考量三類變數：住宅本身屬性、鄰里設施特徵以及年長者醫療可及性。住宅屬性變數包含屋齡、屋齡平方、是否位於一樓、所在樓層、是否有電梯、建物總樓層、房間衛浴數差異值。根據文獻回顧，隨著屋齡增加所造成的建物折舊，房價因而下降 (Goodman and Thibodeau, 1997)，但這個降幅會隨著屋齡增加而減緩 (Wolverton and Senteza, 2000)，因此，本研究將屋齡及屋齡平方納入考量。由於台灣建築是住商混合，位於一樓的建物，除了出入方便之外，更有作為店面之優勢，此特徵理應對房價有加分作用 (Peng, 2019)。所在樓層愈高，往往有噪音低、視野好以及空氣較佳的好處，對房價亦有正向影響 (Jim and Chen, 2009)。電梯的存在，提升居住者出入的便利性，理應提高住宅價值 (Song and Wilhelmsson, 2010)。建物總樓層愈高，同時代表建物本身建材較佳 (方能支撐高樓層) 而且亦需要更多相關公共設施確保安全性，因此建物價格應較高 (Wen et al., 2020)。Chernobai et al. (2011) 指出，房間數與衛浴數原本代表建物的使用空間安排，但若房間數與衛浴數差異大 (例如：5 間房間與 1 間衛浴)，就代表該建物的使用空間效益低，進而影響房價。

鄰里特徵變數包含區位以及附近設施。位於市區的住宅具有工作以及生活娛樂的相關便利性 (Peng and Chiang, 2015)，因此本研究納入是否位於台北市變數；學校這類公共設施對房價亦有影響 (Frenette, 2004; Peng, 2019)，所以本研究考量建物在 300 公尺內是否有小學、國中、高中、大學。台灣的郵局兼具郵政以及金融服務 (Peng, 2019)，捷運提升生活移動便利性 (Peng and Chiang, 2015)，公園提供休憩的環境 (Hoshino and Kuriyama, 2010)，這些設施應能對周遭住宅價格有提升的作用，因此納入建物在 300 公尺內是否有郵局、捷運、公園。由於一些鄰避設施仍存在生活中，因此本研究參考楊宗憲與蘇倖慧 (2011)，考量建物在 1 公里內是否有變電站、焚化爐。

本研究著重在考量長者之三類醫療可及性對房價的影響：長者的醫師可及性、病床可及性以及救護車可及性。面對高齡化趨勢，長者的健康疑慮是以長期慢性疾病為主。以美國為例，約有 8 成的長者有一種或多種慢性疾病，像是高血壓、糖尿病等；在臺灣，這個數值亦高達 81.1%，其中有 85.9% 的長者需定期看醫生以控制慢性疾病病情¹³。由於年長者日常生活中往往需要看醫生取得慢性病處方箋，因此各地區長者之醫師可及性攸關長者們平時的醫療需求；因此，相較於病床（用於較嚴重的疾病狀況）或是救護車（用於緊急狀況），對長者而言，醫生可及性應較為重要。然而，相較於醫師或病床，救護車的存在固然提供緊急醫療服務，但在人口稠密的大台北地區（相較於花東地區），救護車警笛聲若過於頻繁，可能反而產生鄰避效應。

在此補充說明，本研究沒有考慮各鄉鎮市區作為自變數，主要是因為已經根據各村里人口以及市區醫療資源計算上述三類醫療可及性，已反應區域性的差異；此外，由於這三類醫療可及性亦已考慮各地區人口狀況，因此本研究就沒再納入人口密度等相關變數。本研究所考慮之特徵變數的定義以及敘述統計僅列於表 4。

¹³ 衛生福利部，「中華民國 102 年老人狀況調查報告」，取自 <https://www.mohw.gov.tw/dl-4707-e91899c0-d62d-49d9-9d9e-4a2def8e516c.html>，檢索日期：2018/04/16。

表 4 變數之定義及敘述統計

變數	定義	敘述統計值	
		平均值	標準差
地方醫療可及性			
長者之醫師可及性	採用 E2SFCA 計算出各村里長者之醫師可及性 (醫師數/每百人)	1.89	0.63
長者之病床可及性	採用 E2SFCA 計算出各村里長者之病床可及性 (病床數/每百人)	4.89	1.56
長者之救護車可及性	採用 E2SFCA 計算出各村里長者之救護車可及性 (救護車/每百人)	0.01	0.01
住宅屬性特徵			
連續變數		平均值	標準差
屋齡	住宅之屋齡	25.32	11.73
房間與衛浴之差異值	住宅的房間數與衛浴數的差異值	1.25	0.74
建物總樓	住宅所在的建物總樓層	7.07	4.27
所在樓層	住宅所在樓層	4.40	3.10
虛擬變數		是/有(%)	否/無(%)
位於一樓	住宅是否位於一樓(是= 1, 否= 0)	12.3%	88%
電梯	住宅是否有電梯(有= 1)	43%	57%
鄰里屬性特徵 (虛擬變數)		是/有(%)	否/無(%)
台北市	位於台北市=1；位於新北市=0	39%	61%
小學	住宅在 300 公尺內是否有小學 (有= 1)	18%	82%
國中	住宅在 300 公尺內是否有國中 (有= 1)	7%	93%
高中	住宅在 300 公尺內是否有高中 (有= 1)	4%	96%
大學	住宅在 300 公尺內是否有大學 (有= 1)	1%	99%
郵局	住宅在 300 公尺內是否有郵局 (有= 1)	20%	80%
捷運	住宅在 300 公尺內是否有捷運 (有= 1)	8%	92%
公園	住宅在 300 公尺內是否有公園 (有= 1)	26%	74%
焚化爐	住宅在 1 公里內是否有焚化爐 (有= 1)	1%	99%
變電站	住宅在 1 公里內是否有變電站 (有= 1)	6%	94%

伍、實證結果分析

一、實證結果：決策樹迴歸

從整體表現而言 (表 5)，根據測試資料 (test data) 所建構的決策樹迴歸能解釋住宅價格約 50% 的變化，RMSE 值是 44642.52、MAPE 值為 0.36 以及 MAE 值為 30443.20。訓練資料 (train data) 的結果也附上供參。誠如在文獻中提及，過往許多研究，像是 Fan et al. (2006)、Özsoy and Şahin (2009) 或是 Pérez-Rave et al. (2019) 都有採用決策樹迴歸分析房價相關議題，亦在文章中呈現單一決策樹迴歸的樹狀結果。值得注意的是，James et al. (2013) 就曾指出，決策樹的建構容易受到樣本更動的影響，本研究亦同過往文獻研究，呈現的是由一次隨機抽樣所生成的決策樹，但由於訓練資料與測試資料的整體表現相差不大，代表此決策樹模型並無過度擬合的狀況。更重要的是，決策樹迴歸中顯示之影響房價的屬性變數，與隨機森林迴歸 (結合眾多迴歸樹的結果) 具有一致性，由此亦可佐證迴歸樹的表現穩定性 (請見後續實證結果討論說明)。

誠如上述，決策樹是以樹狀為基礎的演算法，採用歸納規則將資料從根結點開始分類，在分支上越是關鍵的特徵屬性，距離根結點愈近。決策樹迴歸的優勢在於能將決策過程中的二元決策輸出以圖像呈現，從樹狀圖中可清楚觀察到模型對於房價單價 (目標變數) 的預測過程，如圖 5 所示。然而在特徵屬性多的狀況下，受限於圖形格式，目標變數的分類數值以及最下方的預測值 (圖 5 最下方) 往往重疊而不易判別。因此在內容不變之下，亦可用圖 6 表列方式呈現，能列出更多的資訊 (像是葉結點涵蓋多少觀察點)。圖 5 下方的葉結點 (node) 標號與圖 6 的標號一致。

表 5 整體表現：決策樹迴歸

整體表現指標	R^2	RMSE	MAPE	MAE
訓練資料	0.52	43543.8	0.33	30133.07
測試資料	0.50	46642.5	0.36	30443.20

如圖 6 所示，在各種屬性的分類下，總計有 23 種房價預測值，也就是葉結點；以下是舉其中一些預測值為例作說明。從圖 6 可見（從左邊開始看起），這棵決策樹的根結點，從「是否為台北市」這個屬性開始分枝成「是否為台北市 < 0.5 」（結點 [2]；亦即：新北市）以及「是否為台北市 ≥ 0.5 」（[15]；亦即：台北市）；從 [2] 結點再根據「長者之醫師可及性」屬性，分枝為「長者之醫生可及性 < 0.01455 」[3] 及「長者之醫師可及性 ≥ 0.01455 」[8]；若再從 [3] 結點，可再分支為「長者之醫師可及性 < 0.00895 」[4] 及「長者之醫師可及性 ≥ 0.00895 」[5]。由於結點 [4] 已經是葉結點，因為其中已經顯示住宅單價的預測值。換言之，從結點 [2]、[3] 及 [4] 可知，若建物位於新北市 [2]，而且該建物的所在地的長者之醫師可及性小於 0.01455 [3]（亦即：每一長者可取得 0.01455 醫師的醫療服務/每 100 位長者可取得 1.455 醫師的醫療服務），而且當這項可及性甚至低於 0.00895，若建物具備這三類屬性，該住宅單價的預測值為 58,951.310（元/每平方公尺）¹⁴，在資料中有 787 筆建物歸類於此葉結點。「長者之醫師可及性 ≥ 0.00895 」的結點 [5]，還可再分枝為兩個葉結點「長者之救護車可及性 ≥ 0.00015 」[6] 與「長者之救護車可及性 < 0.00015 」[7]。值得注意的是，葉結點 [6] 與 [7] 顯現出長者之救護車可及性對房價並無加分作用；換言之，當「長者之醫師可及性 ≥ 0.00895 」時 [5]，相較「長者之救護車可及性 < 0.00015 」[7] 時的住宅單價 (78,095.984)，「長者之救護車可及性 ≥ 0.00015 」[6] 時的住宅單價反而較低 (55,694.613)，而且還低於「長者之醫師可及性 < 0.00895 」[4] 的住宅單價 (58,951.310)。這個結果反應現實生活中都會地區的救護車仍存

¹⁴ 實價登錄中建物面積以平方公尺為單位，因此住宅單價的單位亦是。若以坪為單位，僅需將數值除以 0.3025。例如，58,951.31（元/每平方公尺）等於 194,880.36（元/每坪）。

在一定程度的嫌惡性¹⁵。

若從台北市 [15] 看起，根據長者病床狀況分支為「長者之病床可及性 < 0.06145 」 [16] 以及「長者之病床可及性 ≥ 0.06145 」 [25]。若延續結點 [16]，該點再根據電梯分支為「電梯有無 < 0.5 」 ([17]；無電梯) 以及「電梯有無 ≥ 0.5 」 ([20]；有電梯)；若從 [20] 結點延續，再根據長者之醫師可及性分支為「長者之醫師可及性 < 0.02145 」 [21] 以及「長者之醫師可及性 ≥ 0.02145 」 [24]；從 [21] 開始，根據屋齡，再分支為「屋齡 ≥ 9.5 」 [22] 與「屋齡 < 9.5 」 [23]。由圖可見，[22] 與 [23] 皆是葉結點。從這些分支決策過程可知，若一建物位於台北市 [15]，所在地區的長者之病床可及性小於 0.06145 ([16]；每 100 位長者可取得約 6.1 張病床的服務)，在建物有電梯的情況下 [20]，該建物所在地的長者之醫師可及性小於 0.02145 ([21]；每 100 位長者可取得約 2 位醫師的服務)，再加上該住宅屋齡大於 9.5 年 [22]，這類住宅的單價預測值為 140,918.822 元，在這筆資料中有 405 筆建物屬於這類；但在最後一項屬性中，若住宅屋齡是小於 9.5 年 [23]，住宅的單價預測值就提高到 166,711.637 元，有 248 筆建物屬於這類。

在決策樹分支的過程中，越接近根節點的屬性，其對於整顆樹分支的重要性愈高。從圖 6 可見，區分房價最核心的特徵是「是否位於台北市」，區分成兩類之後，在新北市 [2] 的狀況下，在分支過程中最早出現的醫療變數是「長者之醫師可及性」，在「長者之醫師可及性」小於 0.01455 [3] 的情況下，若該數值又再小於 0.00895，就形成葉結點 [4] (住宅單價為 58,951.310 元)；但若該數值雖小於 0.01455 但大於 0.00895，就再根據「長者之救護車可及性」分支屬性，決定葉結點 [6] 及 [7]，住宅單價各為 55,694.613 元及 78,095.984 元。相對的，若是「長者之醫師可及性」大於 0.01455 [8]，後續的分支屬性則是「位於一樓與否」以及「屋齡」，就達到四個葉結點 ([10]、[11]、[13]、[14])，

¹⁵ 2018 年新北市永和耕莘醫院就曾被鄰里居民連署「到院前 300 公尺需降低警鳴器音量」以及「到院前 100 公尺需關閉警鳴器 (沒有聲音，只能閃燈)」，各方有所爭議因而不了了之。相較於鄉村地區居民苦於缺乏救護車，救護車在人口稠密的都會區 (不缺乏救護車的地區) 仍存在一定程度之嫌惡性。

所預測的住宅單價各為 93,231.275、111,031.802、118,585.943、149,702.539 元。由這兩大分支 [3] 及 [8] 所延伸的葉結點可見，在新北市中，「長者之醫師可及性」高低與否，讓住宅單價有一定程度的差異；某種程度可能反應，新北市地區幅員較廣，能否取得醫生的醫療診斷服務應較為重要。除了長者之醫師可及性以及救護車可及性之外，值得注意的是住宅本身特徵，像是「位於一樓與否」以及「屋齡」對於大台北地區房價的影響仍不容忽視。相對的，在台北市 [15]，最早出現的醫療變數則是「長者之病床可及性」。台北市人口密集，在病床使用頻率較高的情況下，病床可取得性更為重要；這個結果可能與台灣醫療改革基金會所公告的十大醫療民怨之一的「近年健保住院病床一床難求」現象¹⁶，相較於中南部，北部醫學中心病床壅塞程度較為嚴重。因此長者能否取得便利的病床服務在台北地區更顯關鍵。除此之外，住宅本身的特徵像是「屋齡」、「位於一樓與否」、「電梯有無」、「建物總樓層」、「變電站有無」，對台北地區的房價仍有影響力，這也呼應過往研究的結果。

建構此決策樹迴歸的關鍵屬性變數主要是由各變數對樹狀結構的產生之提升程度 (improvement) 來決定；各屬性變數的重要性如圖 7 所示，最關鍵的前 14 項屬性變數依序是：是否位於台北市，建物所在地的長者對病床之可及性、長者對醫師之可及性、長者對救護車之可及性、建物在 300 公尺內有無公園、建物屋齡平方、建物屋齡、住宅所在大樓的總樓層、住宅所在樓層、建物是否位於一樓、建物是否有電梯、建物在 300 公尺內有無捷運、建物在 1 公里內有無變電站、房間與衛浴數差異；其他的屬性變數數值過低，表示在建構迴歸樹之貢獻偏低。

¹⁶ 監察院公報，第 2726 期，取自 <https://reurl.cc/kVz5db>，檢索日期：2019/02/12。

```

[1] root (根節點)
|
| [2] 台北市與否 < 0.5
| |
| | [3] 長者之醫師可及性 < 0.01455
| | |
| | | [4] 長者之醫師可及性 < 0.00895: 58951.310 (n = 787, err = 411151186662.4)
| | | [5] 長者之醫師可及性 >= 0.00895
| | | | [6] 長者之救護車可及性 >= 0.00015: 55694.613 (n = 284, err = 1.08588e+11)
| | | | [7] 長者之救護車可及性 < 0.00015: 78095.984 (n = 1255, err = 572555801853.7)
| | [8] 長者之醫師可及性 >= 0.01455
| | | [9] 位於一樓與否 < 0.5
| | | | [10] 屋齡 >= 17.5: 93231.275 (n = 2913, err = 1.956791e+12)
| | | | [11] 屋齡 < 17.5: 111031.802 (n = 963, err = 772106610789.1)
| | | [12] 位於一樓與否 >= 0.5
| | | | [13] 屋齡 < 37.5: 118585.943 (n = 316, err = 715499692895.0)
| | | | [14] 屋齡 >= 37.5: 149702.539 (n = 206, err = 952344946333.2)
| [15] 台北市與否 >= 0.5
| | [16] 長者之病床可及性 < 0.06145
| | | [17] 電梯有無 < 0.5
| | | | [18] 位於一樓與否 < 0.5: 127186.824 (n = 1016, err = 1.835899e+12)
| | | | [19] 位於一樓與否 >= 0.5: 157841.824 (n = 301, err = 1.715749e+12)
| | | [20] 電梯有無 >= 0.5
| | | | [21] 長者之醫師可及性 < 0.02145
| | | | | [22] 屋齡 >= 9.5: 140918.822 (n = 405, err = 797722739917.2)
| | | | | [23] 屋齡 < 9.5: 166711.637 (n = 248, err = 414545321253.3)
| | | | [24] 長者之醫師可及性 >= 0.02145: 183743.131 (n = 320, err = 1.453864e+12)
| | [25] 長者之病床可及性 >= 0.06145
| | | [26] 電梯有無 < 0.5
| | | | [27] 位於一樓與否 < 0.5
| | | | | [28] 變電站有無 < 0.5: 154304.103 (n = 802, err = 2.45645e+12)
| | | | | [29] 變電站有無 >= 0.5: 203196.190 (n = 63, err = 300928565275.7)
| | | | [30] 位於一樓與否 >= 0.5
| | | | | [31] 變電站有無 < 0.5: 203952.639 (n = 166, err = 1.111898e+12)
| | | | | [32] 變電站有無 >= 0.5: 350748.833 (n = 12, err = 80555157511.7)
| | | [33] 電梯有無 >= 0.5
| | | | [34] 屋齡 >= 12.5: 195571.422 (n = 709, err = 3.653605e+12)
| | | | [35] 屋齡 < 12.5
| | | | | [36] 長者之救護車可及性 >= 0.00025: 163162.714 (n = 28, err = 89841867203.7)
| | | | | [37] 長者之救護車可及性 < 0.00025
| | | | | | [38] 長者之醫師可及性 < 0.02615: 207202.947 (n = 95, err = 390547663216.7)
| | | | | | [39] 長者之醫師可及性 >= 0.02615
| | | | | | [40] 長者之病床可及性 >= 0.0762: 187772.038 (n = 26, err = 49764136351.0)
| | | | | | [41] 長者之病床可及性 < 0.0762
| | | | | | | [42] 屋齡 < 2.5
| | | | | | | | [43] 建物總樓層 >= 13.5: 135848.769 (n = 13, err = 82560957312.3)
| | | | | | | | [44] 建物總樓層 < 13.5: 240245.923 (n = 26, err = 140408642847.8)
| | | | | | [45] 屋齡 >= 2.5: 261363.579 (n = 159, err = 1.00758e+12)

```

圖 6 決策樹迴歸：表列式

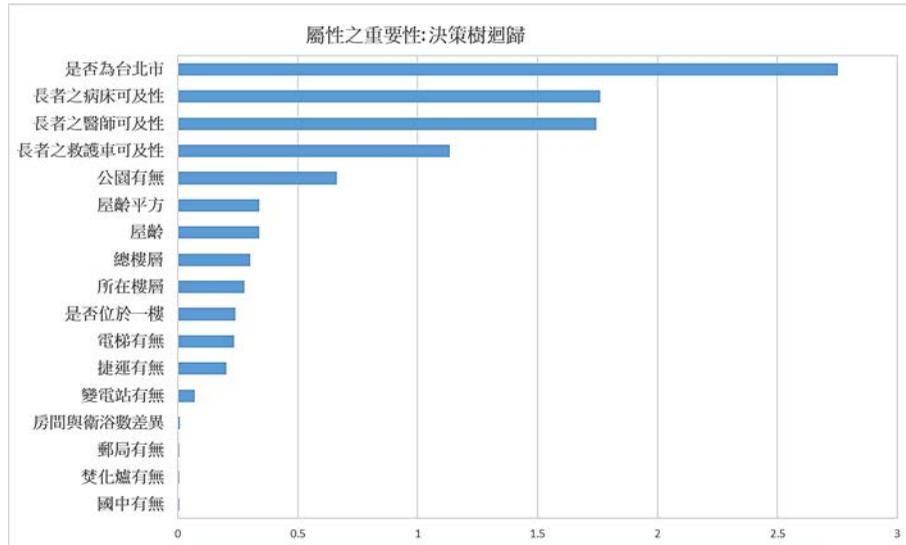


圖 7 屬性之重要性：決策樹迴歸

二、實證結果：隨機森林迴歸

相較於上述迴歸樹為一次隨機抽樣生成的樹，本研究的隨機森林迴歸是 1,000 棵決策樹迴歸的組合，統整眾多決策樹的預測以得到更為穩定的結果。隨機森林迴歸的 R^2 為 0.69，高於決策樹迴歸的 0.50 (表 5)；此外，相較於決策樹迴歸，隨機森林迴歸的 RMSE (40633.68)、MAPE (0.28463) 及 MAE (26174.06) 數值較小，表示估計之誤差值較低，呈現出更有效率且穩定的估計表現。

隨機森林是眾多決策樹的組合，因此無法像決策樹用視覺化的方式匯出內含許多樹的隨機森林，因此在模型解釋上主要是針對各屬性對於建構隨機森林的重要性。進一步查看建構隨機森林迴歸的屬性變數，主要是根據均方誤差增加率 (%IncMSE) 與結點純淨度增加幅度 (IncNodePurity) 這兩個指標來判定，用以衡量各屬性在分支上減少錯誤的幅度 (數值愈高表示對於「減少錯誤」的貢獻愈大)。如圖 8 所示，重要性排序前 14 名的

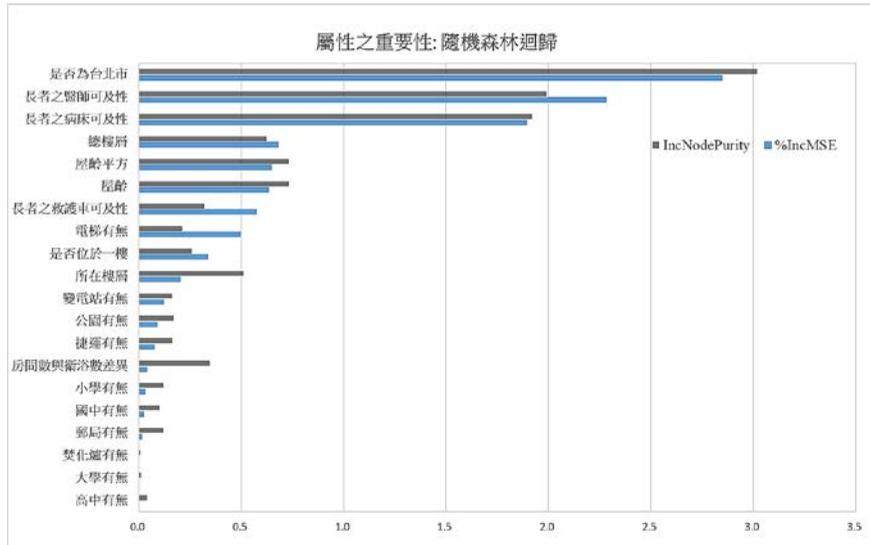


圖 8 屬性之重要性：隨機森林迴歸

變數依序為：是否位於台北市、住宅所在地區的長者對醫師之可及性、長者對病床之可及性、住宅所在之大樓總樓層、屋齡平方、屋齡、長者對救護車之可及性、電梯有無、是否位於一樓、住宅所在樓層、住宅在 1 公里內有無變電站、住宅在 300 公尺內有無公園、有無捷運、房間與衛浴數差異。其他變數影響力則偏低。

各屬性在這兩個指標的表現亦可用圖 9 呈現，縱軸是均方誤差增加率 (%IncMSE)，橫軸是結點純淨度增加幅度 (IncNodePurity)。屬性愈是座落在右上角，表示在這兩個指標的數值愈高，該屬性對建構隨機森林迴歸愈是重要，對房價的影響更為顯著。

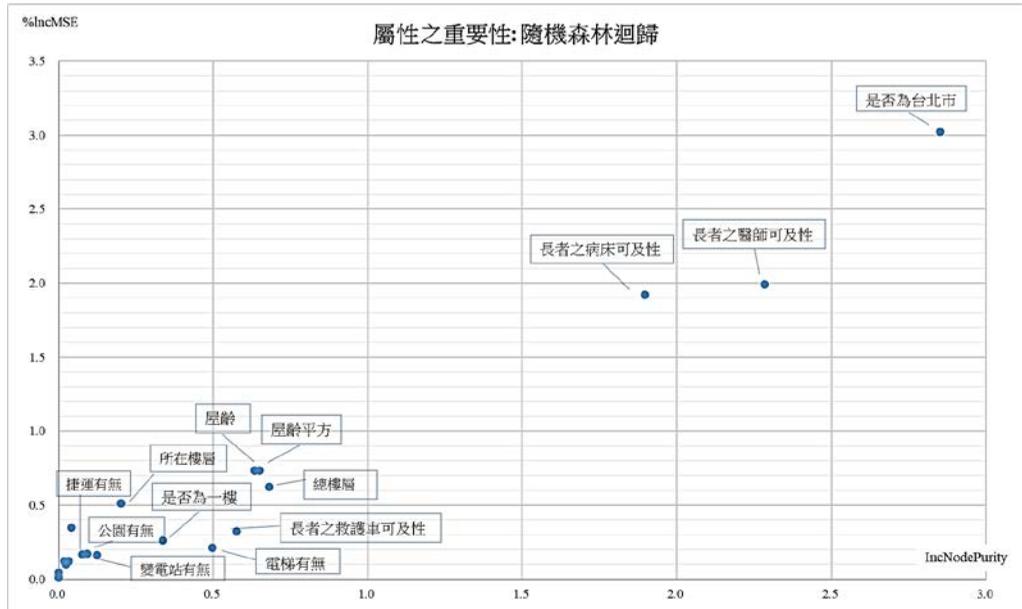


圖 9 屬性之重要性：隨機森林迴歸 (以 x 、 y 軸之概念呈現)

三、實證結果討論

從上述實證結果可見，隨機森林為眾多決策樹的組合表現，因此整體預測的結果比決策樹迴歸更為穩定。更重要的是，就屬性重要性而言，從決策樹迴歸 (圖 7) 與隨機森林迴歸 (圖 8) 比較得知，雖然順序有些不同，但排序前 14 項的屬性相同。由此亦可佐證，即使在不同的模型中，這些左右大台北地區房價的變數仍存在一致性的影響；某種程度亦可反應決策樹與隨機森林迴歸得到的結果，具有一定程度的穩定性。

值得注意的是，建構房價的決策樹與隨機森林迴歸中，最關鍵的屬性皆為「是否為台北市」，緊接著是本研究關注的長者之醫療可及性，尤其是長者對「醫師」與「病床」的可及性。主要的原因在於，在現今老年人口不斷增加的年代，老化趨勢已不可免；年長者往往受慢性病之苦，需要長期透過看診以取得處方簽，或是遇到一些病痛需要住院

(例如：開刀等)，醫師與病床的可及性在所必須，因此對住居選擇以及房價產生影響。但長者對「救護車」的可及性，在決策樹迴歸中的屬性重要性排序第四，在隨機森林迴歸中排序第七 (根據%IncMSE) 或第九 (根據 InNodePurity)；由於隨機森林是由眾多決策樹架構而成，因此可推測，相較於平常需要之醫師資源或病床資源，對一般受慢性病所苦的長者而言，用於緊急醫療的救護車的重要性較為其次。

除了上述屬性之外，決策樹迴歸與隨機森林迴歸在住宅屬性與鄰里設施屬性也是相同；住宅屬性包含屋齡、屋齡平方、位於一樓、電梯、所在樓層、大樓總樓層、房間與衛浴數差異；鄰里設施屬性包含周遭環境是否有捷運、公園或變電站。如同先前文獻提及，隨著屋齡增加所造成的建物折舊，進而影響住宅價值 (Goodman and Thibodeau, 1997)。值得注意的是，大台北地區不僅老年人口占比高，屋齡亦偏高；根據內政部統計，全台 40 年以上老屋存量，台北市占比達 36.9%，新北市為 28.5%。大台北地區老屋偏多，但都更進度卻緩不濟急，因此屋齡仍是影響大台北地區房價的主要特徵。台灣住宅密度高，主要採用住商混合的建築模式，位於一樓的建物，除了出入方便之外，更有作為店面之優勢，Peng (2019) 就曾指出此特徵在大台北地區對房價有加分作用。

以電梯來說，Song and Wilhelmsson (2010) 就發現大廈住宅密集，電梯提升居住者出入便利性，本研究亦證明電梯的存在對於住宅高度密集之大台北地區的房價有影響力。此外，所在樓層愈高，往往有噪音低、視野好以及空氣較佳的好處，對大台北地區房價亦有加分作用；本研究的結果也呼應 Jim and Chen (2009) 的結論。Wen et al. (2020) 指出建物總樓層愈高，表示建物本身建材較佳，因而影響建物價格；本研究的發現亦證實這個現象。就設施而言，Deng et al. (2012) 指出捷運可及性對新加坡房價有決定性的影響；同樣的，大台北地區人民大幅仰賴捷運通勤，居住地點有無捷運對大台北地區房價是個關鍵因素。公園提供休憩的環境，Panduro and Veie (2013) 證明其對房價有提升的效果；楊宗憲與蘇倖慧 (2011) 曾證明變電站對房價的負向影響。

陸、結論

生老病死是人生必經過程，如何讓老年生活過的有品質，不僅攸關現在的年長者，也影響到未來的世代。面對邁入高齡社會的台灣，本研究基於跨領域的精神，結合特徵價格理論以及機器學習演算法，希望就此議題得到更為完整的結論。本研究根據大台北地區相關的醫療、人口資料，採用 Luo and Qi (2009) 進階式兩階段流動搜尋法估計各個村里的三種長者之醫療可及性；進一步結合大台北地區實價登錄資料，在特徵價格理論之下，透過決策樹以及隨機森林迴歸，探討住宅屬性、鄰里屬性以及長者醫療可及性對房價的影響。結果發現，相較於決策樹迴歸，隨機森林迴歸有更高的解釋力；而且兩項演算法呈現相對穩定的結果，也就是建構模型的屬性變數的重要性，具有高度一致性。更重要的是，相較於其他變數，三類長者醫療可及性，也是本研究著眼的關鍵因素，對房價有更為重要的影響力。

根據衛生福利部的報導¹⁷，台灣高達 89% 的長者中有慢性病；但值得注意的是，人口結構已經出現城鄉區域的差異，再加上醫療資源區域分配不均，自然會造成人們傾向隨醫療資源而居。就人口年齡結構而言，早已有「北高南低」的現象，也就是平均壽命由北往南逐漸遞減。根據內政部統計處公告的「簡易生命表」歷年資料，台北市與新北市的平均壽命，長期居各縣市之首；在 2011 年時，臺北市 82.70 歲最高、新北市以 80.46 歲居次；時間推移至 2019 年，台北市仍居首位 (83.63 歲)，新北市仍緊追在後 (81.37 歲)。在此老化時空背景下，都市化地區人口年齡偏高，帶動醫療需求增加，然而醫療資源增長幅度較緩慢，造成醫師與病床數區域分配不均，人們在健康與養老的考量下，在選取住居時，就會更加關注這方面的資源。

在醫療可及性相關文獻中，本研究是少數各別考量醫師數、病床數以及救護車數以

¹⁷ 衛生福利部國民健康署：<https://www.mohw.gov.tw/fp-16-25258-1.html>。檢索日期：2015/10/12。

衡量醫療資源可及性，並採用機器學習演算法探討醫療可及性與大台北地區房價之間的關係。實證結果證明：住宅屬性對於房價的影響固然重要，但長者在空間上能否就近取得醫生或病床服務的重要性，亦不容忽視；這個發現也呼應上述台灣現實層面的現象。畢竟年長者多數有醫療需求，尤其對於台灣高達 9 成有慢性病的長輩來說，能夠就近讓醫生看診、取藥，或是臨時需要住院需求，皆是攸關老年的生活品質，因此醫療可及性自然就會左右當地的住宅價格。從這項研究可知，機器學習之應用並不侷限於資工領域，其他經濟社會領域議題都是往後學者可延伸討論的範疇。

在此首先補充說明的是，或許一般會認為採用機器學習演算法，應採用愈多屬性變數愈好，但過往採用決策樹迴歸或隨機森林迴歸的房價研究，考量的屬性變數主要都在 20 個左右，像是 Pérez-Rave et al. (2019) 根據 22 個屬性變數分析 6 萬多筆的住宅，Özsoy and Şahin (2009) 納入約 23 個屬性變數考量一千多筆的住宅，Yoo et al. (2012) 考量 24 個屬性變數分析四千多筆的住宅。因為住宅本身是異質性的商品，會影響其價格的因素很多，實務研究主要受限於能取得的資料或是相關理論的假設，所以過往眾多文獻考量的特徵因素皆有差異。後續台灣的相關研究若能取得其他影響房價的特徵屬性，像是近年的土壤液化區域資料、犯罪地圖等，應更能提高對住宅價格的解釋能力。此外，誠如前述，固然亦能將進階二階段流動分析法用以計算其他設施的可及性，但此篇文章受限於資料無法取得；若後續研究能取得相關資料透過 E2SFCA 方式計算公園、捷運等設施之可及性，定能對住宅領域議題有一定程度的貢獻。第二，由於本研究著眼在公寓或華廈等建物，因此後續的研究亦可考慮這些醫療可及性的因素對於透天厝或是其他類型的建物，是否也有類似的影響力。

第三，既然人口結構以及醫療資源存在區域差異，若是在非都會地區，長者的醫療可及性是否仍能左右房價，是本研究無法涵蓋的議題，但也是值得關注的重點。因此，在有相關資料的前提下，後續的研究可進一步考慮全台灣的資料，探討長者之醫療可及性的城鄉差距，以及此差距對房價的影響程度。

第四，本研究是透過進階二階段流動分析法探討潛在空間可及性 (potential spatial

accessibility)。過往台灣方面的醫療可及性研究，像是章殷超等 (2011) 透過 2SFCA 衡量台灣各鄉鎮西醫師可及性、廖興中 (2013) 透過 E2SFCA 衡量台灣小兒科醫療資源在鄉鎮層級空間可及性或本研究透過 E2SFCA 估計大台北地區各村里長者之三類醫療可及性，並沒有局限於各醫院的資源，而是以各空間有多少醫療資源來衡量。但有些民眾認知的醫療資源往往還會考量各醫院資源品質或名聲，這是台灣文獻與本研究尚未探討的非空間可及性 (aspatial accessibility) 面向，也就是特定非空間因素 (例如社會經濟因素) 對醫療供給或需求所造成的隔離。如何將一些台灣民眾的大醫院迷思納入考量，這是值得後續研究探討的議題。

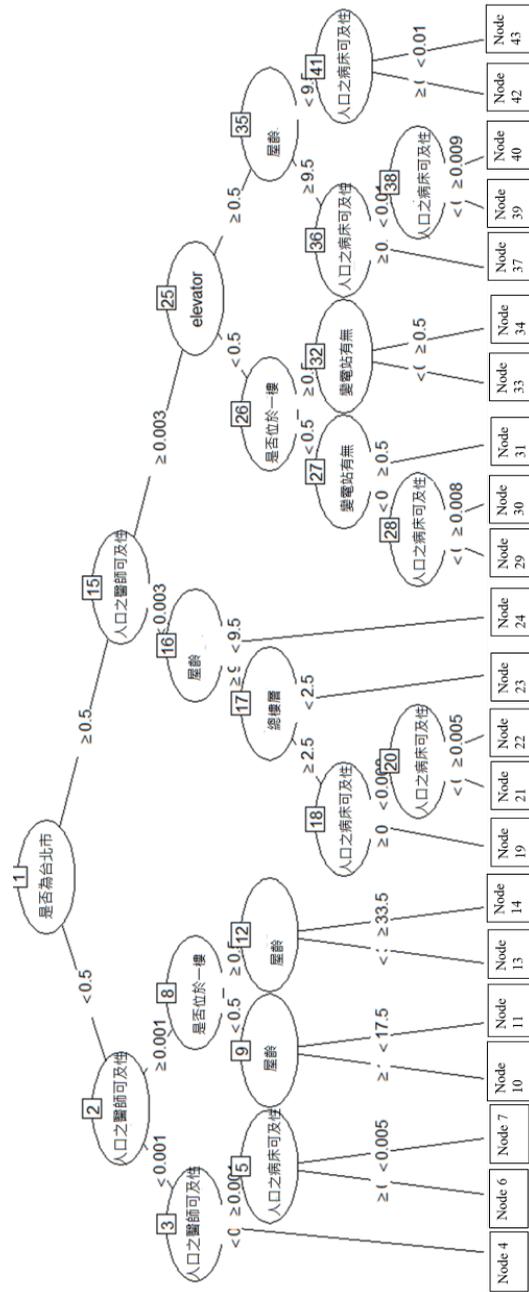
近年醫療資源方面的變化，亦值得關注。爲了因應長者對醫療水準以及生活品質的要求，養生村以另一種型態結合醫療與住宅需求。固然養生村的立意良善，然而入住門檻偏高。目前在台灣的銀髮住宅，是以健康且能行動自理的長者爲對象，而且保證金門檻也讓一般人無法將銀髮住宅視爲養老的考量；例如，集中在大台北地區之新壽養生村的保證金高達 1,500 萬元¹⁸，位於新北市的潤福銀髮住宅的保證金亦須 650-690 萬元¹⁹。而且這類銀髮住宅主要位於都會區 (台北、新北、桃園等)，因此在空間上的普及性較低。換言之，一般人仍然可能傾向就近取得傳統的醫療資源爲住宅選項。

由此可見，在現今的老化趨勢之下，各地醫療資源是否能滿足當地年長者需求，以降低各地醫療資源不均之可能性，將進而左右購屋者的選擇以及後續房價的變化。這將是不容忽視的現象。這些實證結果應能提供相關單位在規劃醫療資源之地理位置時做參考，讓醫療資源達到分配達到最適，極大化居民 (尤其是年長者) 的生活環境品質。

(收件日期爲民國 109 年 10 月 2 日，接受日期爲民國 110 年 1 月 20 日)

¹⁸ 工商時報「新壽養生村入住保證金 1,500 萬」報導日期：2015/07/24。取自 <https://readers.ctee.com.tw/cm/20150724/a05aa5/636829/share>，檢索日期：2019/02/01。

¹⁹ 康健雜誌「高級銀髮出租宅，收費不平價」第 240 期，報導日期：2018/11/01。取自 <https://www.commonhealth.com.tw/article/article.action?nid=78375>，檢索日期：2019/01/01。



附圖 1 決策樹迴歸 (樹狀結構)：總人口之醫療可及性

參考文獻

一、中文部分

- 李泳龍、黃宗誠、戴政安與李善將，2009，「醫學中心對鄰近住宅環境影響之研究」，*建築與規劃學報*，10：75-93。(Lee, Y. L., C. C. Huang, C. A. Tai, and S. C. Lee, 2009, "Effects of Medical Center on Neighboring Residence Environment", *Journal of Architecture and Planning*, 10: 75-93.)
- 洪得洋與林祖嘉，1999，「台北市捷運系統與道路寬度對房屋價格影響之研究」，*住宅學報*，8：47-67。(Hong, D. Y. and C. C. Lin, 1999, "A Study on the Impact of Subway System and Road Width on the Housing Prices of Taipei", *Journal of Housing Studies*, 8: 47-67.)
- 陳信宇與楊育誠，2019，「優質酒品認證標誌之特徵價格研究-以高粱酒為例」，*應用經濟論叢*，106：181-208。(Chen, H. T. and Y. C. Yang, 2019, "Price Premium for Alcohol Quality Certification Label- A Study for Kaoliang Liquor in Taiwan", *Taiwan Journal of Applied Economics*, 106: 181-208.)
- 彭建文、楊宗憲與楊詩韻，2009，「捷運系統對不同區位房價影響分析—以營運階段為例」，*運輸計劃季刊*，38：275-296。(Peng, C. W., C. H. Yang, and S. Y. Yang, 2009, "The Impacts of Subways on Metropolitan Housing Prices in Different Locations-After the Opening of the Taipei Subway System", *Transportation Planning Journal*, 38: 275-296.)
- 章殷超、溫在弘與賴美淑，2011，「比較不同地理可近性分析方法於評估台灣各鄉鎮每萬人口西醫師數之差異」，*台灣衛誌*，30：558-572。(Chang, Y. C., T. H. Wen, and M. S. Lai, 2011, "Comparisons of Different Methods of Geographical Accessibility in Evaluating Township-level Physician-to-Population Ratios in Taiwan", *Taiwan Journal of*

Public Health, 30: 558-572.)

楊宗憲與蘇倬慧，2011，「迎毗設施與鄰避設施對住宅價格影響之研究」，住宅學報，20：61-80。(Yang, C. H. and S. H. Su, “The Impacts of Housing Price in YIMBY and NIMBY Facilities”, *Journal of Housing Studies*, 20: 61-80.)

廖興中，2013，「臺灣小兒科醫療資源空間可接近性分析」，公共行政學報，44：1-39。(Liao, H. C., 2013, “Spatial Accessibility to Pediatric Services in Taiwan”, *Journal of Public Administration*, 44: 1-39.)

二、英文部分

Andersson, H., 2005, “The Value of Safety as Revealed in the Swedish Car Market: An Application of the Hedonic Pricing Approach”, *Journal of Risk and Uncertainty*, 30: 211-239.

Antipov, E. A and E. B. Pokryshevskaya, 2012, “Mass Appraisal of Residential Apartments: An Application of Random forest for Valuation and a CART-based Approach for Model Diagnostics”, *Expert Systems with Applications*, 39: 1772-1778.

Archer, W. R., D. H. Gatzlaff, and D. C. Ling, 1996, “Measuring the Importance of Location in House Price Appreciation”, *Journal of Urban Economics*, 40: 334-353.

Basuchoudhary, A., J. T. Bang, and T. Sen, 2017, *Machine-Learning Techniques in Economics: New Tools for Predicting Economic Growth*, Cham: Springer.

Batalhone, S., J. Nogueira, and B. Mueller, 2002, “Economics of Air Pollution: Hedonic Price Model and Smell Consequences of Sewage Treatment Plants in Urban Areas”, *Working Paper*, No. 234, Department of Economics, University of Brasília.

Berry, M. J. A. and G. S. Linoff, 2011, *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, Indianapolis: Wiley Publishing.

Breiman, L., 1996, “Bagging Predictors”, *Machine Learning*, 24: 123-140.

Breiman, L., 2001, “Random Forests”, *Machine Learning*, 45: 5-32.

Breiman, L., J. Friedman, R. A. Olshen, and C. J. Stone, 1984, *Classification and Regression*

- Trees*, Boca Raton: Chapman & Hall/CRC.
- Bühlmann, P., P. Drineas, M. Kane, and M. van der Laan, 2016, *Handbook of Big Data*, Boca Raton: CRC Press.
- Campbell, R. J., A. M. Ramirez, K. Perez, and R. G. Roetzheim, 2003, "Cervical Cancer Rates and the Supply of Primary Care Physicians in Florida", *Family Medicine*, 35: 60-64.
- Carroll, T. M., T. M. Clauretje, and J. Jensen, 1996, "Living Next to Godliness: Residential Property Values and Churches", *The Journal of Real Estate Finance and Economics*, 12: 319-330.
- Chay, K. Y. and M. Greenstone, 2005, "Does Air Quality Matter? Evidence from the Housing Market", *Journal of Political Economy*, 113: 376-424.
- Chernobai, E., M. Reibel, and M. Carney, 2011, "Nonlinear Spatial and Temporal Effects of Highway Construction on House Prices", *The Journal of Real Estate Finance and Economics*, 42: 348-370.
- Clapp, J. M. and C. Giaccotto, 1998, "Residential Hedonic Models: A Rational Expectations Approach to Age Effects", *Journal of Urban Economics*, 44: 415-437.
- Crompton, J. L., 2001, "The Impact of Parks on Property Values: A Review of the Empirical Evidence", *Journal of Leisure Research*, 33: 1-31.
- Deng, Y., D. P. McMillen, and T. F. Sing, 2012, "Private Residential Price Indices in Singapore: A Matching Approach", *Regional Science and Urban Economics*, 42: 485-494.
- Dony, C. C., E. M. Delmelle, and E. C. Delmelle, 2015, "Re-Conceptualizing Accessibility to Parks in Multi-Modal Cities: A Variable-Width Floating Catchment Area (VFCA) Method", *Landscape and Urban Planning*, 143: 90-99.
- Fan, G. Z., S. E. Ong, and H. C. Koh, 2006, "Determinants of House Price: A Decision Tree Approach", *Urban Studies*, 43: 2301-2315.
- Fransen, K., T. Neutens, P. De Maeyer, and G. Deruyter, 2015, "A Commuter-Based Two-Step Floating Catchment Area Method for Measuring Spatial Accessibility of Daycare Centers", *Health & Place*, 32: 65-73.
- Frenette, M., 2004, "Access to College and University: Does Distance to School Matter?", *Canadian Public Policy/Analyse de Politiques*, 30: 427-443.

- Ghoddusi, H., G. G. Creamer, and N. Rafizadeh, 2019, "Machine Learning in Energy Economics and Finance: A Review", *Energy Economics*, 81: 709-727.
- Goodman, A. C. and T. G. Thibodeau, 1997, "Dwelling-Age-Related Heteroskedasticity in Hedonic House Price Equations: An Extension", *Journal of Housing Research*, 8: 299-317.
- Grömping, U., 2009, "Variable Importance Assessment in Regression: Linear Regression versus Random Forest", *The American Statistician*, 63: 308-319.
- Guagliardo, M. F., 2004, "Spatial Accessibility of Primary Care: Concepts, Methods and Challenges", *International Journal of Health Geographics*, 3: 1-13.
- Hastie, T., R. Tibshirani, and J. Friedman, 2009, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York: Springer.
- Hite, D., W. Chern, F. Hitzhusen, and A. Randall, 2001, "Property-Value Impacts of an Environmental Disamenity: The Case of Landfills", *The Journal of Real Estate Finance and Economics*, 22: 185-202.
- Hoshino, T. and K. Kuriyama, 2010, "Measuring the Benefits of Neighbourhood Park Amenities: Application and Comparison of Spatial Hedonic Approaches", *Environmental and Resource Economics*, 45: 429-444.
- James, G., D. Witten, T. Hastie, and R. Tibshirani, 2013, *An Introduction to Statistical Learning*, New York: Springer.
- Jim, C. Y. and W. Y. Chen, 2009, "Value of Scenic Views: Hedonic Assessment of Private Housing in Hong Kong", *Landscape and Urban Planning*, 91: 226-234.
- Joseph, A. E. and D. R. Phillips, 1984, *Accessibility and Utilization: Geographical Perspectives on Health Care Delivery*, New York: Harper and Row.
- Kanuganti, S., A. K. Sarkar, and A. P. Singh, 2016, "Evaluation of Access to Health Care in Rural Areas Using Enhanced Two-Step Floating Catchment Area (E2SFCA) Method", *Journal of Transport Geography*, 56: 45-52.
- Kelleher, J. D., B. Mac Namee, and A. D'Arcy, 2015, *Fundamentals of Machine Learning for Predictive Data Analytics*, Cambridge: MIT Press.
- Kim, Y., Y. J. Byon, and H. Yeo, 2018, "Enhancing Healthcare Accessibility Measurements Using GIS: A Case Study in Seoul, Korea", *Plos One*, 13: e0193013.

- Landis, J., S. Guhathakurta, and M. Zhang, 1994, "Capitalization of Transit Investments into Single-Family Home Prices: A Comparative Analysis of Five California Rail Transit Systems", *Working Paper*, No. 246, University of California Transportation Center.
- Langford, M., R. Fry, and G. Higgs, 2012, "Measuring Transit System Accessibility using a Modified Two-Step Floating Catchment Technique", *International Journal of Geographical Information Science*, 26: 193-214.
- Leishman, C., 2001, "House Building and Product Differentiation: An Hedonic Price Approach", *Journal of Housing and the Built Environment*, 16: 131-152.
- Li, F., 2010, "The Information Content of Forward-Looking Statements in Corporate Filings -- A Naïve Bayesian Machine Learning Approach", *Journal of Accounting Research*, 48: 1049-1102.
- Love, D. and P. Lindquist, 1995, "The Geographical Accessibility of Hospitals to the Aged: A Geographic Information Systems Analysis within Illinois", *Health Services Research*, 29: 629-651.
- Lovett, A., R. Haynes, G. Sünnerberg, and S. Gale, 2002, "Car Travel Time and Accessibility by Bus to General Practitioner Services: A Study Using Patient Registers and GIS", *Social Science & Medicine*, 55: 97-111.
- Luo, W. and Y. Qi, 2009, "An Enhanced Two-Step Floating Catchment Area (E2SFCA) Method for Measuring Spatial Accessibility to Primary Care Physicians", *Health & Place*, 15: 1100-1107.
- Luo, W. and F. Wang, 2003, "Measures of Spatial Accessibility to Health Care in a GIS Environment: Synthesis and a Case Study in the Chicago Region", *Environment and Planning B: Planning and Design*, 30: 865-884.
- Machin, S., 2011, "Houses and Schools: Valuation of School Quality through the Housing Market", *Labour Economics*, 18: 723-729.
- Mimis, A., A. Rovolis, and M. Stamou, 2013, "Property Valuation with Artificial Neural Network: The Case of Athens", *Journal of Property Research*, 30: 128-143.
- Mullainathan, S. and J. Spiess, 2017, "Machine Learning: An Applied Econometric Approach",

- Journal of Economic Perspectives*, 31: 87-106.
- Overby, J. W. and E. J. Lee, 2006, "The Effects of Utilitarian and Hedonic Online Shopping Value on Consumer Preference and Intentions", *Journal of Business Research*, 59: 1160-1166.
- Özsoy, O. and H. Şahin, 2009, "Housing Price Determinants in Istanbul, Turkey", *International Journal of Housing Markets and Analysis*, 2: 167-178.
- Panduro, T. E. and K. L. Veie, 2013, "Classification and Valuation of Urban Green Spaces—A Hedonic House Price Valuation", *Landscape and Urban Planning*, 120: 119-128.
- Park, B. and J. K. Bae, 2015, "Using Machine Learning Algorithms for Housing Price Prediction: The Case of Fairfax County, Virginia Housing Data", *Expert Systems with Applications*, 42: 2928-2934.
- Peng, Z. R., 1997, "The Jobs-Housing Balance and Urban Commuting", *Urban Studies*, 34: 1215-1235.
- Peng, T. C., 2019, "Does the School Input Quality Matter to Nearby Property Prices in Taipei Metropolis?", *International Journal of Housing Markets and Analysis*, 12: 865-883.
- Peng, T. C. and Y. H. Chiang, 2015, "The Non-Linearity of Hospitals' Proximity on Property Prices: Experiences from Taipei, Taiwan", *Journal of Property Research*, 32: 341-361.
- Pérez-Rave, J. I., J. C. Correa-Morales, and F. González-Echavarría, 2019, "A Machine Learning Approach to Big Data Regression Analysis of Real Estate Prices for Inferential and Predictive Purposes", *Journal of Property Research*, 36: 59-96.
- Plakandaras, V., R. Gupta, P. Gogas, and T. Papadimitriou, 2015, "Forecasting the US Real House Price Index", *Economic Modelling*, 45: 259-267.
- Radke J. and L. Mu, 2000, "Spatial Decomposition, Modeling and Mapping Service Regions to Predict Access to Social Programs", *Geographic Information Sciences*, 6: 105-112.
- Ready, R. C. and C. W. Abdalla, 2005, "The Amenity and Disamenity Impacts of Agriculture: Estimates from a Hedonic Pricing Model", *American Journal of Agricultural Economics*, 87: 314-326.
- Rosen, S., 1974, "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition", *The Journal of Political Economy*, 82: 34-55.

- Sansone, D., 2019, "Beyond Early Warning Indicators: High School Dropout and Machine Learning", *Oxford Bulletin of Economics and Statistics*, 81: 456-485.
- Sirmans, S., D. Macpherson, and E. Zietz, 2005, "The Composition of Hedonic Pricing Models", *Journal of Real Estate Literature*, 13: 1-44.
- Song, H. S. and M. Wilhelmsson, 2010, "Improved Price Index for Condominiums", *Journal of Property Research*, 27: 39-60.
- Storm, H., K. Baylis, and T. Heckelei, 2020, "Machine Learning in Agricultural and Applied Economics", *European Review of Agricultural Economics*, 47: 849-892.
- Varian, H. R., 2014a, "Big Data: New Tricks for Econometrics", *Journal of Economic Perspectives*, 28: 3-28.
- Varian, H. R., 2014b, "Beyond Big Data", *Business Economics*, 49: 27-31.
- Wan, N., F. B. Zhan, B. Zou, and F. Chow, 2012, "A Relative Spatial Access Assessment Approach for Analyzing Potential Spatial Access to Colorectal Cancer Services in Texas", *Applied Geography*, 32: 291-299.
- Wang, F., 2006, *Quantitative Methods and Applications in GIS*, London: Taylor & Francis Group.
- Wang, F. and W. Luo, 2005, "Assessing Spatial and Nonspatial Factors for Healthcare Access: Towards an Integrated Approach to Defining Health Professional Shortage Areas", *Health & Place*, 11: 131-146.
- Wang, G., J. Ma, and S. Yang, 2014, "An Improved Boosting Based on Feature Selection for Corporate Bankruptcy Prediction", *Expert Systems with Applications*, 41: 2353-2361.
- Wen, H., Z. Gui, L. Zhang, and E. C. M. Hui, 2020, "An Empirical Study of the Impact of Vehicular Traffic and Floor Level on Property Price", *Habitat International*, 97: 102132.
- Wolverton, M. and J. Senteza, 2000, "Hedonic Estimates of Regional Constant Quality House Prices", *Journal of Real Estate Research*, 19: 235-253.
- Yoo, S., J. Im, and J. E. Wagner, 2012, "Variable Selection for Hedonic Model Using Machine Learning Approaches: A Case Study in Onondaga County, NY", *Landscape and Urban Planning*, 107: 293-306.

Does Healthcare Accessibility Matter to Housing Prices? Application of Machine- Learning Decision Tree and Random Forest

Ti-Ching Peng*

Abstract

As Taiwan gradually turns into aged society, it highlights the importance of healthcare accessibility (spatial isolation between healthcare demanders and providers). This study applied enhanced two-step floating catchment method (Luo and Qi, 2009) to estimate elders' accessibility to three types of medical resources: physicians, hospital beds and ambulances. Then machine-learning decision tree and random forest regressions are adopted to examine the impact of healthcare accessibility on housing prices in Taipei Metropolis, Taiwan.

Random forest algorithm, as an ensemble of numerous decision trees, performed better in overall model prediction. Both machine-learning algorithms revealed that three types of medical accessibility for elders have significant effects on housing prices. Given that average life expectancy tends to be much higher in Taipei Metropolis than other regions, to elders in need of medical resources for chronic diseases in daily lives, the physician and hospital bed are proven to be more important than ambulance (needed for emergency treatment). These results

* Department of Real Estate and Built Environment, National Taipei University, Taiwan. Email: tcpeng@mail.ntpu.edu.tw.

should provide reference for government institutions to optimally allocate medical resources and maximize elders' living quality.

Keywords: Medical Service Accessibility, Elders, Decision Tree, Random Forest

JEL Classification: I11, O30, R21, R30