



AI於財稅之應用- 以租屋市場查核為例

112.03.15

主計業務創新精進研討會

財政部統計處

簡報 大綱

一

前言

二

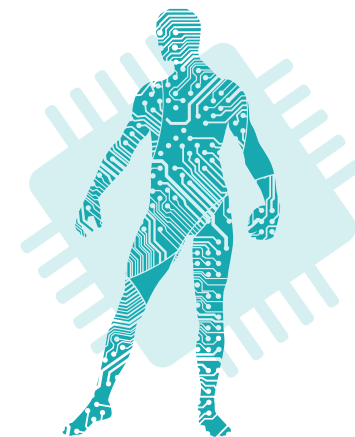
資料蒐集與處理

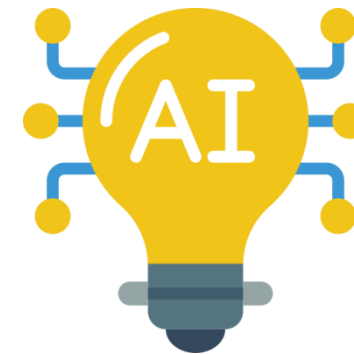
三

實作成果分享

四

結論與建議



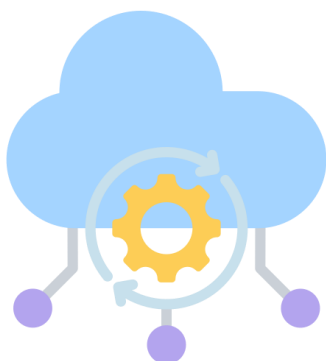


~ 運用人工智慧技術提升稅務行政效能作業計畫 ~

資料科學工作坊

AI PaaS

(人工智慧雲端運算服務)



國稅局

批發零售業短漏開發票
補習班案件其他所得
擴大書審制度分散營業收入

財政資訊中心

....
網紅資料探勘與收入模型分析

賦稅署

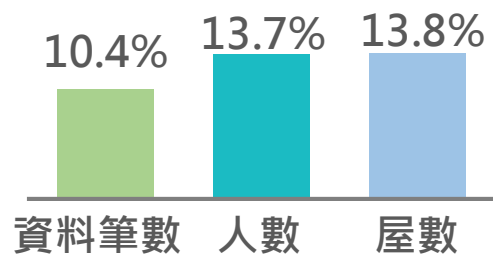
精進營業稅預算編列

統計處

個人租賃所得查核

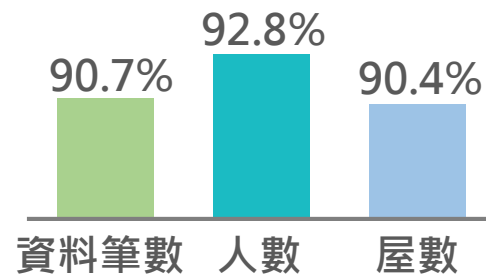
現行租屋所得資料侷限於特定來源

108年租賃所得中
租客屬自然人比重



1成租客屬自然人

108年租賃所得中
屬非扣繳資料比重



9成非主動申報

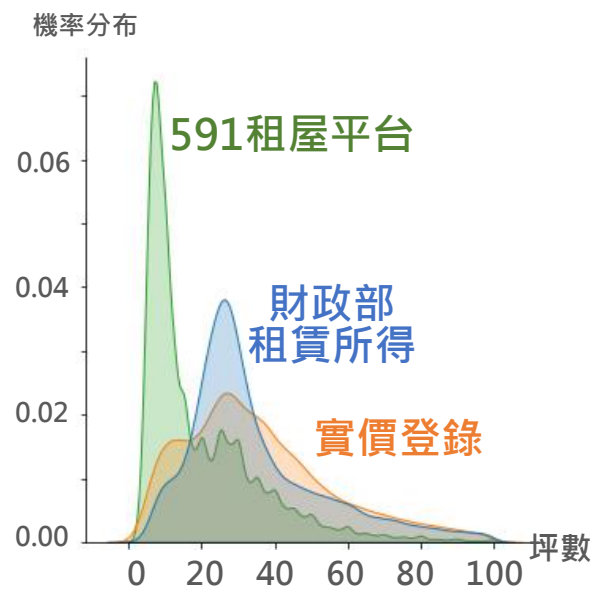
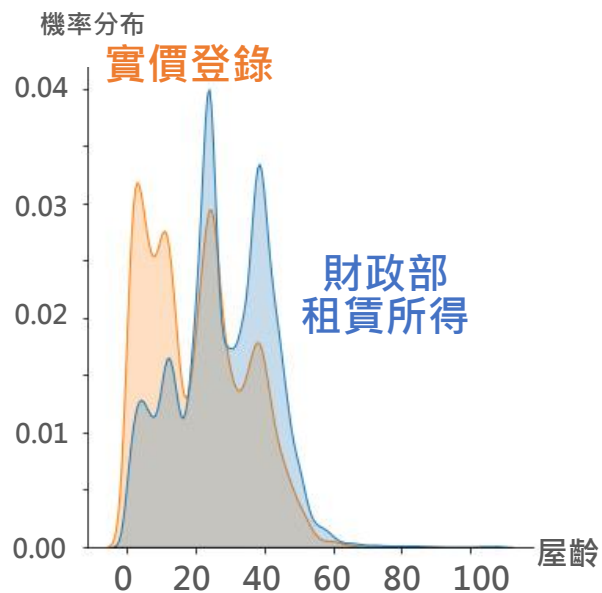
現行非扣繳資料來源

- 租金補貼
- 法院公證
- 營業人交查

非典型租屋者

現行租屋所得資料與外蒐資料態樣不同

排名\類別	財政部租賃所得	實價登錄	591平台
1	新北市板橋區	臺北市中山區	臺中市西屯區
2	臺北市中山區	新北市淡水區	臺中市北區
3	新北市中和區	桃園市桃園區	臺北市中山區
4	臺北市士林區	桃園市中壢區	桃園市桃園區
5	桃園市桃園區	臺北市大安區	桃園市中壢區



地區差異

屋齡差異

坪數差異

廣納變數建模

以現行資料，
利用跨稅及
外部變數
建立預測模型

$$Y = f(x_1 + x_2)$$

擴充查核範圍

比對實價登錄
與租屋平台
擴充現有清單

$$Y_1 \dots Y_n$$

房屋租賃相關資料

實價登錄租屋、民間租屋交易平台(591)
衛生部醫療院所、教育部各級學校

財產估算相關資料

實價登錄房屋交易、內政部公告地價占市價比重、
上市櫃股票價格、中古車交易網站

財稅與戶政資料

房屋稅稅籍檔、財產歸戶檔
戶政資訊檔、綜所稅結算申報核定檔

車輛財產價值估算

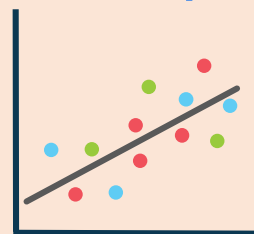
工人智慧



汽車：808品牌
機車：173品牌

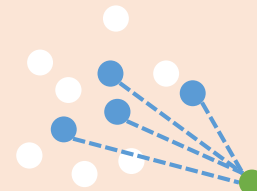
迴歸方程式

$$Y = \alpha + \beta X$$



廠牌
排氣數
出廠年份

整併分群

輪廓係數
(Silhouette
Coefficient)

汽車：63品牌
機車：23品牌



TESLA



模型概念

$$Y = f(x_1 + x_2 + x_3 + \dots)$$

房屋稅籍

出租
 $Y = 1$ 

=

房屋特徵



+

坐落區位特徵



+

屋主特徵

財產及所得
Lag1~2期!!未出租
 $Y = 0$ 

模型架構

輕量化梯度
提升模型

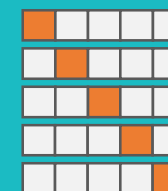


以103~107年
逐年訓練測試



分層交叉
驗證

Stratified
Kfold=5



各年度模型前20大重要特徵值

房屋自身特徵

- 屋齡*
- 非住家用非營業用面積*
- 代表卡序占房屋稅籍面積比重*
- 房屋稅籍面積*
- 住家用面積*
- 營業用減半面積
- 營業用面積

坐落區位特徵

- 房屋坐落村里之房屋稅籍數量*
- 房屋坐落村里之醫療院所數量*
- 房屋坐落村里之教育機構數量
- 房屋坐落村里之高中數量
- 房屋坐落村里之幼兒園數量

屋主特徵

- 屋主持有該村里租賃房屋數比重(Lag1)*
- 屋主持有房屋數量
- 屋主持有純住家用房屋數量
- 屋主持有房屋總市值(Lag1)
- 屋主持有該村里房屋數比重*
- 屋主房屋租賃所得占總所得比重(Lag1)
- 屋主持有該村里租賃房屋數比重(Lag2)
- 屋主持有土地總市值(Lag1)

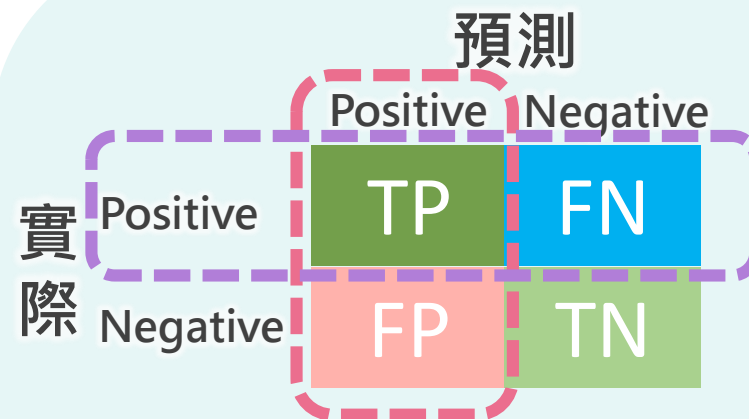
說明：標示*者為各年度前5大重要特徵值

模型解釋效力

指標\年	103	104	105	106	107
ROC AUC	0.91	0.91	0.89	0.90	0.91
Accuracy	0.99	0.99	0.99	0.99	0.99
Precision	0.48	0.70	0.57	0.59	0.65
Recall	0.41	0.34	0.42	0.34	0.36
F1	0.44	0.46	0.48	0.43	0.46

Threshold
= 0.5

精確率5至7成



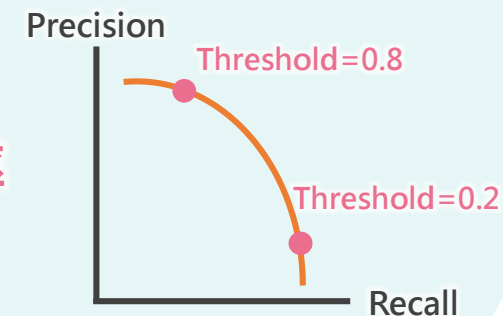
Precision
精確率

預測房屋出租中，
實際為出租的比率
開稅單後，命中比率

Recall
召回率

實際出租房屋中，
預測為出租比率

能掌握到整體
須課稅案件比重

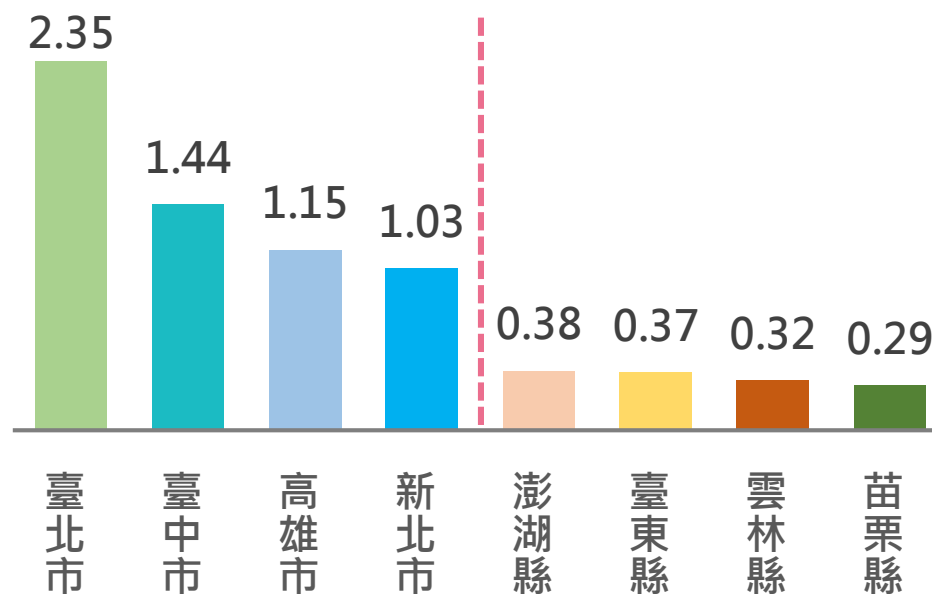


衡量模型效力指標考量因素-資料不平衡

年度	無出租房屋數	出租房屋數	比重
103	887萬	11.1萬	1.24%
104	898萬	8.5萬	0.94%
105	906萬	9.2萬	1.00%
106	914萬	9.1萬	0.98%
107	920萬	9.0萬	0.97%
108	925萬	9.5萬	1.01%

標記資料比重僅1%

各縣市歷年出租房屋數比重(%)



縣市間標記比重差異大

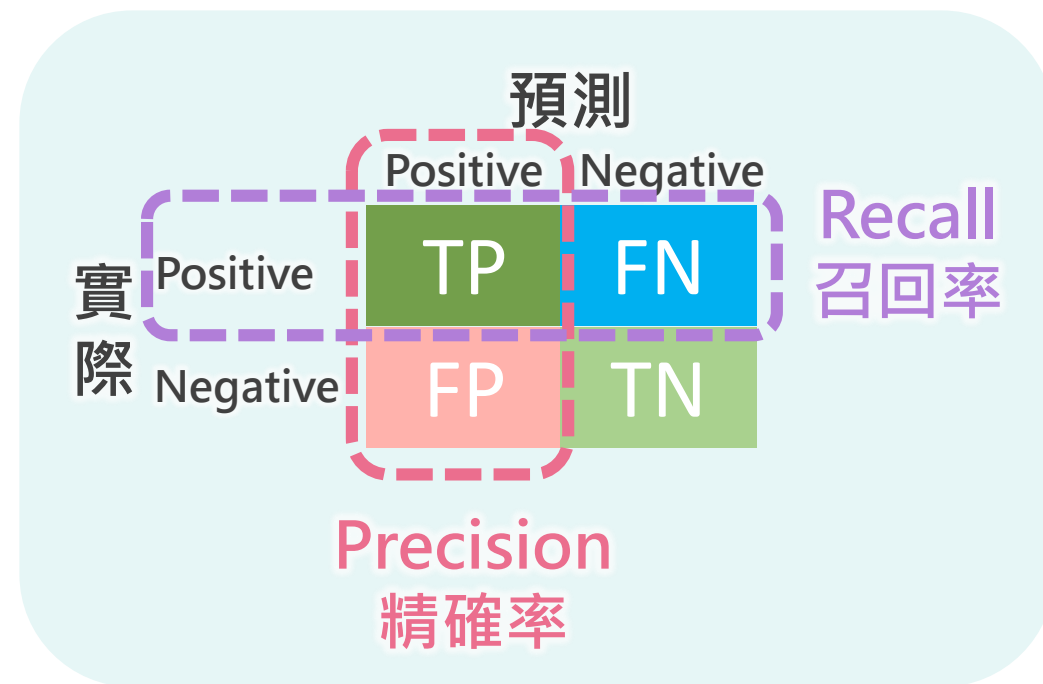
衡量模型效力指標考量因素-問題關注點

Recall
召回率

沒猜到影響重大
產線瑕疵品偵測、流感疫苗接種

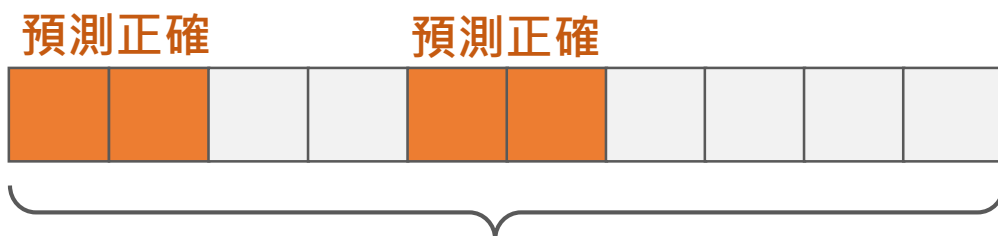
Precision
精確率

猜不對影響重大
疾病檢驗、驗孕測試、發單課稅



Precision at K

機率最高之前k筆資料當中，正確預測比率
the proportion of recommended items in
the top-k set that are relevant.



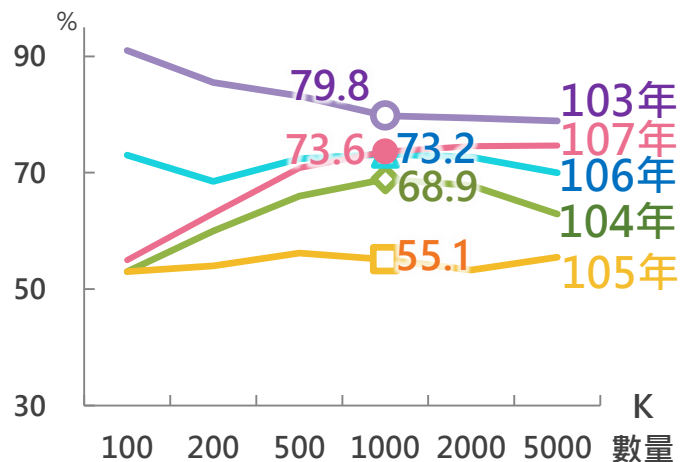
依模型預測機率高至低排序

$$P@10=0.4$$

機率最高10筆資料中
預測正確比率為0.4

模型解釋效力

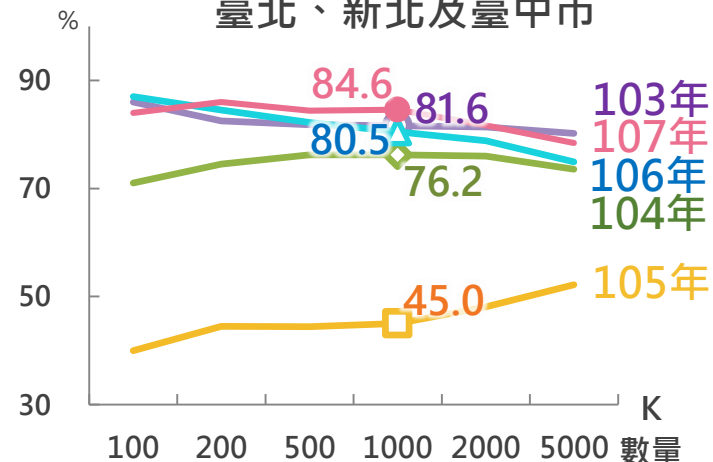
Precision @ K



前千筆精確率達7成


Precision @ K

臺北、新北及臺中市



限縮縣市提升至8成

將外蒐資料轉為潛在查核清單

 內政部
不動產交易實價查詢
服務網
(實際有租屋)

591 租屋
(潛在有租屋)



房屋稅
稅籍檔



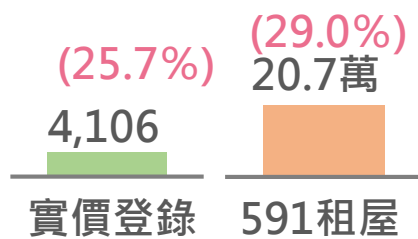
$Y_1 \dots Y_n$

外蒐資料比對做法

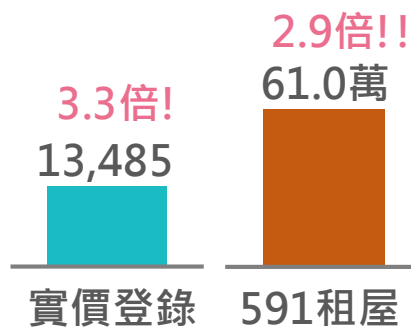
欄位	實價登錄資料內容
交易日期	107年11月1日
地址	臺北市中山區新生北路二段62巷1~30號
建物類型	住宅大樓(11層以上含有電梯)
樓層	12層
總樓層	總樓層數19樓
月租金	7萬元
起建日期	105年3月22日
建物坪數	135.68平方公尺
格局	3房2廳2衛浴
建材	鋼骨鋼筋混凝土造
家具	有
...	...

外蒐資料比對結果

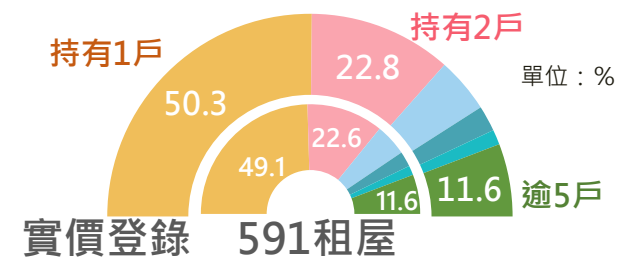
108年
成功比對筆數



108年
房屋稅籍對照筆數



108年
房屋稅籍對照筆數
屋主持有房屋數量



約3成比對成功

清單膨脹2至3倍

5成僅持有1戶

三 實作成果分享

現行租賃 資料來源

- 租金補貼
- 法院公證
- 營業人交查

$$Y_1 = f(x)$$

非自住住家用 房屋專案查核

- 持有逾10戶
- 持有逾5戶

$$Y_2 = f(x)$$

實價登錄 租屋



$$Y_3 = f(x)$$

民間租屋 交易平台

591租屋

$$Y_4 = f(x)$$

廣納不同租賃行為態樣，精進預測模型效力

- 現有模型預測精確率約7成，惟租屋型態較偏頗，**未來是否準確預測，仍有待驗證**。刻正就持有逾10戶房屋者租賃所得建模，並新納入捷運、公車等交通便利性及生活機能等參數，期能持續提升模型預測效能。
- 本部現已就持有多屋者辦理專案查核。惟就實價登錄與房屋稅籍比對結果顯示，其屋主屬**持有1戶者**逾半數，**似可列入選案範圍，以填補目前查核之缺口**。

- 本部已將**實價登錄**資料納入蒐集範圍，未來將**建立定期資料交換機制**，有助強化個人間租賃案件選查。
- 外部資料比對仰賴地址與相關特徵，未來**可借助內政部地址單一識別碼功能**，縮減外部租賃比對清單成本，提升稅務同仁之查核效率。