

2019 中華郵政大數據競賽



告訴你哦

你的信明天十點到



—— 基於深度學習，降低投遞失敗率



隊名：DATA 4MERS

指導教授：吳政隆

隊員：張崱瑋、黃奕鈞、趙奕中、林俐妤

TABLE OF ——— CONTENTS

03	計劃摘要	04	技術應用
05	問卷調查	06	資料分析
08	資料清整	09	模型理論
12	模型成果	14	創意應用
15	系統展示	17	效益評估
18	可行分析	19	計劃總結



計劃摘要



「喂，請問在家嗎？」



「不好意思！我現在不在家。」

是每個收件者與郵差最常見的對話，根據本團隊分析，郵差送件失敗有高達9成2是因為收件者不在家，故本團隊預計引進郵件訊息推播平台來降低投遞失敗率。

收集過去各郵遞區號送達時間等資訊並加以運用。

將清整過的資料輸入DARNN深度學習模型。藉由模型預測時間建置一套內部系統，在適當時機通知收件者收信時間。



技術應用



MSSQL

建置資料



R

資料清洗、
資料分析



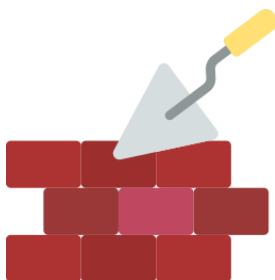
PYTHON

深度學習資料前
處理、環境建置



PYTORCH

深度學習模型使用



FLASK

WEB 開發



POWER BI

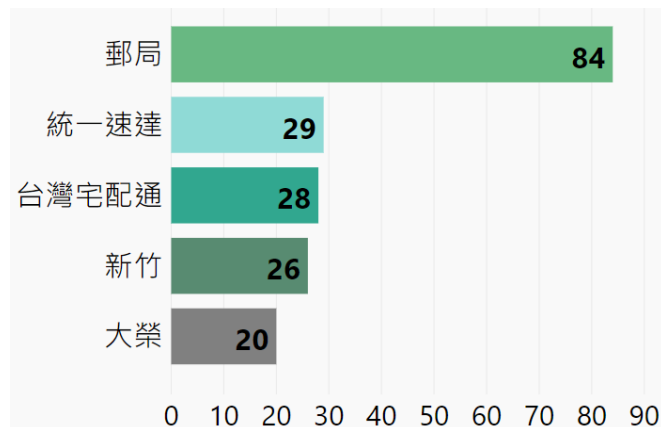
製作視覺化圖表



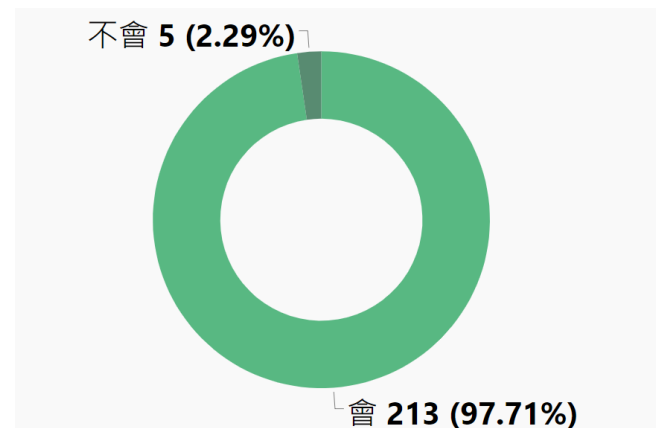
問卷調查

問卷填答人數：218人

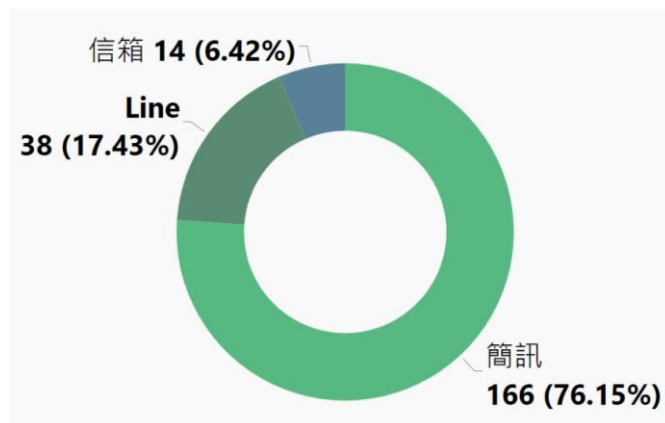
最常使用的物流公司？



會希望有包裹時事先通知嗎？



利用什麼平台事前通知最方便？



事前通知有助於方便收包裹嗎？ (0：毫無幫助——5：極有幫助)

眾數 5 平均數 4.52 中位數 5

可以接受的事前通知時間誤差範圍？ (單位：分鐘)

眾數 30 平均數 32.25 中位數 30



資料分析

1 隨機抽樣100萬筆資料 (ACC28,TT1,TT4)：



1. 結合 Table.TT、Table.ACC

2. TT1 = 投遞成功(I4)或投遞不成功(H4)

3. 取上班時間，早上8點至晚上6點

4. 取郵遞區號第一碼，ACC28 為1到9

ACC28

寄達郵遞區號

TT1

郵件狀態代碼

TT4

處理時間

2 根據問卷，可接受之時間誤差平均為32.25分鐘

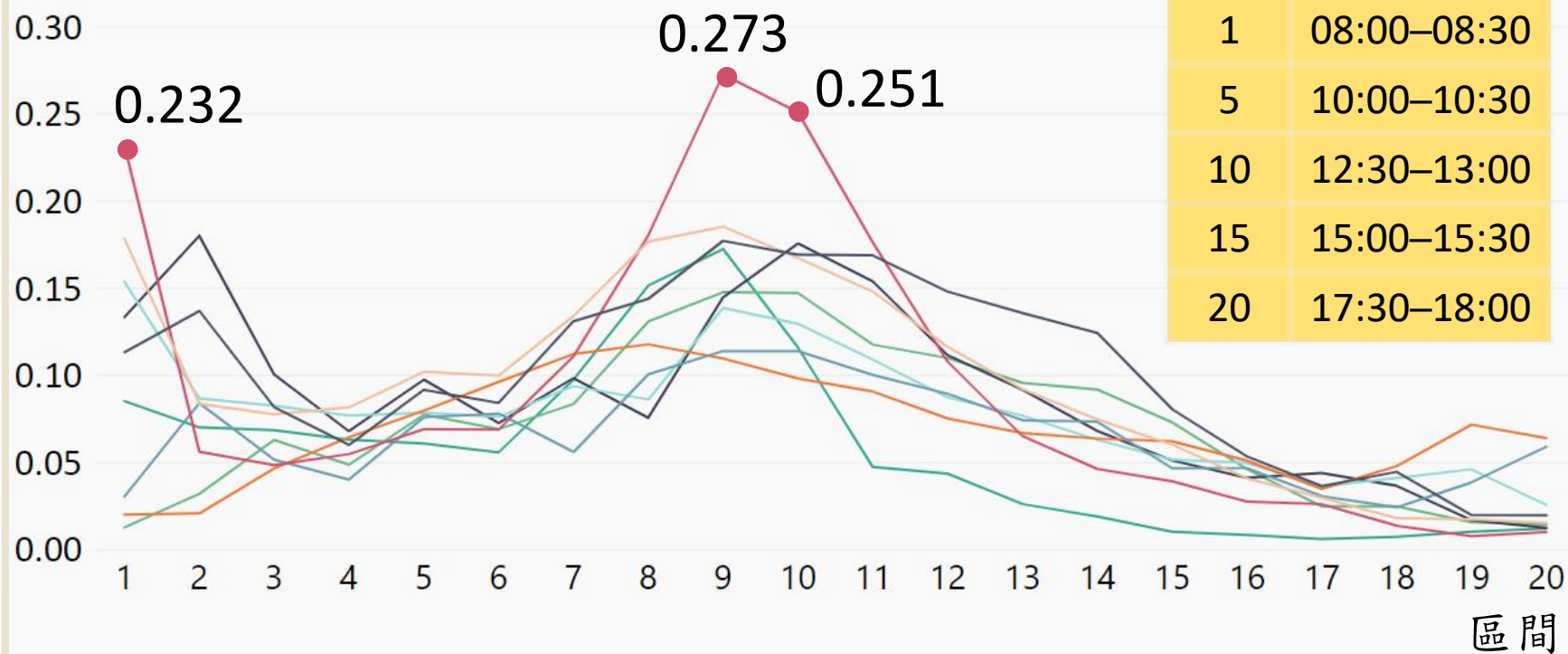
→ 以30分鐘作為一個區間。8:00–8:30為區間一，8:30–9:00為區間二，共20個區間

3 計算 ACC28 二十個區間之投遞失敗率($H4/I4+H4$)



資料分析

投遞失敗率



● 台北 ● 新北、基隆、宜蘭、連江 ● 桃園、新竹、苗栗 ● 台中 ● 彰化
、南投 ● 雲林、嘉義 ● 台南 ● 高雄、澎湖、金門 ● 屏東、花蓮、台東



資料清整

期望透過模型預測出每個郵遞區號未來的送達時間，並提前通知收信人。

1 依據資料分析的結果，台南的投遞失敗率遠高於其他縣市，因此選擇**台南**作為後續模型探討

2 條件1：**包裹及快捷的資料**

條件2：**時間介於8:00–18:00**

依條件隨機抽樣出Q1、Q2、Q3共83445筆當作訓練資料，Q4共20861筆當作測試資料

ACC28,TT3,TT4,TT5,ACC11,ACC12,ACC13,ACC23

3 將時間資料分為區間一至二十，為模型最終預測目標，以區間通知收信人郵件預測抵達時間

ACC28 寄達郵遞區號

TT3 處理日期

TT4 處理時間

TT5 處理局號

ACC11 總件數

ACC12 基本郵件種類

ACC13 細分類

ACC23 速度別



模型理論

A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction (DARNN)

Yao Qin, Dongjin Song, Haifeng Chen, Wei Cheng, Guofei Jiang, Garrison W. Cottrell, A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction, IJCAI, 2017.

- 以深度學習的方式預測送達時間
- DARNN 於時間序列問題中優於傳統 Encoder–Decoder 的模型
- 本團隊此次研究目標即在解時間序列問題，根據目標郵遞區號的歷史資料，預測下一次此郵遞區號應送達之時間
- 根據過往資料預測目標區間外，其還能評估每一個獨立變項對於預測目標的影響



模型理論

- DARNN 為一時間序列模型，屬於 seq2seq 架構(Encoder-Decoder)
- 在輸入 Encoder 之前再加上一層 Input Attention
- 不僅能關注每個獨立變項的重要程度，更可同時兼顧每個不同時間點的重要程度
- 輸入資料必須符合時間順序的原則，且因其參考過往資料，每個郵遞區號需為獨立訓練

Input Data 示意：Input Data 皆為二維

獨立變項X：郵遞區號、處理局號、總件數、包裹種類（針對包裹及快捷）、細分類、速度別（0普通、1限時）

依變項Y：時間區間

▼ 郵遞區號

11111

T (時間點)



獨立變項X

▼ 郵遞區號

11112

T (時間點)

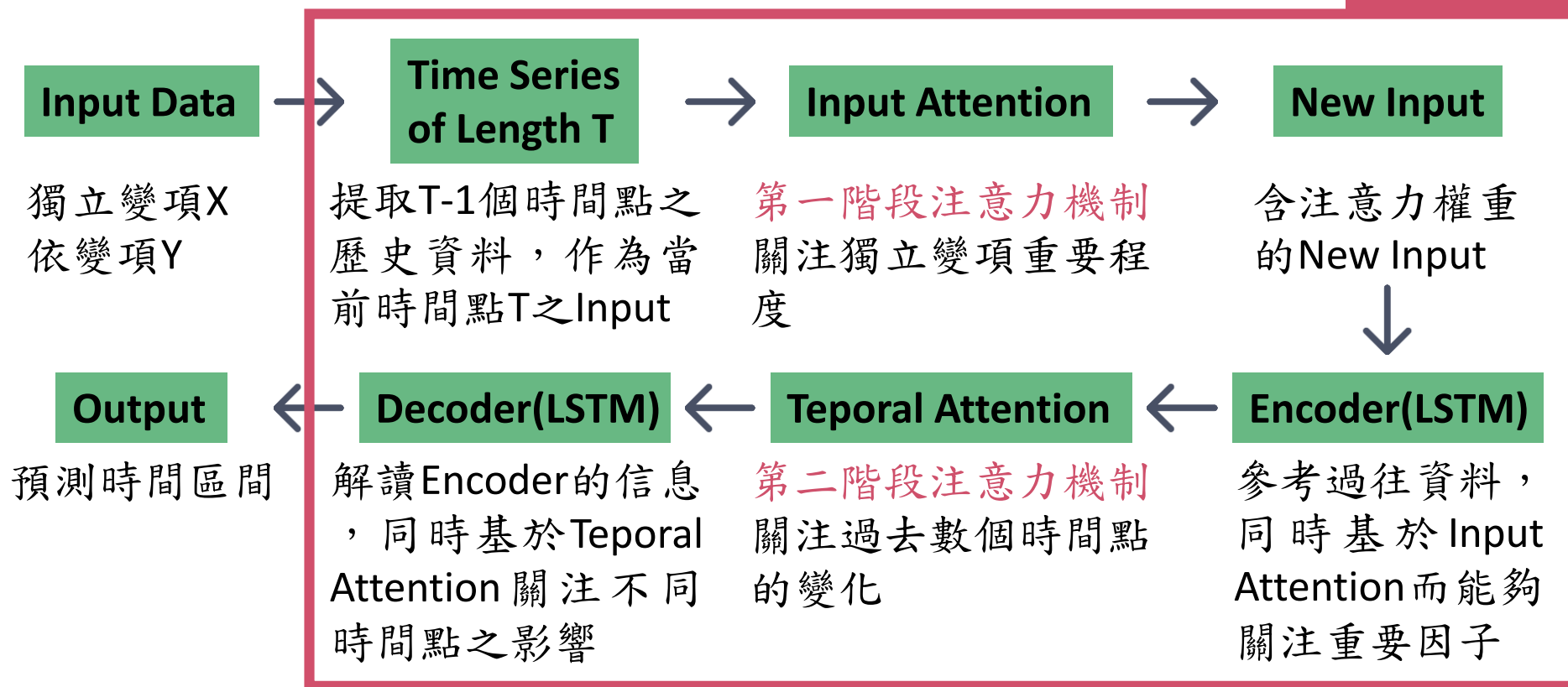


獨立變項X



模型理論

DARNN



初期 Input Attention 關注每一個 Input 變化，如郵遞區號、處理局號，且根據不同時間點產生不同關注程度；第二階段關注過去數個時間點的變化，且每一次 Decoder 預測所關注之重點也會隨著資料不斷的變化。



模型理論

參數設定



DARNN

Encoder Hidden Size = 64

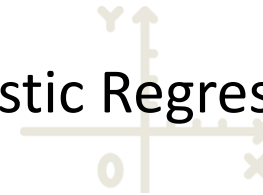
Decoder Hidden Size = 64

Batch Size = 256

Learning Rate = 0.001

Epochs = 6000

Logistic Regression



Hidden Size = 64

Learning Rate = 0.001

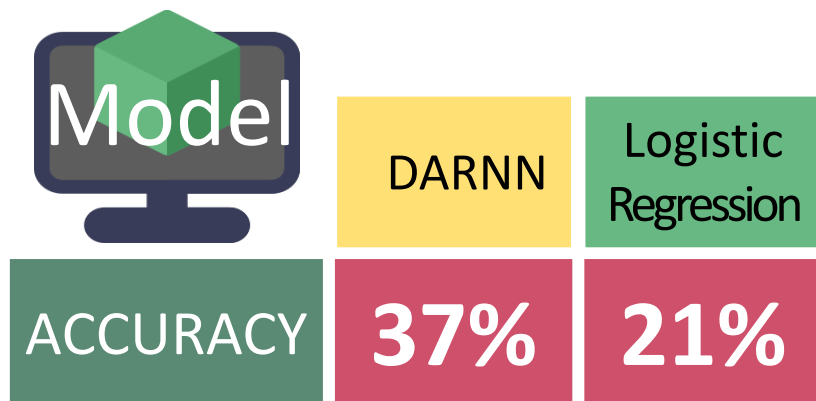
Epochs = 6000

Training Data : 83445 / Test Data : 20861

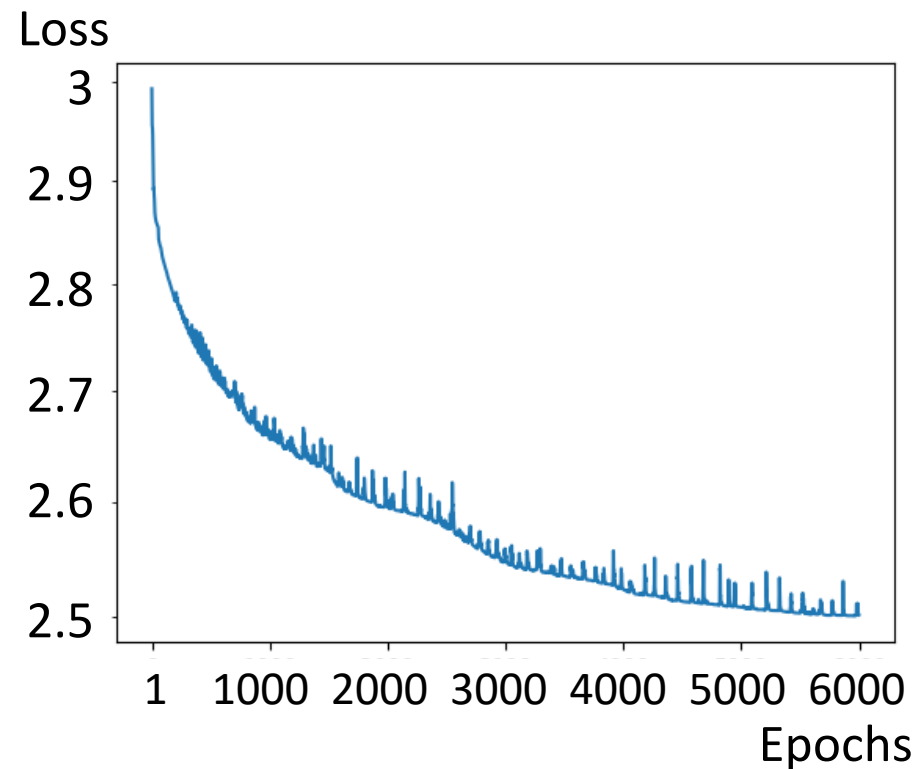


模型成果

為評估模型成效，本團隊將會展示 DARNN 與 Logistic Regression 之差異。



- ▲ 兩種 Model 在相同條件下的表現，DARNN 優於傳統迴歸模型。



- ▲ 本次訓練結果 Loss 收斂。



創意應用

對內：

將明天要投遞的郵件資料輸入至內部系統



系統預測出明天各郵件之送達時間並通知收件者

對外：

收件者在通知之時間範圍內第一次即投遞成功



使用郵局物流

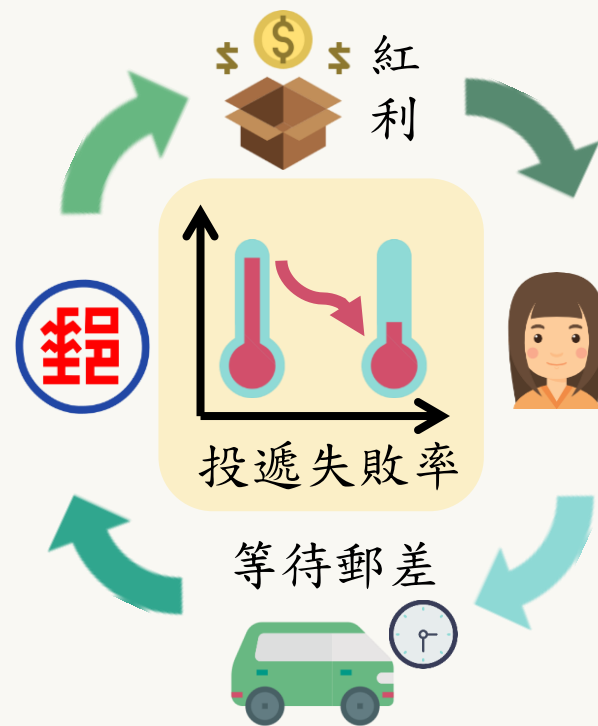


透過點數交換平台刺激商城使用人數



紅利折抵運費

被動滿足收件者收件時間轉換為收件者主動配合郵務收件時間





系統展示

本團隊製作之系統頁面，首頁



中華郵政內部使用系統

輸入日期，索引日期是否有資料

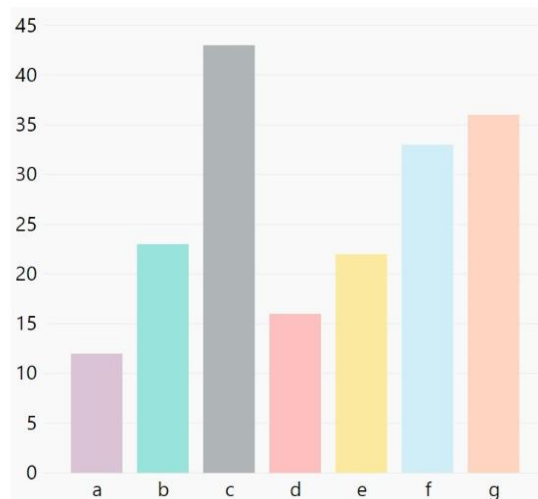
選擇日期

請符合格式：YYYY/MM/DD

模型預測結果：

此處展示簡易分析結果

相關資料分析：



若無資料，將會導向Error頁面



中華郵政內部使用系統



尚未有該日期資料或格式錯誤，請返回。



系統展示



中華郵政內部使用系統

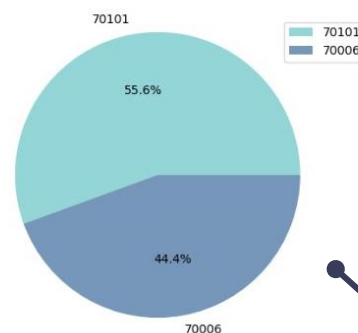
選擇日期

確定

請符合格式: YYYY/MM/DD

模型預測結果:

相關資料分析:



此處以視覺化呈現投遞郵遞區號之比例

展示名字、時間、郵遞區號、預期送達時間

	Name	Date	Number	Predict
0	Arleigh	2019/06/28	70101	[13.0]
1	Tommy	2019/06/28	70101	[13.0]
2	Molly	2019/06/28	70101	[13.0]
3	Toomy	2019/06/28	70101	[13.0]
4	Dino	2019/06/28	70006	[12.0]
5	Dinen	2019/06/28	70006	[12.0]
6	Arlo	2019/06/28	70006	[12.0]
7	Panda	2019/06/28	70006	[12.0]
8	Chaichai	2019/06/28	70006	[12.0]

Date	Number	Predict
2019/06/28	70101	[13.0]
2019/06/28	70101	[13.0]
2019/06/28	70101	[13.0]
2019/06/28	70101	[13.0]
2019/06/28	70006	[12.0]
2019/06/28	70006	[12.0]
2019/06/28	70006	[12.0]
2019/06/28	70006	[12.0]
2019/06/28	70006	[12.0]



效益評估



郵差時薪
181元



一年投遞失敗106173件
一天291件
預期最高降低9成失敗率
平均一天多成功262件



一位郵差一天發送150件
一天省下1.75位郵差工作
量及工錢3168元



油耗0.0595L/km
一台車15km/hr
上班9小時
一天跑135km



油錢30元/L
一台車省下241元
一天省下422元



根據人事成本及油錢，
一天共省下3590元



一年掛號1433677件，一天3925件

	資費/封	總共	省下
LINE	0.15	588	3002
簡訊	0.5	1962	1988



可行分析



投資必要性：

- 會希望有包裹時事先通知嗎？ 97.7% 會 2.3% 不會
- 利用什麼平台事前通知最方便？ 18% Line 76% 簡訊 6% 信箱
- 事前通知4.52分（滿分5分）有助於方便收件
- 可以接受32.25分鐘的事前通知時間誤差範圍



經濟可行性：

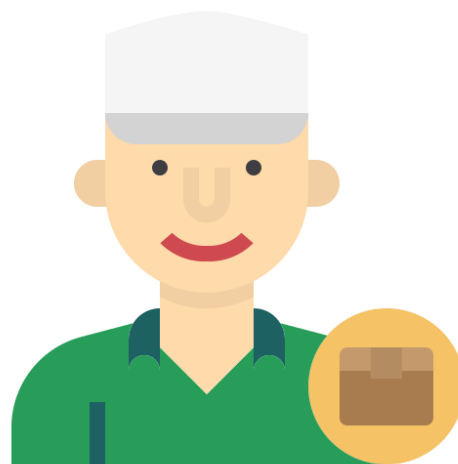
- 輔助人力略過許多繁雜的過程
- 節省不必要的時間和金錢成本
- 對投遞資訊的掌握有實質提升





計劃總結

- 基於深度學習改善投遞失敗率。
- 預期藉此降低投遞成本並將之回饋於客戶，共創雙贏局面。





**THANK YOU
FOR YOUR TIME**