

# 《成果應用企劃書》

隊名：你愛我對不隊

組員：林祐陞、張泳樺、郭又嘉

指導老師：陳律閔教授

## 目錄

一、	作品名稱	3
二、	摘要	3
1.	提案動機	3
2.	市場分析-「i 郵箱」優勢	3
3.	問題分析	3
4.	願景	4
三、	大數據分析流程及結果	4
1.	資料處理	4
2.	模型架構	7
3.	模型訓練	8
4.	模型驗證	9
5.	模型結果	9
6.	模型優點	9
四、	創意應用及服務設計（效益評估、可行性分析）	10
1.	分辨潛在客戶	10
2.	潛在客戶據點分析	10
3.	Iber Mail	12
五、	結論與未來展望	13
六、	參考資料	13

一、作品名稱：用深度學習挖掘使用「i 郵箱」的潛在客戶

## 二、摘要

### 1. 提案動機

「i 郵箱」為郵局近年來大力推廣的一項發展重點，提供多元且便利的寄件與取件服務，並擴展至改投、多元支付的服務，積極轉型為智慧物流，智慧城市的一大象徵。

即便如此，在使用率上卻始終寥寥無幾，實際向民眾調查後，發現可歸納出兩大潛在客群並未使用「i 郵箱」的原因，一為住家附近沒有「i 郵箱」的存在，二為並不了解自己的包裹其實適合使用「i 郵箱」收寄的民眾。

為找出這些潛在客戶，我們以大數據深度學習的方式找出合適使用「i 郵箱」的包裹，繪製潛在客戶地區分佈圖，搭配現有的「i 郵箱」位置，以架設新的據點，並針對該潛在客戶進行「i 郵箱」的推廣，同時達到提升使用率的效果。

### 2. 市場分析 - 「i 郵箱」優勢

#### A. 交貨取貨不受限制

滿足新時代民眾的配送需求，不再受限於必須配合郵政上下班時段，同時亦可在某程度上降低人力成本，另外，取貨過程無需跟人接觸，民眾不用擔心購買隱私商品時的尷尬。

#### B. 多元的取件服務

提供更加便利多元的取件服務，民眾除了普通的收寄國內一般包裹、快捷及 3 號便利包之外，若經一次投遞不成功之國內包裹、快捷等掛號郵件，於收件人同意後，亦可改投「i 郵箱」3 天內自助取件服務，達到更加便利的改投服務。

#### C. 價格實惠、付款多元

在資費上僅介於 60 至 70 元間，可透過悠遊卡、一卡通、icash 等電子票證輕鬆支付。

#### D. 省去排隊時間

「i 郵箱」相對原先臨櫃辦理業務上來講，可達到分散人流的效果，使部分業務轉移至「i 郵箱」輔助辦理，民眾在投遞包裹上可不再花冗長排隊時間，亦同時可達到整體郵政業務效率的上升。

#### E. 智慧城市的象徵

「i 郵箱」相較於其他業者的智慧型物流 ATM 來說，並且和政府相關企業合作上有相對的優勢，例如在公共場合，如車站、學校、政府機關等地點設置新站相對容易。

### 3. 問題分析

#### A. 設置據點少且不符需求

目前「i 郵箱」因設點容易而大多設置於公共場所，然而較少深入民眾的生活中心，例如市場、公司附近，許多民眾即便包裹符合使用「i 郵箱」的特徵，也會

因為「不順路」而未選擇使用「i 郵箱」。

B. 推廣不足

因據點少導致民眾對「i 郵箱」的陌生，多數人還不明白如何使用與其優點，導致許多潛在客戶的流失。

C. 櫃位設計不符需求

目前「i 郵箱」有提供多種不同大小的櫃位，但每格只能放置一件包裹，且也無提供大型的包裹交寄，若需要寄送多件或大件包裹的使用者較於不方便。

4. 願景

A. 擴展 i 郵箱據點

在這次創意競賽上，我們將使用人工智慧深度學習從包裹資料中分辨出哪些為「i 郵箱」的潛在客戶，並根據此映射至地圖上，找出「i 郵箱」目前據點與潛在客戶分佈圖，以此做為「i 郵箱」新據點的設置依據。

B. 分辨潛在客戶

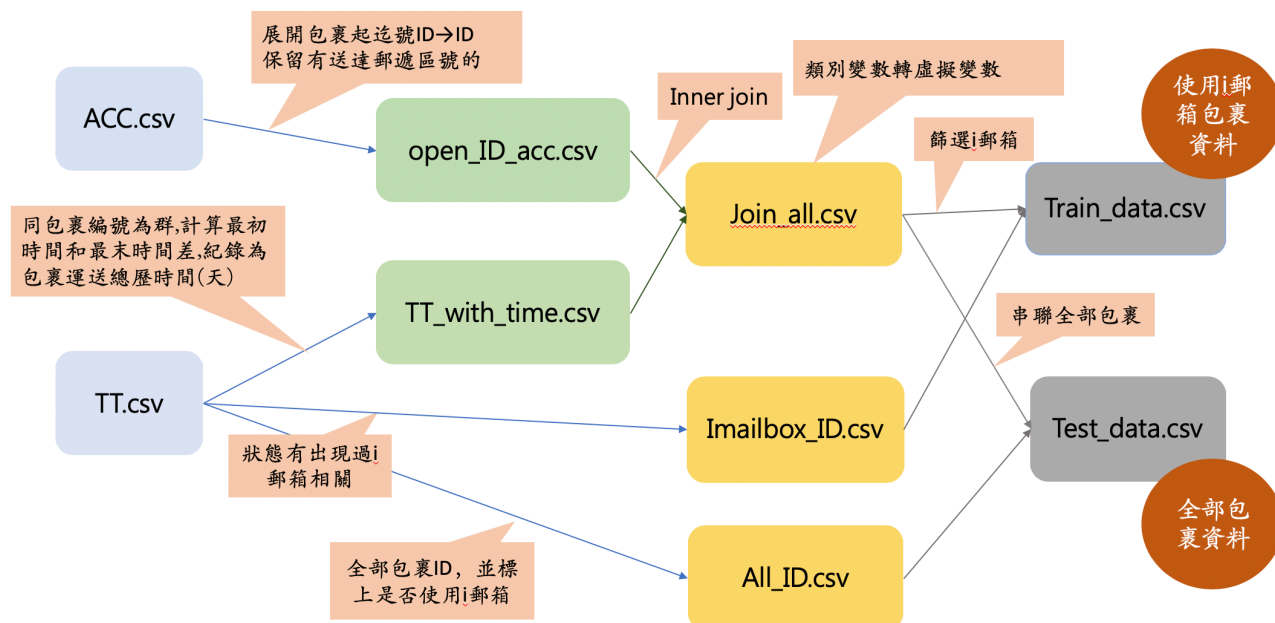
找出一個合適的模型，從特約客戶中找尋經常寄出合適使用「i 郵箱」包裹的潛在特約客戶，同時當民眾至臨櫃辦理寄件包裹時，馬上分辨是否合適使用「i 郵箱」寄件，再加以向這些潛在客戶推廣「i 郵箱」，提升使用率。

C. Iber Mail

設計一個 Iber Mail 的整合性手機應用軟件，結合「i 郵箱」潛在客戶推薦系統，並可事先輸入包裹寄件資訊，或是使用常用聯絡人資料快速一鍵代入資料，此外亦可隨時透過 Iber Mail 查看包裹即時資訊，使在收寄包裹使用體驗上更加便利流暢。

三、大數據分析流程及結果

1. 資料處理



## Step 1

在資料處理部分上主要使用收寄資料明細檔(ACC)、特種郵件追蹤查詢資料(TT)，首先將 ACC 中包裹編號(ID)由包裹起迄號(ACC33、ACC34)展開，並針對有送達郵遞區號的包裹進行分析，得到新的表格為 open\_ID\_acc。

交寄局號	...	本段件數	掛號號碼起號	掛號號碼迄號	...	收寄月份
100257	...	3	00037610025701	00037810025701	...	01
220067	...	1	67146800100470	67146800100470	...	01



將[包裹編號起號]與[包裹編號迄號]，擴展為每筆編號有自己的一系列數據

交寄局號	...	本段件數	包裹ID	...	收寄月份
100257	...	3	00037610025701	...	01
100257	...	3	00037710025701	...	01
100257	...	3	00037810025701	...	01
220067	...	1	67146800100470	...	01

圖二 ACC 與 TT 資料對應展開說明

## Step 2

將特種郵件追蹤查詢資料(TT)進行通包裹編號為群，計算最初紀錄的時間與最末時間的時間差，作為一新變數為包裹運送總歷時間(diff\_time)得到新的表格為 TT\_with\_time。

郵件狀態代碼	掛號號碼	處理日期	處理時間	處理局號	...
A1(交寄郵件,支局)	67146800100470	2018-01-02	11:03:00	220067	...
Z4(運輸途中)	67146800100470	2018-01-02	19:28:44	091814	...
Y4(郵件投遞中)	67146800100470	2018-01-03	08:41:53	220584	...
H4(投遞不成功)	67146800100470	2018-01-03	14:37:31	220584	...
Z1(郵件轉運中)	67146800100470	2018-01-03	17:53:38	600006	...
G3(到達i郵箱)	67146800100470	2018-01-04	10:41:01	220584	...

max 2018-01-04 10:41:01  
 - min 2018-01-02 11:03:00  
 2.35974

將[處理日期]合併[處理時間]，並以同個掛號號碼分群，計算[處理時間差]

掛號號碼	處理時間差(天)	...
67146800100470	2.35974	...
...	...	...

圖三 新增 diff\_time 變數說明

## Step 3

同時將特種郵件追蹤資料(TT)中，狀態出現過與「i 郵箱」相關狀態的包裹，進行篩選，分辨出包裹 ID 是否和「i 郵箱」相關，製作 Imailbox\_ID 的表格。

#### Step 4

將特種郵件追蹤資料(TT)中出現過的所有 ID 進行彙整為 All\_ID 表格，並一樣標明該 ID 是否有經過「i 郵箱」，之後將以此作為與 ACC 串聯時的依據，避開那些紀錄上資料的缺失。

#### Step 5

以 open\_ID\_acc 與 TT\_with\_time 兩表格進行以 ID 為鍵值的 Inner join 得到新表格。同時將大多數類別型變數進行虛擬變量的轉換，其中要注意的是不論該類別是否出現過，再這邊都先補上所有資料說明檔中的類別名稱，避免之後出現時無法使用，而這個表格名稱定為 Join\_table。



圖四 兩表格進行 Inner join 說明

#### Step 6

##### 訓練資料集

與剛剛的 Imailbox\_ID 再做一次篩選動作，篩出全是使用「i 郵箱」且含有新變數包裹運送總歷時間的表格。

##### 測試資料集

使用 Join\_all 和 All\_ID 進行串聯，得到全台資料，並含有標記是否經過「i 郵箱」的變數。

	ACC1	ACC2	ACC5	ACC6	ACC7	ACC15	ACC17	ACC19	ACC21	ACC22	...
ID											
00602860101020	600028	0	0	1070102	084338	0	0	0	0	1	...
00603760101020	600028	0	0	1070102	084338	0	0	0	0	1	...
23366300104170	400030	0	0	1070102	084717	0	0	0	0	1	...
04160410617018	100181	0	0	1070102	084719	0	0	0	0	1	...
92955230002410303002	300024	0	0	1070102	085045	0	0	0	0	1	...

圖五 最終輸出結果

## 2. 模型架構

- A. 目的：利用使用「i 郵箱」資料的特性，挖掘未使用「i 郵箱」包裹當中的潛在客戶。
- B. 資料特性：
- 郵政的資料中包含可能的情況太多，有使用「i 郵箱」的情況有可能包含只嘗試一次但不適用的包裹，而在未使用「i 郵箱」當中，存在適合使用「i 郵箱」，只因不熟悉而未使用的情況，所以在此標籤對於有無使用「i 郵箱」無法準確的區分，因此不使用二元分類，需要利用 One-class Classifier 模型歸納出使用「i 郵箱」的特性，再進行判斷。
  - 大部分的包裹變數皆為類別變數，且變數與變數之間可能存在著交互作用，因此將所有變數考慮其交互作用，並利用模型進行篩選。
- C. 使用模型：**One-class Classifier Least Square Autoencoder GAN (以 LSAGAN 簡稱)**
- D. 使用模型解釋：

### One-class Classifier

由於資料包含情況複雜，只關心使用「i 郵箱」的包裹特徵為何，所以針對單一目標類別進行訓練，希望從這裡面找出未來會長期使用「i 郵箱」的資料，也利用 huber loss 期望結果不受離群值的影響。

### 原始 GAN [註一]

由 Discriminator network(D)、Generator network(G)構成，G 利用真資料製造偽真的假資料，而 D 要分辨這個假資料的真假，並給出分數(接近 1 為真、接近 0 為假)。G 利用這個分數來「訓練」資料，直到最後偽真資料很接近真實，可讓 D 和 G 相互制衡。

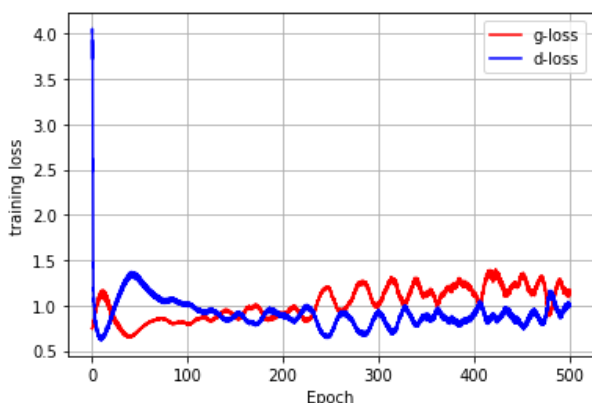
G 努力做出逼近真實的假資料  
D 努力分辨真實與假資料的差別

相互對抗

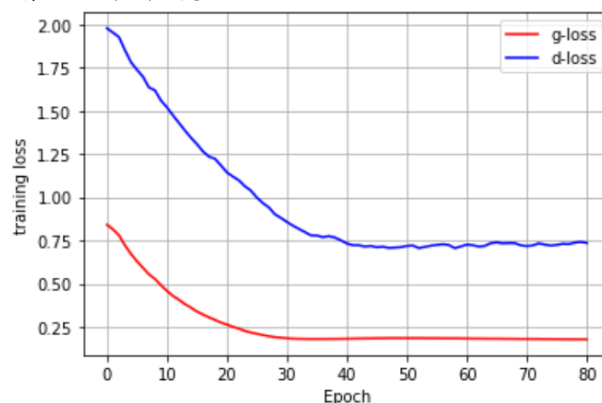
G 產生最真實的資料  
D 可精確分別真假 (此使用目的)

### 利用 LSAGAN 做改善 [註二]

改善 GAN 訓練多次後容易因為梯度的消失，模型無法持續精進的很好，而使製造出來的假資料即是被判斷為真也並不十分相像，而不易判斷停止的正確時間，因此由 Sigmoid cross entropy 替換 Least Square 損失函數，分數的設定改為接近 1 為真、接近 -1 為假，由圖七可看出模型較穩定。



圖六 GAN G 和 D Loss 比較 (不穩定)



圖七 LSAGAN G 和 D Loss 比較



### E. LSAGAN 模型詳細結構

#### Generator

在原始資料中裡包含希望找出未來會使用「i 郵箱」的資料，在摻雜一定比例只試用但未來不會再使用的資料中，利用 Autoencoder 搭建出 robust 的複製器。

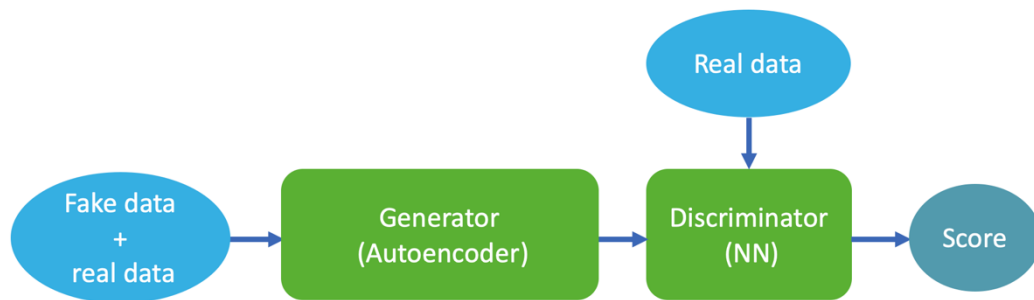
目的：找出未來會長期使用「i 郵箱」的資料與特徵。

#### Discriminator

I. 利用 NN 架構搭建 Discriminator，進行部分特徵取樣

II. 加入 L1 的懲罰項，歸納出重要的特徵

目的：藉由 Discriminator 辨別 Generator 生成的好壞，判斷是否可能為使用「i 郵箱」的潛在客戶。



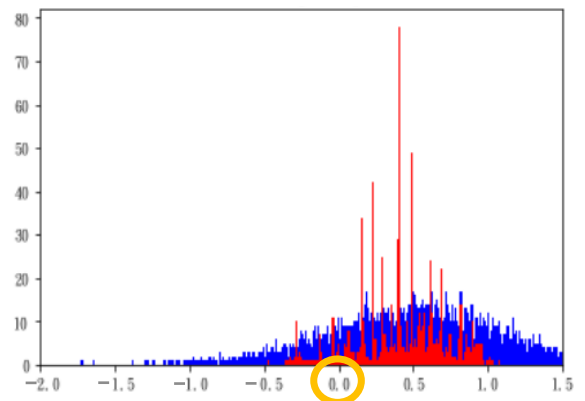
圖八 LSAGAN 簡易架構圖

### 3. 模型訓練：

利用現有使用「i 郵箱」的資料，歸類出包裹的相同的特徵，利用此模型架構可以判斷此包裹是否適合，更可以藉由模型推斷最適合的新「i 郵箱」據點。

- A. 利用使用「i 郵箱」的資料分為 train V.S validation。
- B. 模型輸出為一個 Score，分數越高判斷此包裹越適合使用「i 郵箱」，當作選取的門檻值。
- C. 利用模型檢驗出的結果，分數應為一個左偏的分配。
  - ⇒ 小於門檻值：視為此包裹可能為只使用一次「i 郵箱」，即為異常值。
  - ⇒ 大於門檻值：視為此包裹適合使用「i 郵箱」，即為穩定長期使用的包裹。
- D. 利用 validation data 代入判別器，形狀與 training data 無太大差異，可見模型無 Overfitting。

將所有使用「i 郵箱」的包裹利用 discriminator 判別器給定的分數畫出 score 圖，如圖九所示，藍色的部分為 training data，紅色的部分為 validation data，可發現兩者具有相同分佈，而且大於 0 的包裹佔大多數，符合我們的假設。

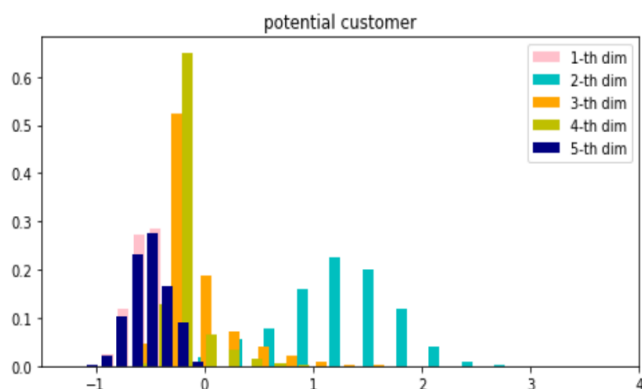


圖九 train/validation score 比較

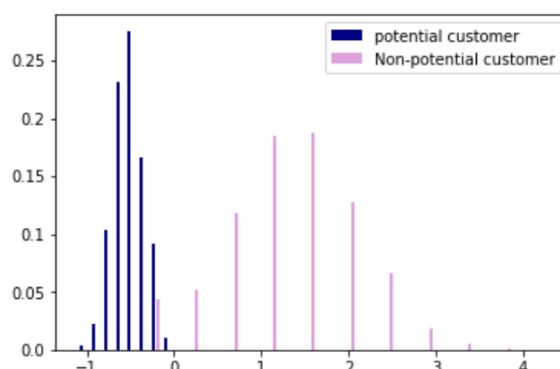


#### 4. 模型驗證：

在LSAGAN 中的 Generator 由 autoencoder 的 encoder 和 decoder 合成，因此藉由圖十證明在 training data 中，判別為潛在客戶的包裹在經過 encoder 後變為五維的向量（分別為五個不同的顏色），這五維所畫出的長條圖有各自不同的集中趨勢，而圖十一中拿出第五維特徵與非潛在「i 郵箱」比較，兩者趨勢不同，代表被判斷為長期使用「i 郵箱」的包裹，經由 encoder 轉換之後特徵有相同的特性，便能判定模型的成功。



圖十 潛在包裹 encoder 後的特徵分佈

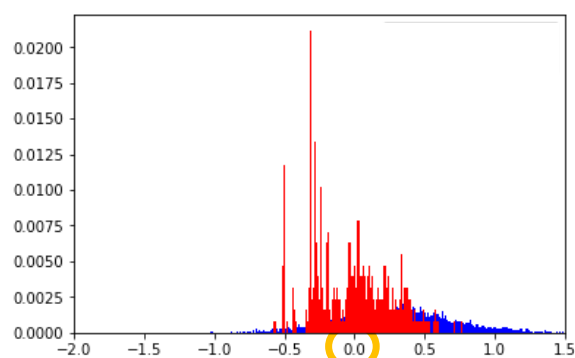


圖十一 潛在與非潛在包裹第五維的特徵比較

#### 5. 模型結果：

在圖十二中，藍色為全部使用「i 郵箱」的資料，而紅色為未使用「i 郵箱」的資料，將其放入訓練好的模型，可發現判斷後的分數依然界在 0 附近，利用模型的特性，將分數大於 0 的包裹判斷為潛在客戶。

最後模型所判斷出的結果為大致有 63.3 % 為適合使用「i 郵箱」的潛在客戶，而在圖十三中可看到所判斷出來的潛在客戶特徵都大致相符，可利用潛在客戶做推廣及用此依據來設立新的據點。



圖十二 train/test score 比較

#### 6. 模型優點：

LSAGAN 比起其他傳統的方法，更能從潛在變數中尋找其中非線性的關係，利用 autoencoder 找出潛在包裹的特徵，可從混合資料中找出未來會繼續使用「i 郵箱」的包裹。

ACC7_0	ACC35_00	ACC24_2	ACC35_12	diff_day	pre
1	0	0	0	1.329329	0.229
1	1	0	0	1.216898	0.954
1	1	0	0	1.187303	0.912
1	1	0	0	6.348611	1.178
1	1	0	0	1.289456	0.912
1	1	0	0	1.141227	0.172
1	1	1	0	0.420127	0.944
1	1	0	0	0.451042	0.181
1	1	0	0	1.226852	0.889
1	1	0	0	1.239803	0.160
1	1	0	0	2.112500	0.185
1	1	0	0	2.081481	0.228
1	1	0	0	1.072940	0.483
1	0	0	0	1.001354	0.201

圖十三 潛在客戶資料特徵

#### 四、創意應用及服務設計（效益評估、可行性分析）

##### 1. 分辨潛在客戶

郵局未來收到的每一筆包裹都可依照此包裹的特徵，利用模型分辨是否為使用「i 郵箱」的潛在客戶，在分出的潛在客戶當中可分為一般民眾及特約客戶兩種類型，我們對兩種不同類型的客群做推廣。

##### A. 一般民眾

「i 郵箱」比郵局設點容易、據點也較多，對於一般民眾來說，距離通常為交寄包裹的一大考量，利用模型找出適合設點的地區，增加「i 郵箱」據點，讓民眾可更方便的收寄信件與包裹。

許多民眾在網上會自行買賣或二手交易，通常會需要頻繁的寄送商品，包裹體積偏小且少量，也是適合使用「i 郵箱」來寄送包裹，若郵局提供合作方案，提供長期租賃櫃位服務與郵資優惠，可吸引民眾使用「i 郵箱」。

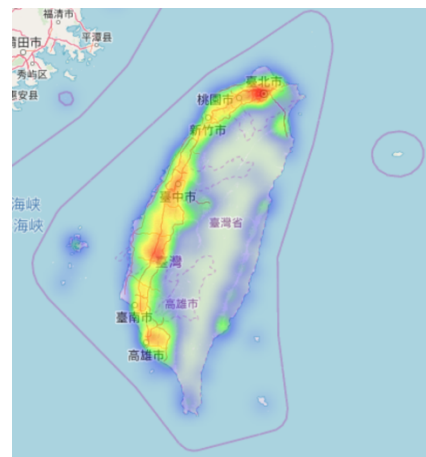
##### B. 特約客戶

資料中顯示，特約客戶中比起預約郵局收件自行交寄的佔大多數，而且需要收寄大量包裹，郵局可與特約客戶合作、放置專屬的「i 郵箱」，為客戶增加方便性，也為「i 郵箱」增加據點與曝光性，營造雙贏。

##### 2. 潛在客戶據點分析

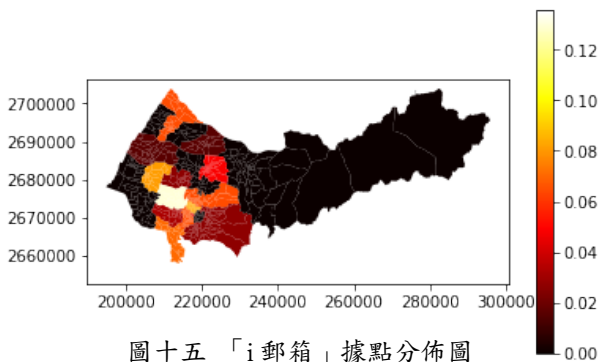
利用模型篩出的潛在包裹來繪製分佈圖，由於資料中並無對應的地址，所以選擇包裹來源的最小單位-交寄局號，結合外部資料-郵局據點畫出全國潛在客戶分佈熱力圖。

圖十四可看到紅色為潛在客戶較多的地區，台北、台中、嘉義和高雄皆是潛在集中區，因為組員都來自台中，進而想探討台中地區的「i 郵箱」的分佈地點及潛在據點。

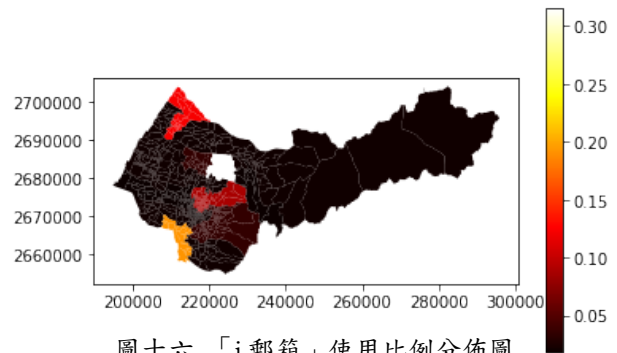


圖十四 全國潛在分佈圖

針對台中地區的潛在包裹資料，比較各個郵遞區號的潛在客戶比例、「i 郵箱」據點比例，圖十五結合政府資料開放平臺上的「i 郵箱」據點，畫出分佈圖來表示使用情況，可看出「i 郵箱」分佈與使用比例不相符。

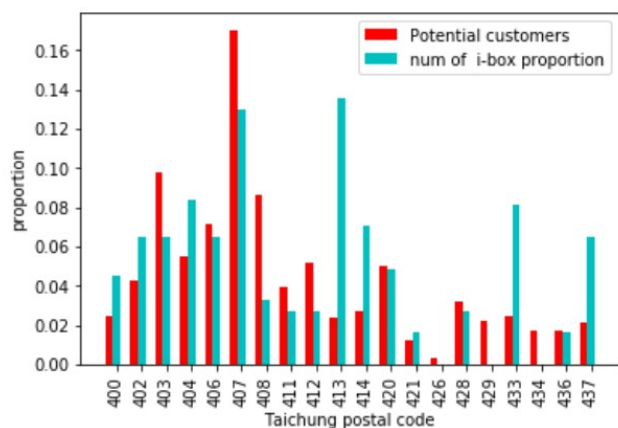


圖十五 「i 郵箱」據點分佈圖

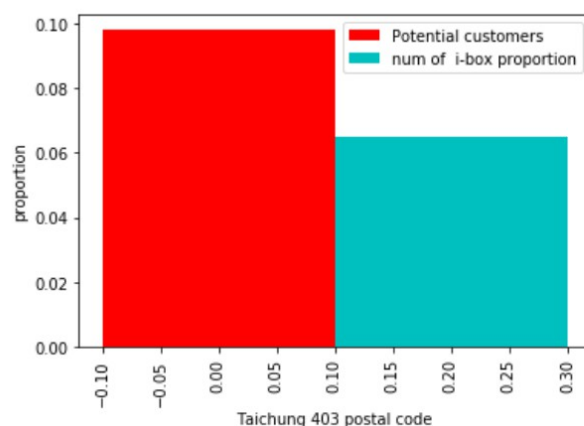


圖十六 「i 郵箱」使用比例分佈圖

由長條圖可更清楚比較兩者的關係，若是要設立新的「i 郵箱」據點，應要選擇潛在客戶比例高，但「i 郵箱」據點相對少的地區，所以從圖十七可觀察到郵遞區號 403、407 皆符合上述的現象，即為台中的西區與西屯區，並選擇西區 403 再更詳細的探討「i 郵箱」的新據點。

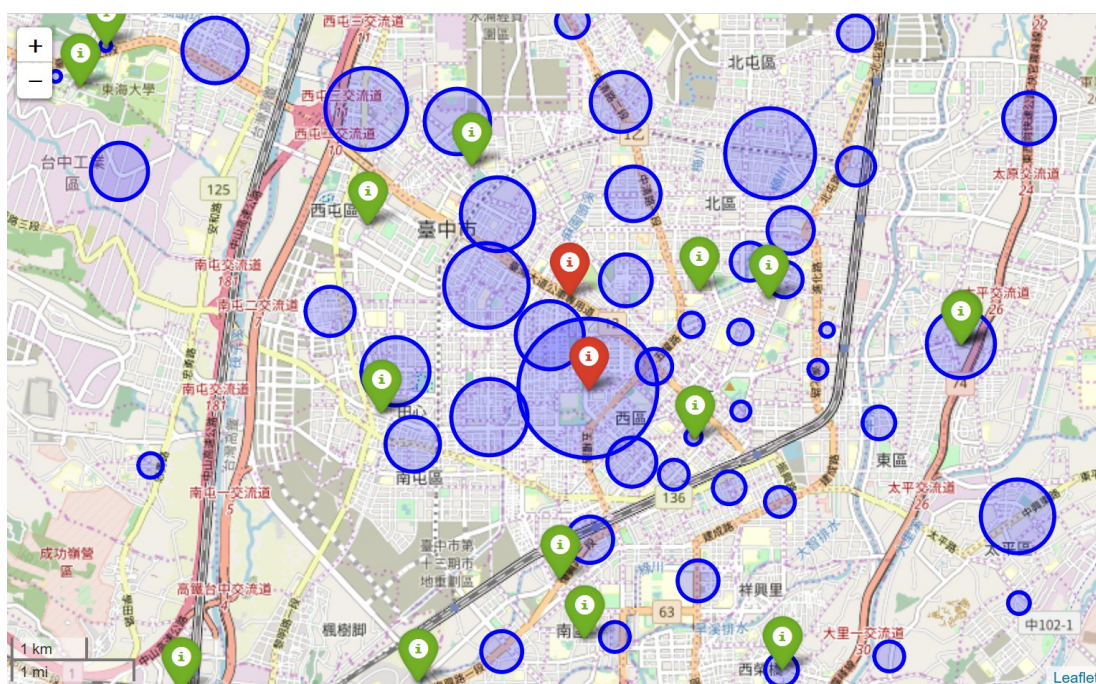


圖十七 台中地區潛在數、櫃位數比較圖



圖十八 西區潛在數、櫃位數比較圖

利用相同的資料畫出潛在客戶分佈的泡泡圖，半徑越大的圓圈代表潛在客戶越多，所以可建議郵局優先選擇在圈圈最大的地區選擇設立新的「i 郵箱」。



圖十七 西區潛在客戶泡泡圖

目前西區只有在英才郵局及金典綠園道商場設立「i 郵箱」(紅色座標)，其餘「i 郵箱」為綠色座標，考慮到使用「i 郵箱」的客群應大多為住在附近的民眾，並非為只來一兩次的觀光客，應要把新的「i 郵箱」據點選擇在住宅區附近，以利住戶使用「i 郵箱」。

利用模型找出潛在客戶都是擁有使用「i 郵箱」包裹的特徵，依照潛在客戶的位置來

設立「i 郵箱」據點，比起隨機選擇或是單考慮人流多寡來選擇設置的點，是為更準確且更有依據，這樣可減少某些「i 郵箱」被設立，卻較無人使用的「i 郵箱」成本，而新建立的「i 郵箱」可為郵局補足人力不足的問題，也提供郵局非營業時段的交寄服務，讓民眾 24 小時皆能使用「i 郵箱」。

### 3. Iber Mail

開發一款專為「i 郵箱」設計的 APP-Iber Mail，提供更方便且完善的服務，簡易操作就能完成，更能打進年輕市場。我們使用 android studio 設計的 Iber Mail 主要擁有六大功能：寄送包裹、追蹤包裹、推薦最佳交寄據點、查看「i 郵箱」空置櫃位、預約櫃位、會員資料，下面依照寄件與收件來描述 Iber Mail 特別的功能。

#### A. 寄件方面：

##### **選擇寄送對象**

可新增寄送人資料或者選擇已填寫過的人做寄送，在 Iber Mail 上會保留以往寄送人資料，可不用再填寫一次。

##### **填寫包裹資料**

填寫包裹相關資料，包括大小、重量、內容物（是否為易碎品）等等，Iber Mail 會依照包裹的特徵顯示應使用何者寄送方式，也利用判別潛在客戶模型自動推薦給使用者可利用「i 郵箱」寄送包裹，不但可幫助使用者做選擇，也可增加「i 郵箱」使用率。

##### **選擇寄達地址類型**

寄送位置包括地址、「i 郵箱」、郵局信箱、軍事特種信箱可做選擇，使用者可依照推薦系統建議，選擇想要的寄送方式。

##### **建議鄰近的寄送地點/預約郵差收件**

依照使用者位置建議鄰近的郵局、「i 郵箱」、郵筒位置，可在 Iber Mail 預先抽郵局號碼牌，減少排隊時間，也可查詢「i 郵箱」的空置櫃位及預約「i 郵箱」櫃位，避免無櫃位可使用，另外對於需要寄送大型包裹的使用者，可選擇預約郵差到府收件。

##### **預約寄達時間**

使用者可在 Iber Mail 上預約送達時間，依普通或限時包裹選擇想要送達的時間，減少包裹未送達比率，也避免到達與收件時間落差。

##### **選擇支付方式**

Iber Mail 提供多種支付郵資的方式，可利用郵政的金融卡或多家銀行的金融卡與信用卡、到場付款也可利用現金、悠遊卡、apple pay 等多種方式，方便使用者做選擇。

#### B. 收件方面：

##### **追蹤包裹**

收寄件者可利用 Iber Mail 追蹤包裹狀態，寄送當中 Iber Mail 也會自動通知收件者目前的寄送狀態，在寄件時若有先預約寄達時間，收件者提早安排取貨時間，避免取

貨時間與安排時間不相符，造成投遞不成功的狀況。

#### **收貨通知**

若是郵差寄送包裹，在包裹快送達時，Iber Mail 會提前通知收件者準備收件，提早讓收件者得知收件時間，減少郵差等待時間。

而若是「i 郵箱」及其他信箱類，Iber Mail 主動通知收件者及告知可取貨時間。若有投遞不成功的狀況，一樣由 Iber Mail 來通知收件者。

### **五、結論與未來展望**

「i 郵箱」一直都是郵局推廣的重點，目前「i 郵箱」正處在於發展期的階段，而以我們目前拿到為去識別化的資料，模型可分辨哪些包裹為潛在客戶，在特約客戶方面，郵局方面可得知客戶 ID 和包裹的對應，即可直接針對潛在特約客戶推廣使用「i 郵箱」。

除了可對特約客戶做推廣，郵局可利用潛在客戶分佈圖新增「i 郵箱」據點，由於資料中並無包裹交寄地址，目前的分佈圖是以郵局局號為單位做繪製，若能在資料中記載更詳細的地址，規劃「i 郵箱」的據點解析度可為更高。而我們相信若擁有「i 郵箱」每個櫃位的使用率和詳細的「i 郵箱」資訊，此資訊可跟我們分析出的潛在客戶相比對，進而設立出更準確的據點。

模型的分析結果對郵局保有更準確的推廣對象與設點目標，再加上 Iber Mail 的推廣，可讓使用者能更方便交寄包裹，Iber Mail 不但增加了「i 郵箱」的曝光率，也讓使用者更認識「i 郵箱」、了解「i 郵箱」的使用方式，進而會把「i 郵箱」列為寄送或收件的選項之一。

### **六、參考資料**

[註一] Adversarially Learned One-Class Classifier for Novelty Detection

[註二] Least Squares Generative Adversarial Networks

使用外部資料：

政府資料開放平臺

A. 「i 郵箱」據點

B. 郵局據點