



資料科學程式設計 |

中華郵政大數據分析

組名 | 消失的章魚

代碼 | 140006

成員 | 國立台灣大學 土木工程研究所 交通組

蕭妤安 施安隆 張耀仁 俞又瑄

指導教授 | 國立臺灣大學 朱致遠教授



目錄

- 01 現況分析
- 02 創意設計理念
- 03 應用設計方法
- 04 效益與可行性評估
- 05 未來建議

An aerial photograph of a suburban area, showing a mix of residential housing, roads, and green spaces. The image is in grayscale, with a semi-transparent white box overlaid on the right side containing text.

01

現況 分析

在投遞失敗的案例中，收件者不在與拒收為主要失敗原因，根據特種郵件追蹤查詢資料(TT)107年第一季統計，失敗率可達**24.31%**，即17,982,933件。其中收件者拒收為不可抗力之因素，此處不予以討論；而收件者不在家之案例，主要分布在非商辦之住宅區，其中又以無大廈管理員之區域為主。在此以投遞郵局**高雄林園郵局(830024)**為例，此投遞郵局服務區域之特色包含：較少商業區土地使用、較少服務人數與每日只派一個班遞送，以利模擬成果呈現與理解。

An aerial photograph of a landscape. In the foreground, there's a road with a crosswalk. To the right, there are some industrial-looking buildings. The background shows a vast, flat landscape under a cloudy sky.

02

An aerial photograph of a landscape. In the foreground, there's a road with a crosswalk. To the right, there are some industrial-looking buildings. The background shows a vast, flat landscape under a cloudy sky.

創意設計理念

設計理念

資料管理 |

投遞失敗有諸多因素，而收件者不在為主要原因之一，為了使投遞成功率上升，透過歷年數據統計，加入了客戶投遞成功機率的因子，連結客戶與其收件之特性，了解在週間客戶收件成功的機率，以在路線規劃上加入隨機規劃的部分，增加投遞成功的機率。

隨機規劃 |

隨機規劃在路網規劃上主要加入機率的因子，但對於龐大的路網規劃會耗費多時，為了加速規劃時間，我們將考慮投遞貨物之投遞郵局、地理位置與投遞時間等因素加以分群，加入機率因子後再進行路線規劃，以最大化投遞的成功率，並最小化投遞的成本。

設計理念

路線視覺化與資料的數位化 |

將資料統計及規劃後，路線的視覺化才是連結規劃與投遞的重要橋梁。將規劃後的路線資料轉換到行動裝置所裝載的路線導航上，能夠快速有效地使投遞人員能夠了解投遞路線，數位化的路線規劃也能使工作交接或是轉換能夠更加順暢。而透過數位的簽收，能夠即時更新資料，也能使資料更加完善可靠，並快速加入規劃當中。

投遞失敗的處理 |

除了避免時間錯開之原因，透過大數據分析，也可以找出其他影響投遞失敗的原因，加以轉化為因子加入路線規劃中。對於遲遲無法成功投遞的客戶，也可以結合i郵箱的服務，透過不受限制的取貨時間，除了能夠增加成功投遞的機率、郵箱使用的推廣，也可以增加郵局與客戶之間的互動，將單方面的互動轉為雙向的流通，使物流系統整體運作能夠更加提升。

An aerial photograph of a landscape. In the foreground, there is a road with a crosswalk and a building with a flat roof. The background shows a vast, flat area with some distant structures and a cloudy sky.

03

應用設計方法

資料
處理

郵局端

未來資料庫 建議建置方向

目前由於個資法因素，中華郵政並未記錄每筆郵件資料之配送地址，因此本提案會透過資料處理推估顧客位置進行分析。本段為對於郵局端日後可朝向之資料蒐集調整方案之建議，以完整資料庫之建置。

特種郵件追蹤資料(TT)

- 現況：目前僅有紀錄郵件之狀態，無法得知狀態發生的地址與時間。
- 建議：可以透過會員制收集收件人之地址資訊，並提升現行寄送包裹和快捷郵差的PDA使用率以即時記錄收件人的取件時間與數位建檔。

四輪車輛GPS軌跡資料

1. 現況：目前GPS的資訊並無法得知車型、寄送貨品等資訊。
2. 建議：可以利用TT資料中的處理局號與GPS資料串接，以時間的紀錄點與該時間點的運送郵件內容推測車輛移動路徑的合理性。

資料
處理

規劃端

透過現有資料 推估位置與缺失值

1. 收件人位置：從107年第一季GPS資料中篩選出所有車機狀態顯示為「在外暫停」的紀錄，假設「在外暫停」為郵差正在為收信人服務的位置，建置該區域所有客戶推估之地址資料。為了排除郵差服務的對象相同，但每次暫停位置些許不同的情況，我們使用k-means分群找出最適合的分群數。
2. 每位客戶星期一至六之投遞成功率與失敗率：從107年第一季TT資料中可以找出該局號所負責區域所有客戶星期一至六之投遞成功與失敗件數，再將其每一封郵件採取均勻分布或是三角分布分給所有的收信人。
3. 收件人之間的距離時間成本：考量到道路的交通量、速限與蜿蜒程度，因此採用google map 的 direction API，計算所有點到點之間的距離和時間，而提案中分別計算上午10點和下午2點的資料，採取開車駕駛的模式並避開收費路段。

數據規劃 資料流程圖

資料處理

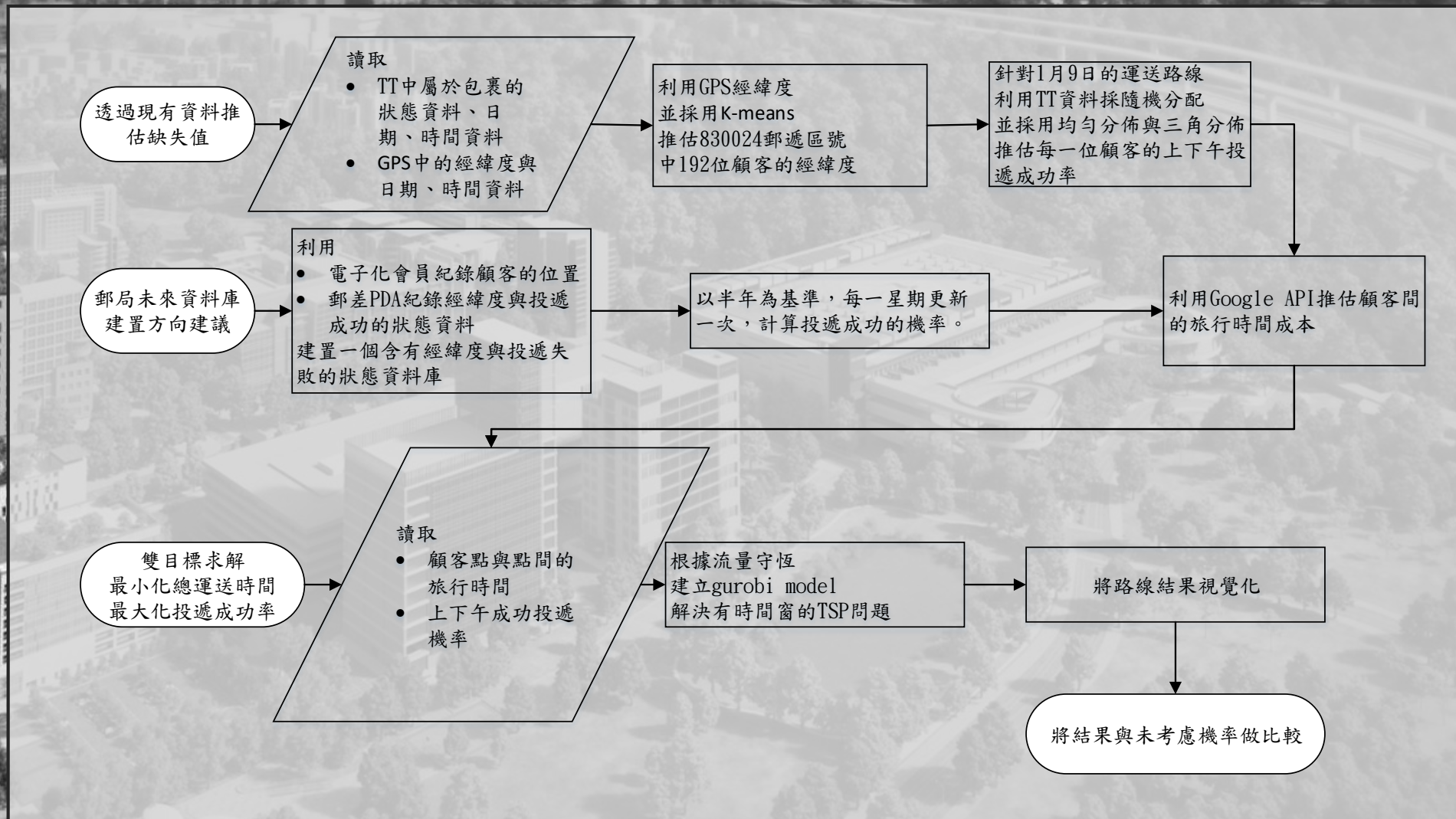
使用資料：TT, GPS
取用欄位：

- TT：包裹的狀態、日期、時間資料
- GPS：車輛經緯度、日期、時間資料

數據分析

使用方法：

- K-means分群
- Google API
計算旅行成本
- TSP旅行行銷員
問題透過MIP
作路徑規劃

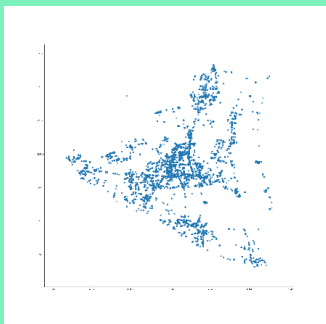


資料
處理

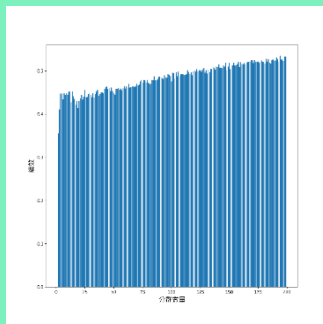
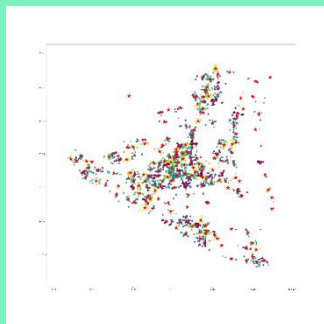
規劃端

以局號830024林園郵局作為案例分析－以107年1月9日為例

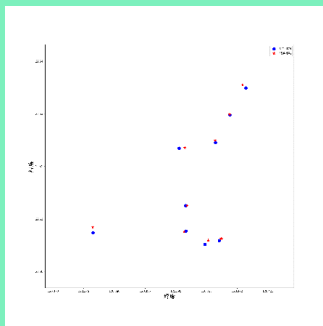
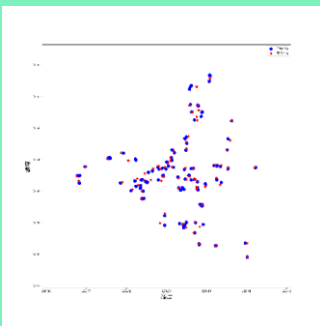
原始收件人位置分布



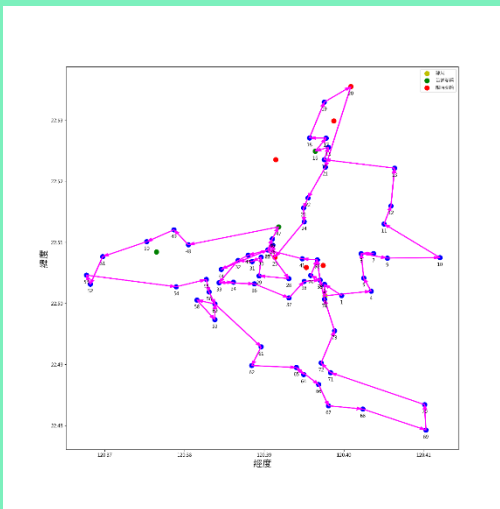
Kmeans分群之收件人位置
(紅色點)分布(左)及分群績效圖(右)



107年1月9日收信人座標轉換圖(左)
及郵筒座標轉換圖(右)



107年1月9日郵務車路線圖



在案例分析中，透過以下兩項資料處理轉換過程，以推算前段所列之所需資料，其運算結果列圖於左：

1. 判斷運送郵件為快捷或是包裹
利用TT資料中郵件的掛號號碼第13、14碼辨別當天郵件為快捷或是包裹的數量，按照兩者的比率隨機分配給當天所有的收信人。

2. 轉換實際收信人和郵筒座標位置
將當天所有GPS狀態顯示為「在外暫停」的點和郵筒的經緯度歸類到離假設收件人所在位置最近的點

路線
規劃

郵局端 / 規劃端

進行模型建置並將處理後資料輸入至模型中進行路線規劃

模型選用：旅行行銷員問題
(Traveling Salesman Problem, TSP)

在此問題中，郵局車輛的起訖點皆為當地的寄送郵局，一位客戶一天只配送一次，經詢問過確認車輛不會有返回取貨及容量限制的問題，另外，選擇的地區經過確認只由一輛車進行配送，判斷符合TSP問題的前提，因此採用此方法求解。

規劃目標：採雙目標規劃

最小化總旅行時間＋最大化投遞成功的機率

由於是雙目標問題，所以需要一個權重。目前先將權重定為0.1，希望還是主要由距離影響配送結果，另外會在結果比較權重為0及0.1造成的差別。

規劃結果與比較

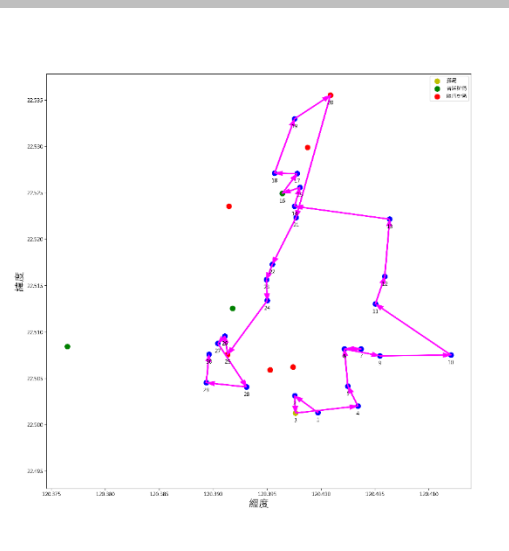
原始路線安排 |

原始路網：取01/09 郵務車行駛前30個點

旅行距離：19.9 km

旅行時間(不考慮服務時間)：58 min

當天遞送時間：9:49 - 11:29



此處設計三種不同目標之模式
比較其路線規劃之差異：

Model 1：僅考慮最短時間目標式
未考慮機率

Model 2：考慮最短時間目標式和
遞送成功率(三角分布)

Model 3：考慮最短時間目標式和
遞送成功率(均勻分布)

路線
規劃

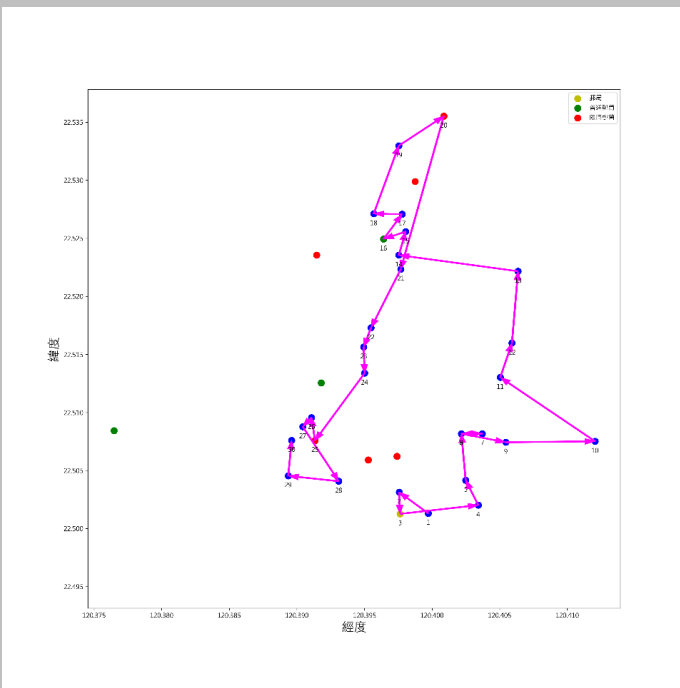
郵局端 / 規劃端

進行模型建置並將處理後資料輸入至模型中進行路線規劃

規劃結果與比較

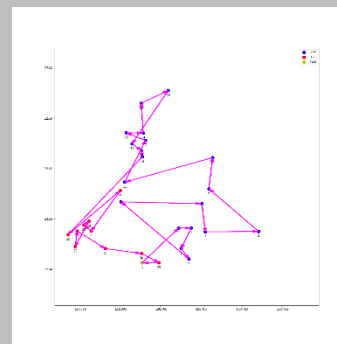
原始路線安排

旅行時間(不考慮服務時間)：58 min



Model 1

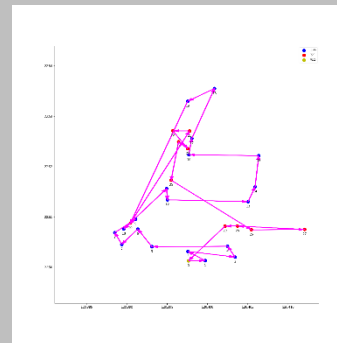
考慮最短時間目標式
未考慮機率



服務時間：10 min
機率權重：0
旅行時間：70min
(不包含服務時間)

Model 2

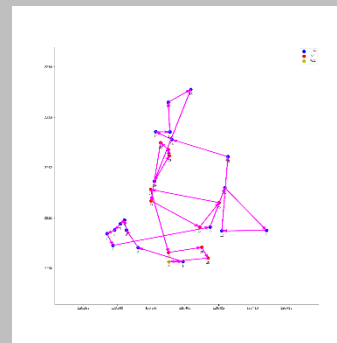
考慮最短時間目標式
和遞送成功率
每位收件人投遞成功
率採取三角分布



服務時間：10 min
機率權重：0.1
旅行時間：77min
(不包含服務時間)

Model 3

考慮最短時間目標式
和遞送成功率
每位收件人投遞成功
率採取均勻分布



服務時間：10 min
機率權重：0.1
旅行時間：70min
(不包含服務時間)

後續
處理

投遞失敗處理 i 郵箱補足 服務時間的缺口

為成功服務無法在遞送時間內簽收之顧客，可在配送系統中加入i郵箱之服務，將無法成功投遞之貨物放置至i郵箱，透過簡訊通知顧客取貨，藉由i郵箱無服務時間限制之特性，除了可以增加投遞的成功率，減少多次投遞成本，還可以有效推廣i郵箱之使用。除此之外，在顧客取貨時，可以透過電子屏幕簡單詢問顧客取貨失敗之原因，蒐集顧客回饋，以改善整體服務品質。

透過數據分析統計，可找出投遞成功率較低之區域，並將該區域之使用者組成及分析加入考慮，增設i郵箱之點位，提升使用者對於i郵箱使用之方便性，增加服務最後一哩路之廣度及深度。

後續
處理

投遞失敗處理 增加雙向互動

過去中華郵政服務顧客往往以單向服務為主要方式，較難獲得顧客回饋、了解顧客需求，因此如前述所言，未來可以朝向將服務電子化，如透過i郵箱電子屏幕進行簡易問卷調查，或是建立會員制並結合手機app，使顧客在接受服務當下可以立即給予回饋，除了可以增加整體服務品質，也可以增加郵局與客戶之間的互動，將單方面的互動轉為雙向的流通，使物流系統整體運作能夠更加提升。

模型優化

郵局端 / 規劃端

透過統計模型與機器學習

找出其餘影響之因素以優化模型

使用資料：TT, ACC
將TT中與投遞成功
(I4)與失敗(H4)有關
之郵件狀態，與ACC
以掛號號碼合併

使用方法：
- 相關係數矩陣
- 二元羅吉斯
回歸分析
- 機器學習

分析結果範例：

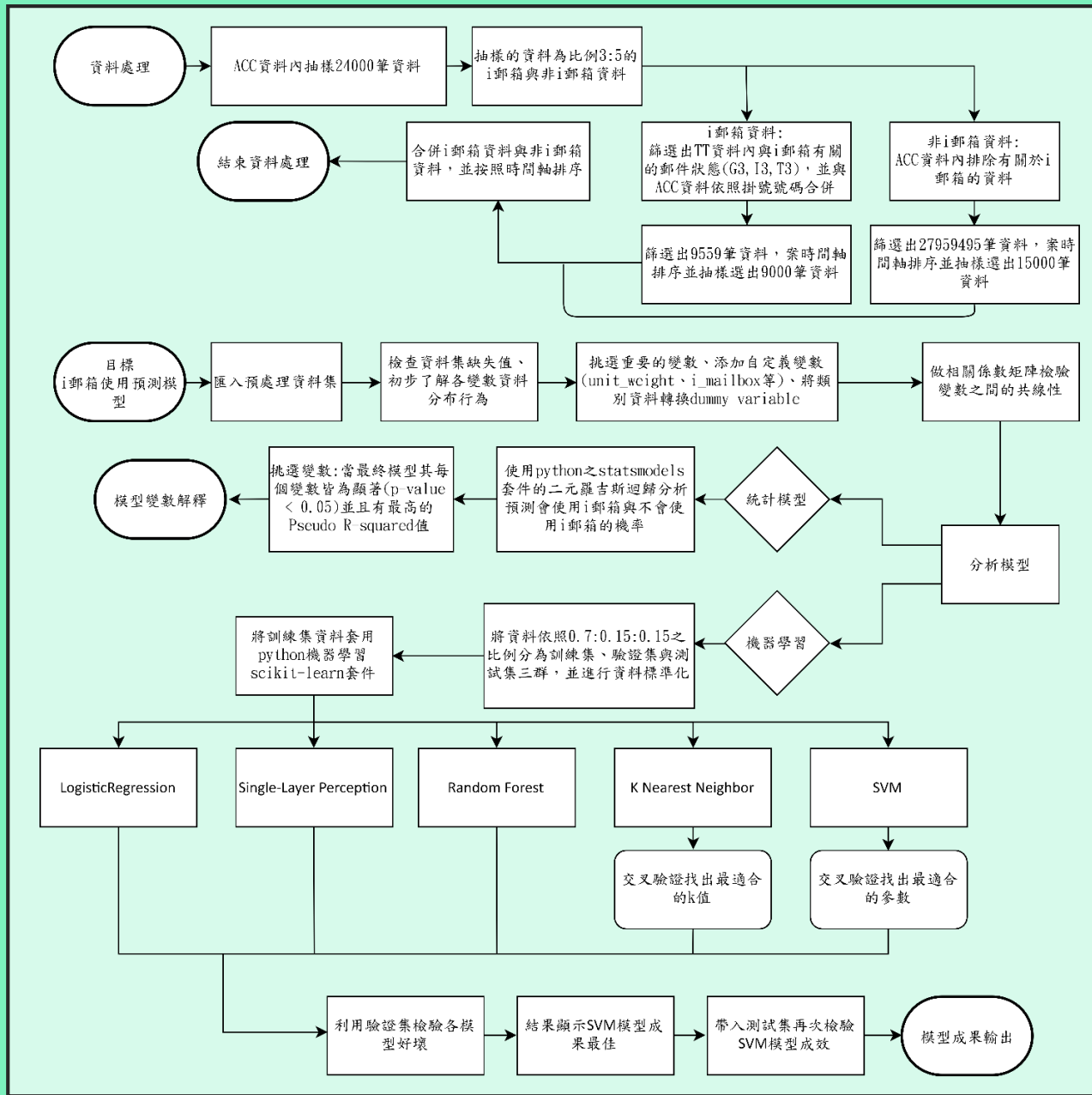
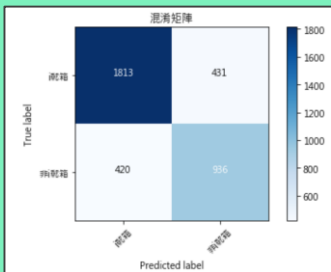
以預測是否使用 i 郵箱之分類模型及評估為例
註：此處使用初賽企畫書所使用之分析結果

統計模型

$$f(x) = \ln\left(\frac{P_{\text{使用i郵箱}}}{P_{\text{不使用i郵箱}}}\right) = -3.0185 + 0.8147 * \text{acc}29_5 + 2.6832 * \text{acc}29_4 + 2.4041 * \text{acc}29_3 + 0.3901 * \text{acc}29_1 - 2.595 * \text{acc}24_2 + 1.6279 * \text{acc}16_0 + 1.6536 * \text{acc}12_2 - 0.5564 * \text{acc}12_1 + 0.2854 * \text{officetime} - 2.5811 * \text{acc}21 - 0.0059 * \text{acc}32 + 0.6626 * \text{acc}2 - 0.00009957 * \text{unitweight}$$

機器學習

使用模式	Logistic Regression	Single-Layer Perception
準確率	0.7568	0.71
Random-Forest	K Nearest Neighbor	SVM
0.74	0.75	0.77



An aerial photograph of a landscape. In the foreground, there's a road with a crosswalk. To the right, there are some industrial or commercial buildings. The background shows a vast, flat area, possibly a field or a large parking lot, under a cloudy sky.

04

效益與可行性評估

效益評估

	不考慮投遞機率	投遞機率呈三角分佈	投遞機率呈均勻分佈
投遞成功機率 在目標式的權重	0	0.1	0.1
旅行時間 (不包含服務)	70	77	70
總旅行距離(公里)	26.4	27.7	27.3
成本上升(公里)	0	1.3	0.9
所有包裹投遞成功機率	0.268465909198	0.305702528295	0.287642045276
所有包裹投遞成功提升率	0	13.87%	7.14%

投遞成功機率在目標式的權重：用來衡量兩目標(旅行時間及投遞成功率)間的重要性，在此我們將投遞成功率的重要性降低，主要將目標放在最小化旅行時間。

旅行時間：車子行駛於服務點之間所需花費的總時間(利用google API推估而得)，不包含停下來服務的時間。

總旅行距離：車子行駛於服務點之間所行經的距離(利用google API推估而得)，不包含停下來服務的時間。

成本上升：相對僅考慮最小化旅行時間單一目標，雙目標使得成本上升多少行駛距離(利用google API推估而得)。

所有包裹投遞成功機率：根據該包裹到達該點為上/下午，乘上該點上/下午投遞成功的機率推估而得。

所有包裹投遞成功提升率：與單一目標式相對應的提升率。

可行性評估

經濟層面 可行性

此方案大部分的資訊可以透過郵局的大數據提取，因此在路線規劃時僅需要考量使用google map API求兩點之間距離的額外成本。由於郵局為大宗業務需要，因此可以採用企業優惠方案。

技術層面 可行性

有關於路線規劃的程式碼以及使用環境建置已完成。然而我們發現使用gurobi在時間限制下求解精確解的效率並不高，因此對於顧客點的數量有其限制。實際上在求解此類mip的問題，可以考量將model改成啟發式演算法(ex.插入法、基因演算法)藉此提高規劃效率。

問題規模層面 可行性

目前的路徑規劃有考量時窗限制，因此求解的時程表能涵蓋到郵務車的服務時間、午休時間與營業時間，因此我們可以彈性調整各項參數。此外，目前的model只考量單一路線的小規模郵局。針對大規模多路線的郵局可以考量使用VRPTW問題求解。

模型應用層面 可行性

目前郵差是依照固定路線運送郵件，而我們提供的路線行駛規劃，雖與郵差過往經驗不同，但能透過導航系統給予行駛建議，由郵差自行判斷是否要使用該建議路線。

05

未來 建議

- 資料收集 |

GPS資料可與郵件資料結合，建立共同的索引，以利後續追蹤與分析；此外郵件資料可以藉由會員制，增加地址項目以助於提升服務品質與投遞成功率；而數位化的簽收可以有助於即時的資料更新，完整後端資料的管理與後續資料的分析。

- 資料結合 |

結合未來郵局電子會員制，與郵差PDA資訊，可以在未來建置每位顧客上下午投遞成功機率的資料庫，有助於本方案的執行，進而提升投遞成功的機率，且針對投遞成功機率較低的區域可以納入未來設置i郵箱的規劃區。

An aerial photograph of a modern urban development, featuring a mix of high-rise and mid-rise buildings, green spaces, and a network of roads. A large, semi-transparent green triangular overlay covers the right side of the image, serving as a background for the text.

簡報結束
敬請指教