

摘要

隨者每年來台工作的移工人數逐漸上升，移工逃逸比例居高不下，儼然已成為一個急需解決的社會問題。並且中央移工資料庫過於陽春、台南市政府勞工局工作繁重，迫切需要一個方便的移工管理系統。本研究旨在導入AI機器學習技術對台南地區的移工逃逸問題進行深入分析，首先預測移工的逃逸風險：除了利用機器學習資料分析方法，也結合地理資訊，找出造成移工逃逸的關鍵因素。再者開發台南市移工管理平台、建立移工資料庫：包含整體的統計表格、結合地圖的地理數據可視化、高風險移工的預警系統，期望能夠即時掌握移工的動向，作為勞工局的施政助手。

研究動機與研究問題

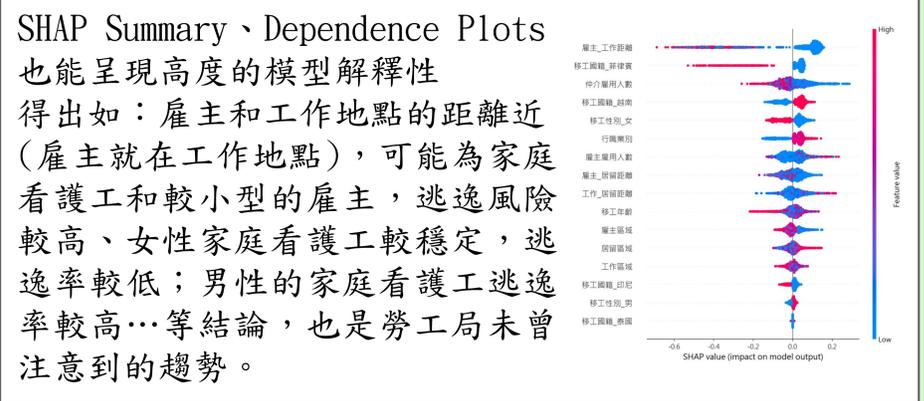
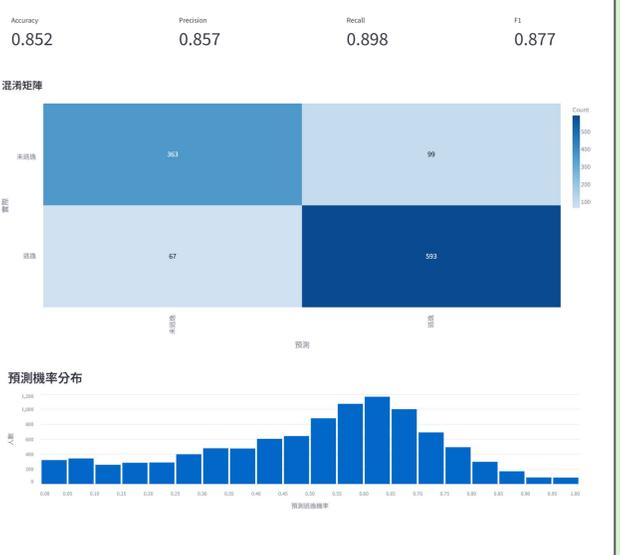
目前，勞工局對於移工問題的處理方式是接獲勞工或是雇主的申訴、或是移工逃逸以後才被動的去處理，缺乏防範未然的能力。而且，台南市勞工局僅依靠人員手動登打excel紀錄與中央統整資料，中央的移工資料系統僅有簡單的資料匯出功能，缺乏分析、統計功能與歷史資料的檢視，手打資料也常有缺漏重複，且局處人員較無資訊相關背景，統計也只依賴excel與手動計算進行，因而需要一套自動化工具。

研究方法

因本計畫涉及個資法問題(如移工姓名、居留證號等)，故資料先委託台南市政府智慧發展中心進行個資的掩蓋與替換成編號、地址留下○○路、街，無門牌號碼，故所使用的資料無違法之虞。本計畫首先統整勞工局提供的資料來源，進行資料前處理整理多份表格、合併、刪除重複項、補齊缺失欄位等，同時使用python的pandas、numpy等工具開發自動化工具。再者，嘗試許多機器學習模型：SVM、Random Forest、LightGBM，最後選定XGBoost與使用GridSearchCV調整參數預測移工的逃逸風險。接續，使用SHAP與LIME進行模型解釋，了解導致移工逃逸的因素，以及解讀單一移工為何被判定為高風險。其次，結合外部資料，如政府資料開放平台的GIS資料：醫院、診所位置、高齡化人口和勞動部勞動統計查詢網的私立就業服務機構評鑑結果。提高模型的準確度與解釋性。最後建置mysql資料庫，並且提供完整的分析報告、降低移工逃逸機率的建議，使用Streamlit開發圖形化介面網頁，建立一套完整的風險移工統計與預測系統。

研究結果

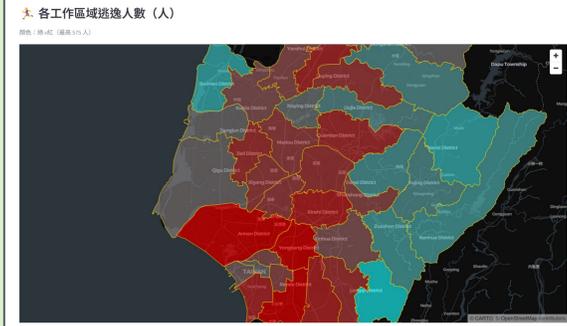
本研究的模型各項評估指標都已達0.85以上，於人流相關研究已經是相當高準確。對於預測資料集(尚在工作，不確定是否逃逸)預測機率分布約呈常態分佈，則可以統計學標準設置標準差產生高風險名單。



移工管理系統實作



主頁面 以表格方式列出現有資料 並提供篩選、檢索，此功能也適用後續頁面



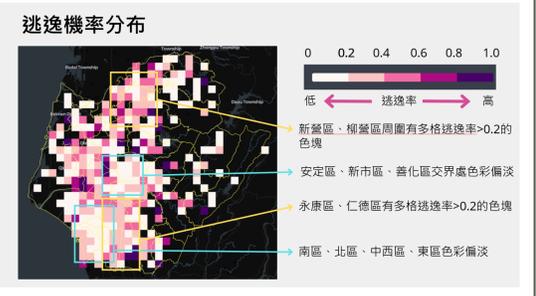
高風險雇主

雇主ID	雇主名稱	雇工人數	逃遷人數	逃遷比例%
M000001	...	236	78	33.05
M000002	...	100	54	54
M000003	...	34	32	94.12
M000004	...	49	30	61.22
M000005	...	17	17	100
M000006	...	14	14	100
M000007	...	13	13	100
M000008	...	11	11	100
M000009	...	20	17	85.00
M000010	...	11	11	100
M000011	...	16	16	100

高風險雇主 除了移工本身外，雇主也是重要的逃逸因素，此頁面顯示雇主與仲介過去的雇用紀錄，以及有無違法紀錄等



逃逸風險預測 除了提供高風險名單，也製作local的LIME和SHAP waterfall plot、global的feature importance和SHAP Dependence Plots，讓勞工局能了解整體逃逸趨勢與針對個案了解為何被判定為逃逸，從而做出對應的處理。並且保留模型訓練的部分，有新資料進入資料庫時可更新模型，也能選取時間範圍剔除舊資料，達到讓預測能與時俱進的目的。



地圖-方塊圖 有時區里界線對人流分析效果不佳，改用方塊圖可以更明顯看出都市與近郊的差異

結論與日後系統導入

本研究計畫與台南市政府勞工局與智慧發展中心合作，並已經有初步的研究成果報告予勞工局並且聽取回饋。於建立機器學習模型方面，已經找出適當、高準確率的模型以及與勞工局人員經驗相符且合理的結果，待日後有新的數據加入，反覆驗證後可正式用於預測。於移工管理系統方面，已經完成大部分，預計近期即可交予智慧發展中心進行個資處理的恢復、使用API接入中央移工資料系統或和勞工局確認資料導入方式，即可投入使用。