

參賽作品名稱：

基於大型語言模型的交通壅塞語音人工智慧助理與系統
A Large Language Model-Boosted Conversational AI-
Agent and System for Traffic-Jam

隊伍名稱：國道一路通

目錄

| | |
|--|----|
| 本文摘要 | 3 |
| 一、研究目的 | 4 |
| 二、研究範疇 | 5 |
| 三、現況與問題盤點 | 7 |
| 四、資料來源 | 8 |
| 五、系統架構 | 9 |
| 五-1 用於預測事故排除所需時間之決策樹模型 (LIGHTGBM) | 9 |
| 五-2 用於實時車流量統計之物件辨識及追蹤模型 (PRB-FPN 與 SMILETRACK)..... | 10 |
| 五-3 可互動式的大型語言模型代理 (BREEZE-7B-INSTRUCT) | 12 |
| 六、實驗與分析結果 | 15 |
| 六-1 事故排除時間預測模型之預測結果與實驗分析 | 15 |
| 六-2 用於車流量統計之物件辨識與追蹤模型實驗結果..... | 20 |
| 七、預期成果 | 21 |
| 八、相關建議與結論 | 23 |
| 九、參考文獻 | 24 |

本文摘要

在當前快速變化的交通環境中，國道交通流量的優化是確保交通順暢和安全的關鍵因素。由於國道交通涉及眾多複雜的變量與精細的流量控制，任何微小的變化都可能導致交通堵塞和安全隱患。隨著交通需求的日益增長，如何在維持或提升交通流量的同時提高道路安全，成為亟待解決的問題。

近年來，機器學習和數據分析技術的發展為交通流量優化提供了新的解決方案。透過對交通數據的分析，人工智慧模型能夠預測交通流量與各項變數之間的關係，從而提供更精準的管理策略。從而構建端到端的系統以實時監控交通情況，自動識別交通事故異常，並提出用戶即時的調整建議與回饋。

本研究旨在如何利用大型語言模型及先進人工智慧演算法來舒緩台灣國道交通流量問題，以提高道路通行效率。藉由綜合分析和應用先進的數據處理技術。並期望能夠建立一套智慧交通系統，實現交通管理的自動化和智慧化，進而提升整體交通效能與安全性。藉由這種方式，不僅能提高道路使用效率，減少交通堵塞，還能在更大程度上滿足民眾對高效交通的迫切需求。

一、研究目的

現今針對交通壅塞的解決方案仍然大量依賴人為處理或傳統儀器檢測，隨著人口的密集程度日益提高，這些方案的有效性日漸不足。隨著人工智慧與深度學習技術的發展，將傳統方案進行優化並整合的可能性也隨之提高，這項技術的發展可以大幅降低相關單位的運營成本，並改善用路人及國道的用路環境。

事故排除是影響交通狀況的重要因素，現有的方法聚焦在蒐集如天氣、路況、過往發生事故等資料以訓練 LightGBM [1] 等機器學習模型，對事故排除時間進行預測，但其應用的整體性及泛化性仍存在提升及發展空間。

有鑑於上述內容，我們的提案內容聚焦在如何偕同大型語言模型及人工智慧相關技術，提供用路人及相關單位即時查詢國道壅塞情況的即時且強大的系統，以利交通單位與民眾兩方快速應對與決策，從而達到舒緩交通壅塞的問題，並期望於實際部屬後能夠配合 1968 等官方平台並增強其的使用便利性和普及性。通過精確的數據分析和模型訓練，本研究期望實現國道交通管理的自動化和智慧化，提升道路使用效率，減少交通堵塞，滿足民眾對高效交通的需求。

二、研究範疇

隨著科技的快速發展，交通運輸領域也面臨著日益複雜的挑戰，尤其是在國道交通流量管理方面。國道壅塞問題困擾許多駕駛人，其原因在於駕駛人員未能善用交通局設置的 1968 平台，該平台定時更新國道各路段的壅塞情況，但由於駕駛人未充分利用，平台的潛力未能發揮，導致駕駛人僅憑經驗或偏好選擇行車路段，最終造成壅塞。此外，現有的車流量相關統計及檢測儀器精度不足，無法在極端天候或特殊情況下有效檢測並統計車流量，進而造成誤判。

再者，事故排除及其所需時間的估計也是重要問題，現有的模型即便在既定資料上有良好性能，但於實際應用時會因模型泛化能力不足及資料分佈可能不一致而導致其效果受限。故僅部屬用於預測事故排除模型可能並不周全。

為了解決這些問題，我們提出了一項綜合性的研究計劃，旨在利用人工智慧 技術來優化國道交通流量管理。我們的提案將分為兩個主要部分：

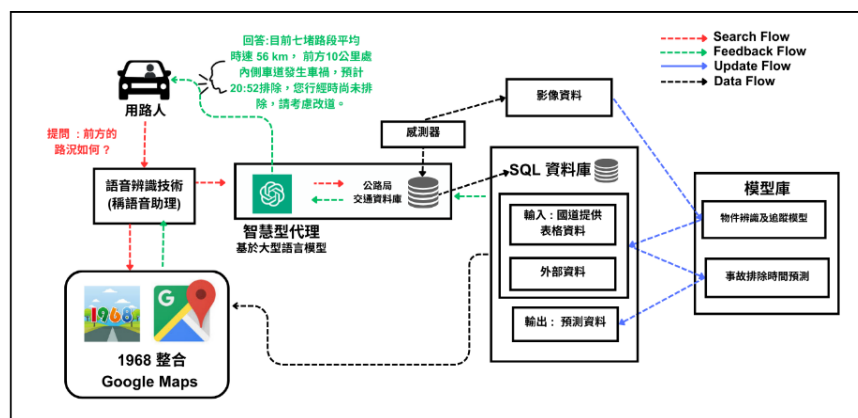


圖 1：系統架構示意圖

- 偕同大型語言模型以提高民眾查詢交通路況及相關內容的方便性：
 - 將資料清理並整合以微調大型語言模型，希望能夠回答駕駛人對於國道壅塞情況的提問，讓駕駛人能夠實時查詢特定路段的壅塞情況。
 - 配合語音交互方式，以提升駕駛人使用 1968 平台的便利性和實用性，藉此提升現有平台或後續模型的利用率及普及性。
- 提升車流量預測在全局或局部路段的精準度：
 - 引用以 PRB-FPN [5]系列等對於不同尺度物件上均能有良好效能並且能夠實時推理的輕量物件辨識模型進行車輛預測，並且引入 SmileTrack [6]等多物件追蹤模型來建立實時的車流量統計及追蹤系統。
 - 對表格數據進行數據預處理，處理缺失值、離群值等，並設置以 LightGBM [1]等具有代表性的人工智慧模型，並引入集成學習及模型堆疊等策略來提升預測事故排除所需時間的模型效能。

本計畫的目標是藉由精確的數據分析和模型訓練，逐步降低預測誤差率，並將這一技術應用於國道交通管理中，提升整體交通效能與安全性。藉由這兩個部分的綜合應用，我們期望能夠建立一套智慧交通系統，實現交通管理的自動化和智慧化，不僅能提高道路使用效率，減少交通堵塞，還能在更大程度上滿足民眾對高效交通的需求。

三、現況與問題盤點

隨著人口增加，我國國道高速公路在尖峰時段、惡劣天氣或交通事故下常發生壅塞，給民眾帶來不便，解決此問題迫在眉睫。本章節透過問卷和社交論壇調查，介紹國道壅塞的可能原因及交通部交通資訊服務的現有問題。

- **國道高速公路：**

- 交流道紅綠燈等待時間及加開路肩時段為固定設置，無法根據路況調整，增加人力成本並潛在造成壅塞。
- 相關單位處理國道突發事件流程不及時，依賴民眾通報，處理時間長，易造成單車道回堵，增加壅塞舒緩難度。

- **1968 平台：**

- 多線道路段即時車速為平均車速，無法準確反映各車道實際路況，系統精確度及有效性存疑，具體情況如圖 2 所示。
- 依賴 VD 感測器，天候不佳時可能出現感測失準或設備異常，數據與實際車流量不符，且大量部署感測器存在運營成本問題。
- 用路人未充分利用交通局設置的 1968 平台。



圖 2：現 1968 平台的即時路況系統

四、資料來源

用於訓練及預測之表格資料(此處特指預測事故排除時間，如第五-1 節)分別來自於交通部於此次競賽所提供之 112 年 1-10 月交通事故簡訊通報資料以及 113 年 1-2 月交通事故簡訊通報資料。

車流量統計預測(第五-2 節)來自於新竹市警察局所提供之 Fisheye-8K[10] 資料集，其包含了 80K 的透過 18 個不同規格的魚眼相機在新竹市不同重要交通路段拍攝的連續影像，其對公車、車輛、自行車、行人、卡車等不同類表進行標註，總共包含了 157k 個物件。其雖不全然涵蓋國道場景，但據我們所知，這是現有極少數公開、用於物件辨識模型、且高品質的台灣交通影像資料集。此外，此資料集也在 CVPR 等公認頂尖的人工智能與電腦視覺相關的國際會議上用於由 NVIDIA 主辦的 AI-city Challenge 國際競賽[11]，因此也有不

少相關對於該資料集的研究報告，因此使用該資料集作為比較的基準會為我們的實驗內容廣泛地提高全面性及有效性。

五、系統架構

於本章節中，分別介紹系統架構的各個模塊，如上文圖 1 所示。主要會分成四個部分進行細部說明，分別如下：(1)決策樹模型對於表格資料的處理，用於預測事故排除所需時間，(2)物件辨識及追蹤模型用於車流量統計，(3)大型語言模型集成上述三者以供用路人進行路況及交通相關內容查詢。因需要考量實際部屬的各種情況，提案內容除了方法的可行性外，亦會考慮訓練或部屬成本，以及推理時間等重要問題。

五-1、用於預測事故排除所需時間之決策樹模型

決策樹算法在處理表格數據，尤其是數據規模受限的情況下，具有顯著優勢。其結構直觀，易於理解和解釋，每個節點代表特徵，分支對應特徵取值，葉節點則為預測結果。此外，決策樹能夠自動選擇最重要的特徵，避免因過度擬合導致新數據表現下降，這對數據規模有限的情況尤為重要。

我們選擇使用 LightGBM (Light Gradient Boosting Machine) [1]作為人工智慧系統預測事故排除的代表模型。LightGBM 在表格資料預測中具有卓越的準確性與泛化能力，並且支援平行訓練，能有效處理大規模資料，訓練成本低且推

理速度快，適合實際應用。為進一步提升模型性能與泛化能力，我們還應用了集成學習和模型堆疊技術，以提高預測準確性，從而改善交通壅塞狀況。

五-2、用於實時車流量統計之物件辨識及追蹤模型

國道現有的影像管理系統的會在固定時段(如每兩分鐘為單位)進行拍攝並回傳到中央伺服器，這其中存在大量優化空間，如圖 4 所示。首先，隨著人工智慧與電腦視覺等技術的發展，越來越多的國家使用物件辨識追蹤技術來改善交通路況，因此部屬相關模型並整合至系統中勢在必行。二者，部屬物件辨識與追蹤模型的效益除了提供車流量統計資訊以改善事故排除模型準確度以外，其相對低廉的部屬成本也使得取代傳統設備成為可能，也易於大量部屬，從而得以構建更完備的車流量監控系統。



圖 4：車輛偵測器(VD)所獲取的資料的變數解釋(左)；CCTV 即時影像內容(右)

我們提議使用 PRB-FPN 作為物件辨識模型，其特點在於 (1)繼承了

yolo(You-Only-Look-Once)[17]系列的單階段辨識模型的特點，推理速度極快且功耗低，易於大量部署（2）其網路架構與特性適合多尺度物件檢測，尤其是對於檢測極小型物件。考量到車體有大有小，對於小物件檢測優化是必須的。

基於檢測的結果，我們提議接續使用 SmileTrack 作為物件追蹤模型，其與 PRB-FPN 為同個作者團隊提出，並且同樣有著推理速度極快且功耗低的特性，其已在多個用於物件追蹤的代表性大規模資料集(如 MOT17、MOT20) [18]上取得 state-of-the-art 的效能，其有效性已充分得到認證。

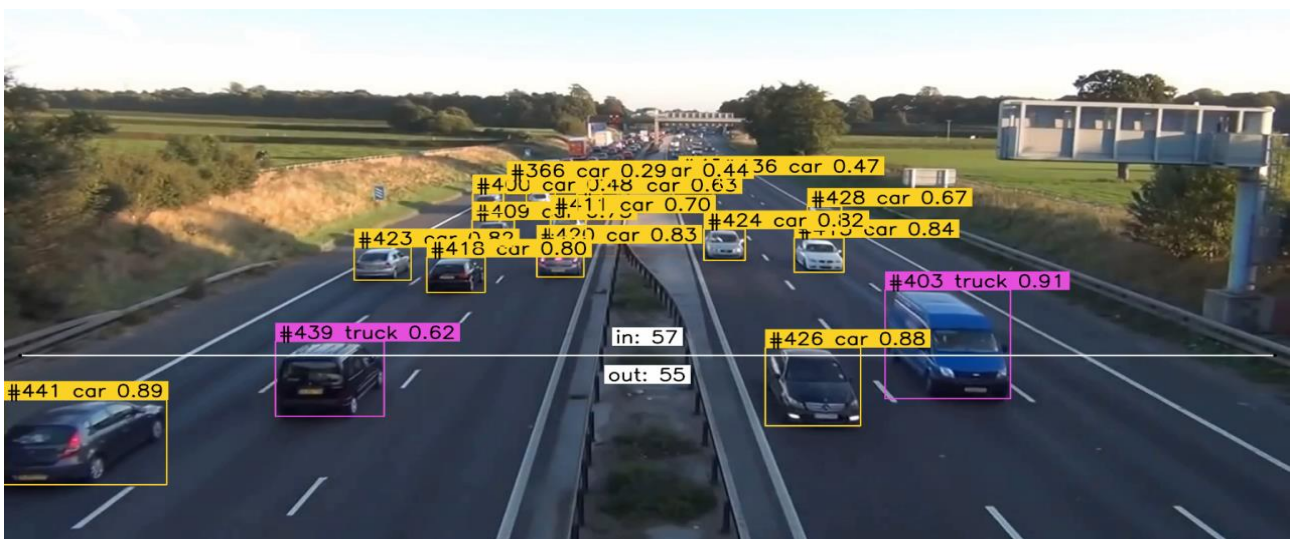


圖 5：車輛辨識追蹤與車流量統計系統示意圖

我們提議在國道路段部署相機，將影像回傳給交控中心後，使用上述提到的物件辨識與追蹤技術，進而判別車輛並統計「即時」車流量狀況（藉由進出某個區域）。此系統亦能用於辨識掉落物及車禍，並即時通報交管中心，相關單位提前行動，減少排除事件的時間。此外，基於影像的方法往往低成本，不

須部屬大量昂貴或不穩健的感測器，僅需部屬相對便宜的相機在國道中即可完成布署。現今網路上亦有大量的公開資料集可供模型預訓練，微調此類模型的成本亦不高。車流量統計亦能有如下幫助：

- 調整上、下交流道的紅綠燈 [3]、[4]：目前多數的紅綠燈控制系統多半是基於 VD 檢測器感知的車流量進行調控，一來 VD 檢測器的精度尚有提升空間，二來無法根據其他資訊來自適應地改變紅綠燈的停等時間。有鑑於此，基於影像的車流量統計有潛力偕同其他統計資料(如天氣、車種、日期等)的來即時調整各紅綠燈的時間，以改善上下國道壅塞的現象。
- 動態調整路肩開放時間 [3]：不同於傳統固定路肩開放時間的方式，在符合現行高公局法規的路肩管制策略，依照車流量統計模型判讀預測整體車流變化來判斷是否開放路肩通行，於安全範圍內最小化與原本無管制車流的壅塞時間差異，達到最安全且最有效率的路肩通行環境。

五-3、可互動式的大型語言模型代理

如前文所示，此項目除了提出基於 LightGBM 的事故排除時間預測模型和具有潛力的車流量統計模型外，旨在幫助用路人語音查詢欲知道的路況，使駕駛人不需手動操作 1968 平台的 app，並解決年長者視字不便以及不熟悉手機操作的問題。有鑑於被廣泛認為功能強大的大型語言模型(Large Language Model, LLM)盛行，已有越來越多的公司基於 LLM 設計或改良現行系統，因此將其整合進未來的交通系統以提升服務品質及可靠度是勢在必行的趨勢。

此處我們簡要地介紹 LLM，其泛指基於 GPT (Generative Pretraining Transformer)[13]的生成式深度學習模型架構。其基於透過自注意力機制[14]理解文本中上下文之間的語意訊息，並透過掩碼預訓練[15]等技術提高學習到的特徵表示品質，最後在推理階段進行文字的預測與生成(Next token prediction)。隨著深度學習的發展與資料規模和設備的指數級增加，GPT 模型逐漸能夠與用戶進行交流與互動，最後發展成了現今的 LLM。最具代表性的 LLM 是家喻戶曉的 ChatGPT、Claude 與 Mistral 等。

- Breeze-7B-Instruct

此處使用的 LLM 為 Breeze-7B-Instruct[7]，是聯發創新基地 (MediaTek Research) 開源的 MediaTek Research Breeze-7B 模型，他是基於 Mistral-7B 開發的語言模型，是一款針對繁體中文語言理解優化的開源語言模型，具備出色的語言理解和生成能力。

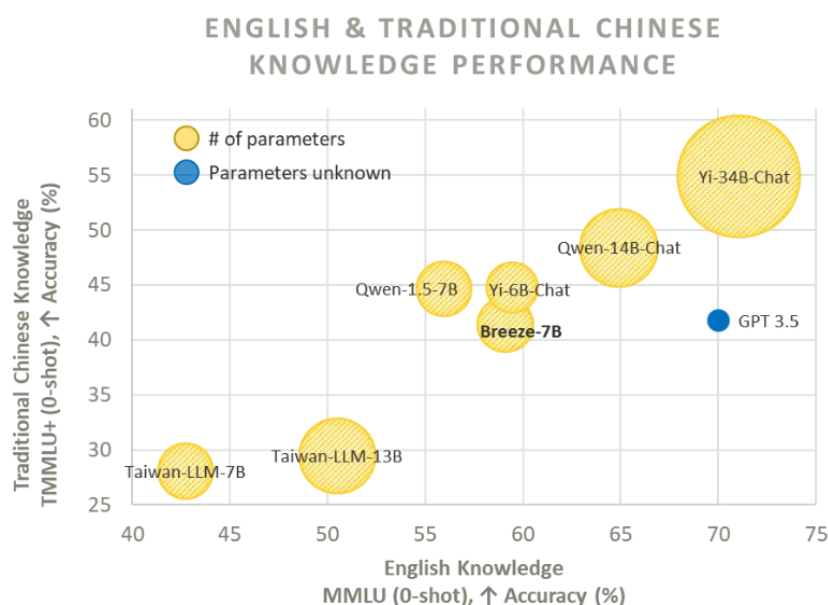


圖 6: Breeze-7B 在繁體中文知識評測數據集(TMMLU+) 和英文知識評測數據集(MMLU)的測試結果比較圖[7]。

- LoRA (Low-Rank adaptation)

LLM 在我們的系統中扮演舉足輕重的角色，為了使其充份理解國道相關資料，以正確地回答用戶問題，對預訓練的 LLM 進行微調是不可或缺的。為了有效減少計算成本並使其適應相關任務，我們使用 LoRA[12]來對 LLM 進行微調。LoRA 是一種高效的模型微調技術，其旨在將凍結住的模型中(此處特指 GPT)的 feedforward network 旁額外設置一個可訓練的參數矩陣，在微調時只對該矩陣回傳梯度並更新參數，實現對特定任務的適應。

具體運作流程如下文：由用路人在 1968 平台的 app 內使用語音輸入，再把輸入的音訊藉由 RNNoise[2]進行語音雜訊降噪，並且將其轉成文字。接著輸入本研究經由 LoRA 等高效微調技術所微調後的大型語言模型 Breeze-7B，使其能對用路人語音輸入的問題做出回答。此外，透過讓 LLM 透過 LangChain 等公開 package 調用 SQL 資料庫以查詢國道現部屬之儀器與本提案所敘的事故排除預測模型及車流量統計模型所得到的資料，即時回覆使用者路況資訊。最後，將文字轉成語音輸出給使用者。

六、實驗與分析結果

六-1、事故排除時間預測模型之預測結果與實驗分析

資料前處理與特徵工程的詳細過程，請見附錄 A、B。

以下為模型對於測試資料的處理時間預測，結果如圖 7，以 LightGBM 和 AutoGluon[16]的集成模型擁有較高的準確率，能夠以 10 分鐘的誤差預測測試資料的處理時間。而特徵的重要性排名如圖 8。

| | model | score_test | score_val | eval_metric |
|---|-----------------------|------------|------------|---------------------|
| 0 | LightGBMXT_BAG_L1 | -10.527705 | -5.894180 | mean_absolute_error |
| 1 | WeightedEnsemble_L2 | -10.527705 | -5.894180 | mean_absolute_error |
| 2 | WeightedEnsemble_L3 | -10.527705 | -5.894180 | mean_absolute_error |
| 3 | KNeighborsUnif_BAG_L1 | -12.723994 | -11.551474 | mean_absolute_error |
| 4 | KNeighborsDist_BAG_L1 | -12.723994 | -11.551474 | mean_absolute_error |
| 5 | LightGBMXT_BAG_L2 | -15.614480 | -9.151550 | mean_absolute_error |
| 6 | LightGBM_BAG_L1 | -16.157606 | -8.455140 | mean_absolute_error |
| 7 | LightGBM_BAG_L2 | -18.173811 | -8.661640 | mean_absolute_error |

圖 7：模型預測分數

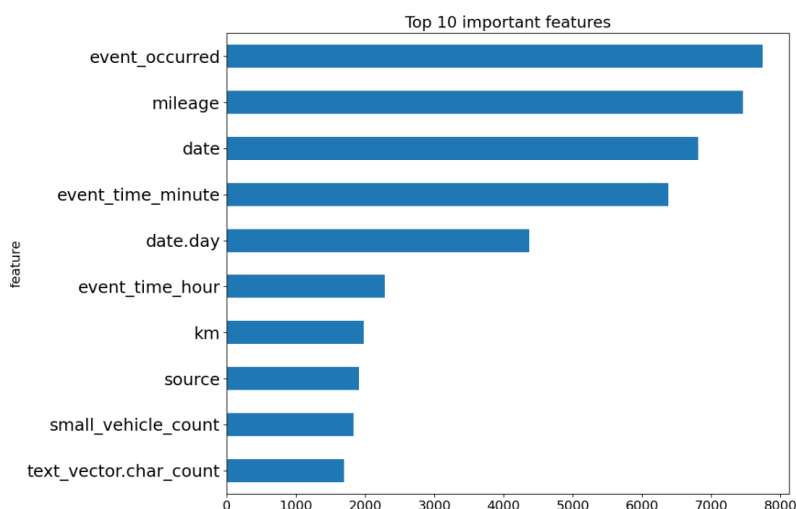


圖 8：特徵重要性排名

最重要的特徵為事件發生(event_occurred)，顯示事故發生的時段會大幅影響處理時間。圖 9 為不同時段的處理時間統計數據和平均處理時間長條圖。

可以觀察到凌晨 6 點前的平均事故處理時間雖然較少，卻擁有較長的處理時間，可能是因為凌晨人力較少的緣故。

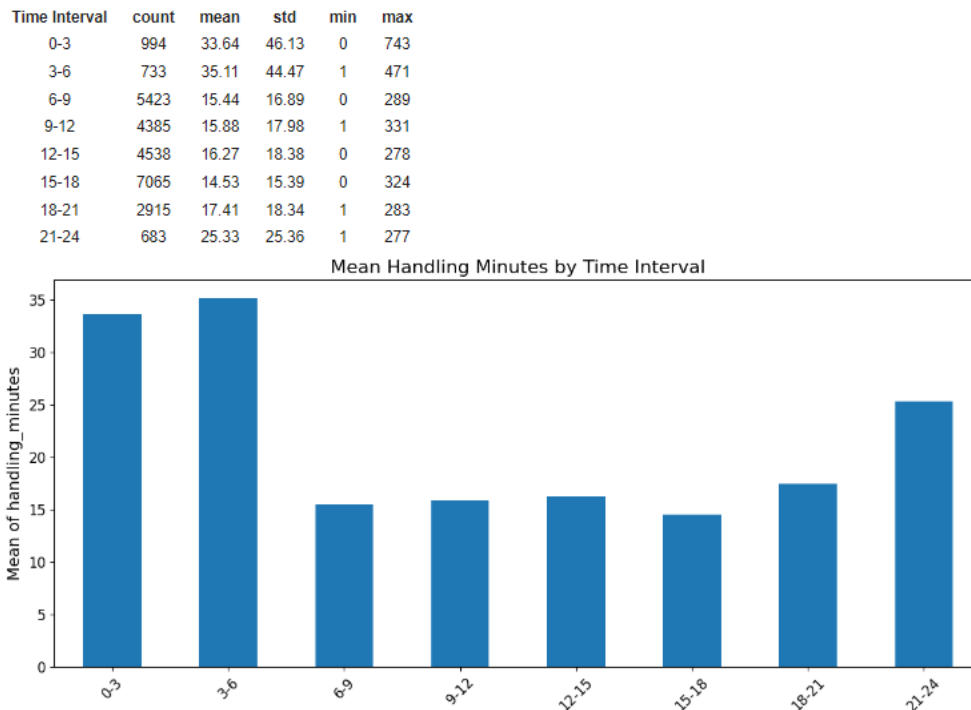


圖 9：不同時間段處理時間統計數據和平均處理時間長條圖

其次是里程(mileage)，圖 10 為不同里程區間的處理時間統計數據和平均處理時間長條圖。可以觀察到由北到南的事故數量呈現遞減，但平均處理時間則呈現遞增。原因可能是南北警力佈署的人力差異，因此可考慮結合全台公安局在不同路段以及不同時段的警力部屬資料來增加預測準確度。

| Mileage Group | count | mean | std | min | max |
|---------------|-------|-------|-------|-----|-----|
| 0-100 | 19066 | 16.12 | 19.47 | 0 | 743 |
| 100-200 | 2290 | 19.88 | 22.68 | 1 | 399 |
| 200-300 | 1790 | 21.33 | 29.94 | 0 | 471 |
| 300-400 | 3451 | 18.17 | 21.82 | 1 | 324 |
| 400-500 | 138 | 24.03 | 18.00 | 2 | 81 |

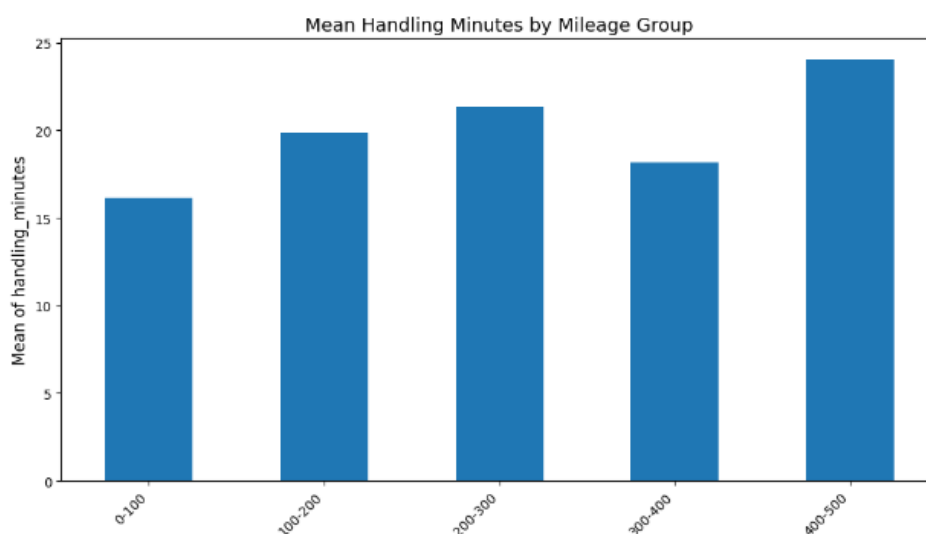


圖 10：不同里程區間處理時間統計數據和平均處理時間長條圖

第三則是日期(date)，為了觀察此現象，我們還另外繪製了圖 11 和圖 12 評估平均事故處理時長是否具有週期性。在圖 11 我們針對每個月的事故處理時間進行了分析，展示了各個星期的平均處理時間隨著星期的變化趨勢。由圖 11 可觀察出以下兩點：1. 週期性趨勢：在折線圖中，每個月的事故處理時間顯示出一定的週期性趨勢。例如，一些月份在週一到週五的處理時間較為穩定，而週末則可能出現較大變化。2. 波動性：折線圖展示了每個月內不同星期的波動情況，有些月份的不同星期之間處理時間的波動較大，而有些月份則表現出較為一致的趨勢。再來圖 12 展示了每個月中完整星期之間的皮爾森相關係數(PCC)熱圖。每個子圖對應一個月份，顯示了該月份中各個星期的相關性。由圖 12 可針對相關性做出以下兩點猜測與分析：高相關性月份：在一些月份中（例如 3 月和 4 月），相關係數熱圖顯示四個星期之間有很高的正相

關，這意味著這些月份的各個星期之間的事務處理時間變化趨勢非常相似。反映了穩定的交通模式，這可能與這些月份中沒有短天數的特殊事件或季節性影響有關。低相關性或負相關月份：在其他月份（例如 5 月和 6 月），部分星期之間顯示出低相關性或負相關，這意味著這些星期之間的事務處理時間變化趨勢差異較大。這種情況通常出現在具有不穩定因素或異常事件的月份，如短天數的節假日、突發的交通事故增加或天氣變化。

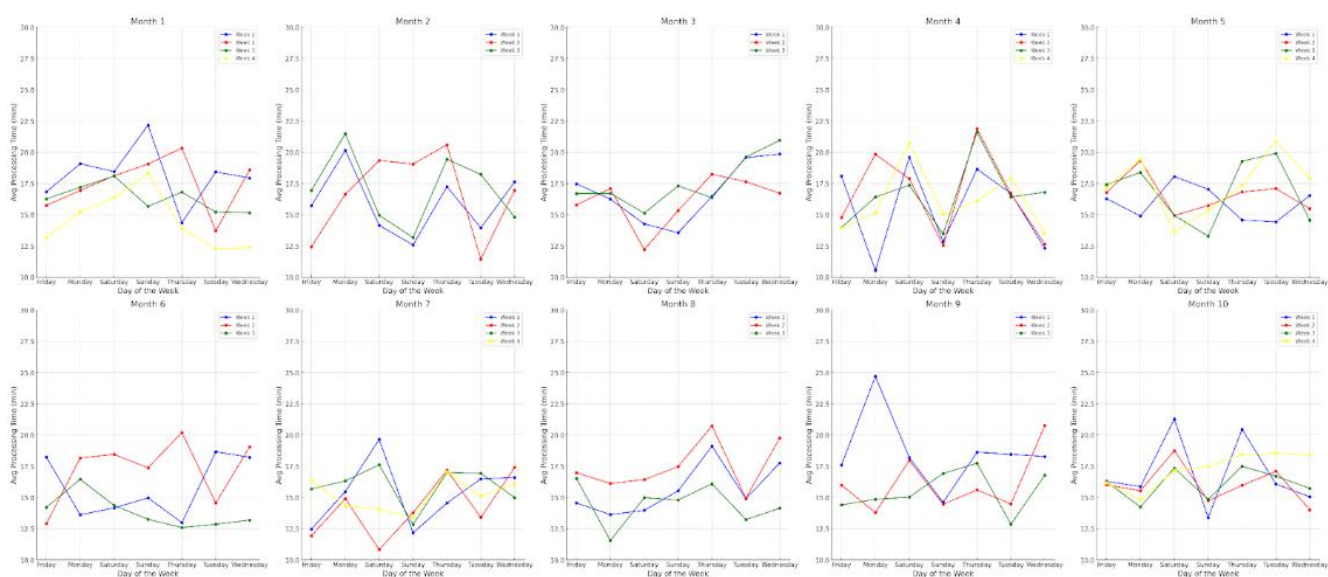


圖 11：不同月份按週的事務處理時長折線圖

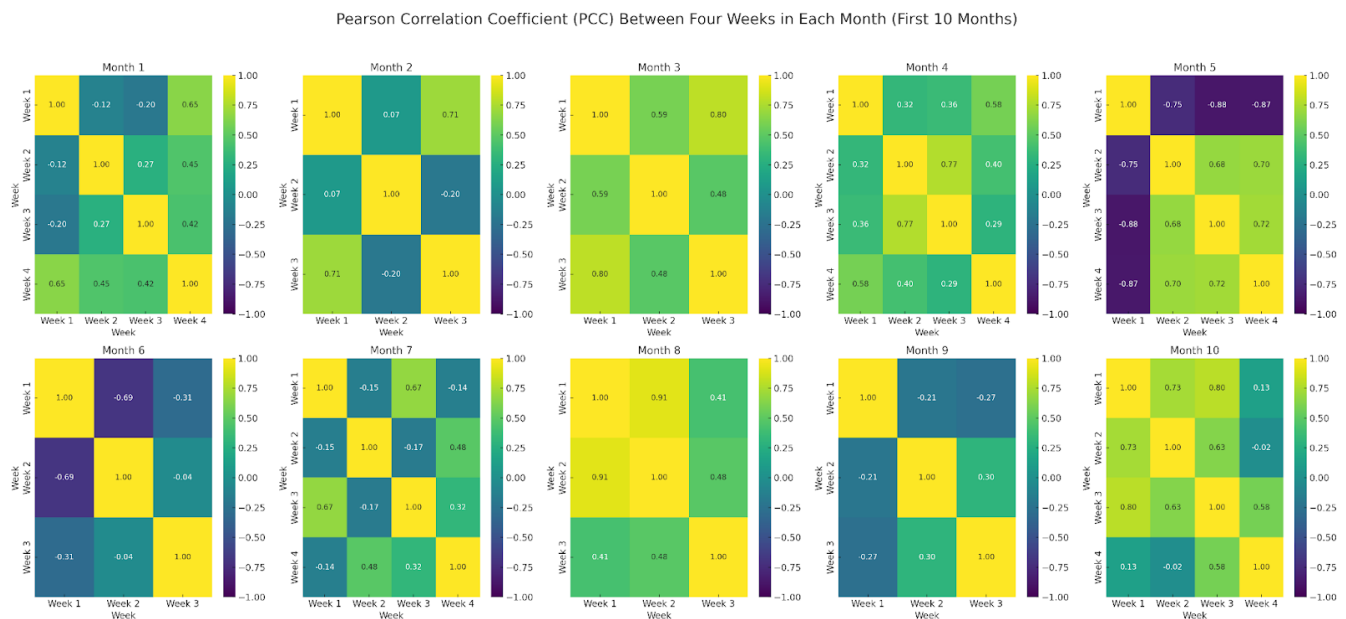


圖 12: 不同月份按週的事故處理時長皮爾森相關係數(PCC)圖

通過觀察這些趨勢，交通管理部門可以根據每個月的具體情況，調整事故應對策略，例如在特定的星期天或星期五加強監控和應對措施，以應對可能的高峰時段。

最後，圖 13 為 LightGBM 的決策樹圖，描述了模型是如何透過特徵進行分裂並預測。可以觀察到，第一個分裂節點便是事故發生時間(event_time_hour)在凌晨五點前後，符合先前對於事故發生時間段的統計描述。里程(mileage)大小則多次出現在決策樹圖中，對應了不同地區大幅的影響了處理時間的長短。而日期(date.day)則出現在部分條件下，分裂的末端處，也同樣對應了特定日期，會對處理時間造成影響。

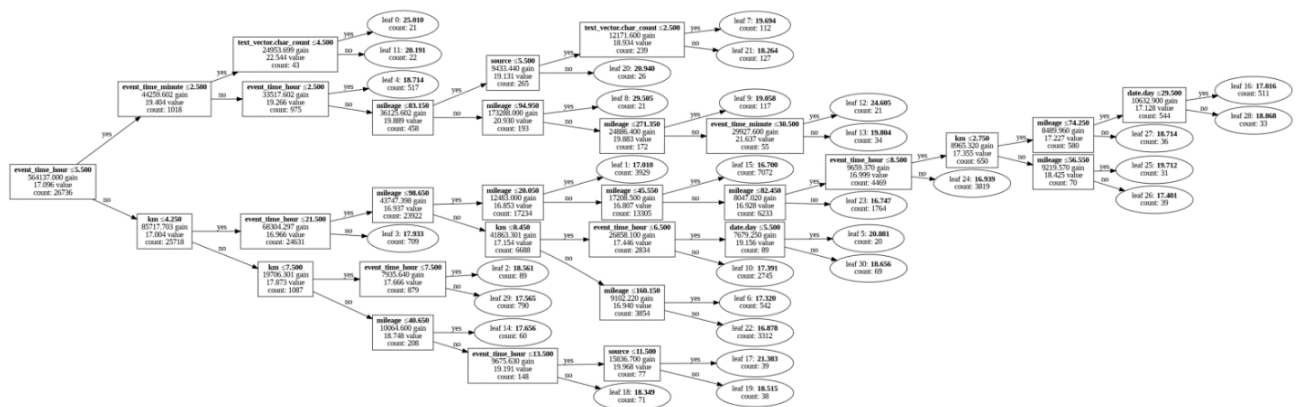


圖 13: LightGBM 之決策條件圖

六-2、用於車流量統計之物件辨識與追蹤模型實驗結果

由於缺乏經標註的國道實時影像資料集，將使用由新竹市警察局提供的

Fisheye-8K 資料集，用於模擬台灣的交通場景。以期望我們的提案能夠有潛力

遷移在台灣國道資料集上以舒緩交通壅塞問題並得到更多實務上的應用。

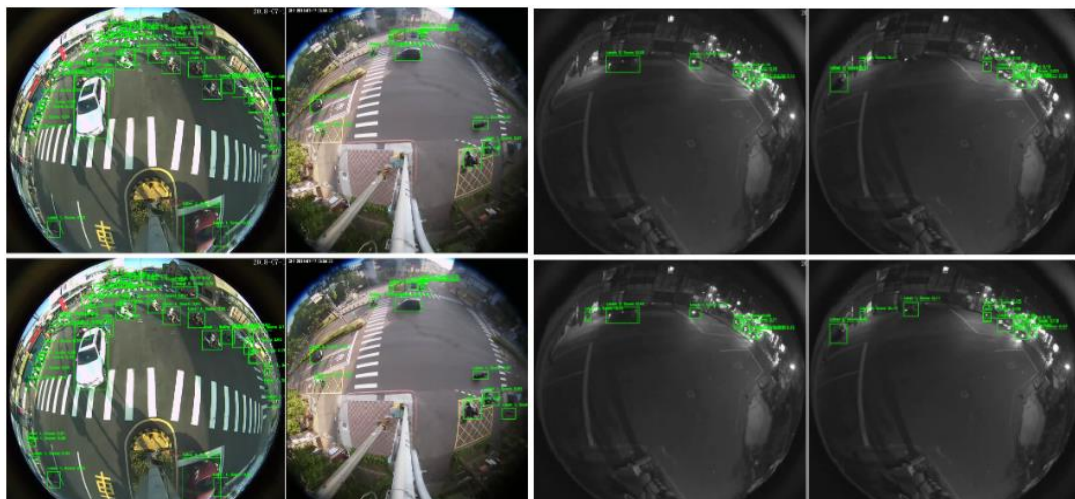


圖 14: PRB-FPN 在 Fisheye-8K 資料集上的預測結果示意圖，為了說明方法的

有效性及泛化性，左圖為白天的影像資料，右圖為晚上的影像資料

| Model | AP(0.50:0.95) | AP(0.50) | AP(Small Obj) | AP(Large Obj) |
|-----------------------|---------------|----------|---------------|---------------|
| Yolo-v7 (baseline) | 0.308 | 0.4739 | 0.1994 | 0.3307 |
| PRB-FPN (proposed) | 0.459 | 0.6437 | 0.2708 | 0.5991 |

表 1: 物件辨識模型的預測結果，各項指標皆是越高越好

受限於 Fisheye-8K 缺乏物件 ID 標籤，故而無法呈現全面的物件追蹤相關實驗結果，因此我們僅使用 IDS 作為本文評估實驗結果好壞的指標。需要注意的是，物件追蹤是基於物件辨識的結果進行關聯，辨識的結果的好壞會大幅影響追蹤的效能。

| Model | ID-switch |
|--------------------------|-----------|
| ByteTrack (baseline) | 108 |
| SmileTrack (proposed) | 54 |

表 2: 物件追蹤模型的預測結果，指標是越低越好

七、預期成果

受限於 1968 平台所使用的交通資料庫為非公開資料，且在 1968 平台中用作時間預測的計算公式及模型無法得知，因此我們此處模擬的資料庫為自行抓取 1968 平台上呈現出的各項資料，如下文來做模擬測試。未來若要實際應用，可直接串接到 1968 平台的資料庫做使用，以利及時更新。我們藉由以下不同情境，說明用路人實際上路狀況。

情境一

用路人提問：我所行駛的路程會在國道哪段塞車，哪個車道比較流暢？

語音助理回答：在國道一號南向 6 公里處平均時速 66 公里，內側車道平均時速 100 公里，車流量較少(見圖 15)。

語音助理提示：國道三號 2.1 公里處，內側車道發生車禍，全車道平均時速 50 公里，預計事故排除時間上午 9:20，抵達時尚未排除，全車道速度驟降，請小心行駛(見圖 16)。

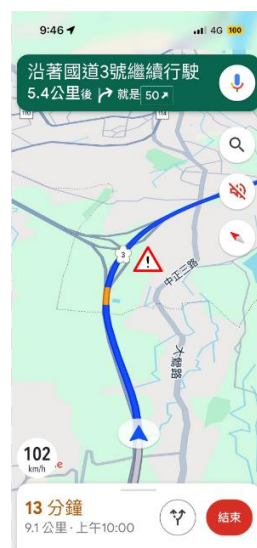
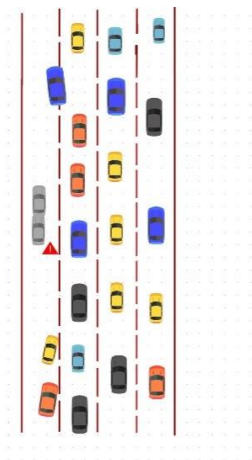
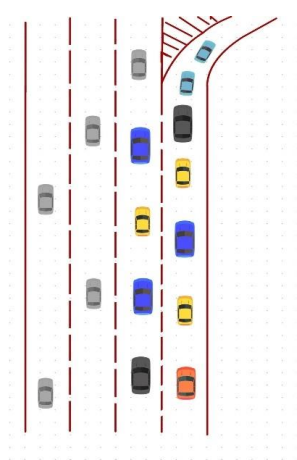


圖 15: 車道示意圖 圖 16: 內側車道發生車禍 圖 17: Google Maps 示意圖

情境二

用路人在國道上持續直行，但前方有突發事故，如圖 17，Google Maps 上出現紅色驚嘆號，表示有突發狀況請注意，此時語音助理將進行路況提示。

語音助理提示:前方 2 公里處，內側車道有車禍，全車道平均時速 50 公里，請用路人提早變換車道、減速慢行，注意與前方的安全距離。

八、相關建議與結論

本研究利用 LightGBM 模型預測事故排除時間，並提出整合 PRB-FPN 與 SmileTrack 等物件辨識與追蹤技術以部屬在國道上，進而高效地獲取國道上的局部車流量統計訊息，受惠於其可實時部屬並且有效的特點，對於路況控管的成本將有機會大幅降低，進而潛在地減少交通壅塞的可能。本文也提出結合大型語言模型相關的技術構建 AI-agent 語音助理，讓用路人可以即時取得國道路況資訊，也期許 1968 平台能結合 Google Maps，從而整合上路相關平台的有效利用率，並促進用路人擁有更好的行車安全與環境。在日後優化模型部分，可以嘗試利用生成國道事故模擬資料，利用基礎模型(foundation model)去大規模生成模擬資料 然後在模擬環境訓練一個基礎任務特定模型(Task-specific foundation model)，減少收集資料之成本，並透過遷移式學習(transfer learning)使模型更擬合現實狀況，並連結更多外部資料，達到多模態模型的應用，使模型能更配適真實世界樣貌。期望我們所提出的一套國道智慧系統，不論是交通部高速公路局、用路人等等，都能因此受惠於此，讓國道行政業務處理更有效率，並希望在未來能夠成為台灣交通單位建造智慧都市系統的重要里程碑。

九、參考文獻

- [1] G. Ke et al., “LightGBM: A Highly Efficient Gradient Boosting Decision Tree” , Advances in Neural Information Processing Systems (NeurIPS), 2018.
- [2] J.-M. Valin, “A Hybrid DSP/Deep Learning Approach to Real-Time Full-Band Speech Enhancement” , IEEE Multimedia Signal Processing Workshops (MMSPW), 2018.
- [3] J. Luo et al., “Design of Variable Traffic Light Control Systems for Preventing Two-Way Grid Network Traffic Jams Using Timed Petri Nets, “ IEEE Transactions on Intelligent Transportation Systems (TITS), Vol. 21, Issue 7, July, 2020.
- [4] T. Cheng et al., “Reinforcement Learning Approach for Adaptive Road Shoulder Traffic Control,” National Digital Library of Theses and Dissertations in Taiwan, 2020.
- [5] P. Chen et al., “Parallel Residual Bi-Fusion Feature Pyramid Network for Accurate Single-Shot Object Detection” , IEEE Transactions on Image Processing (TIP), 2023.

- [6] Y. Wang et al., "SMILEtrack: SiMilarity LEarning for Multiple Object Tracking", Annual AAAI Conference on Artificial Intelligence (AAAI), 2024
- [7] C. Hsu et al, Breeze-7B Technical Report, arXiv:2403.02712, 2022
- [8] 雨量觀測資料, 中央氣象署,
<https://opendata.cwa.gov.tw/dataset/observation/0-A0002-001>
- [9] 日交通量參考值(匝道):
<https://www.freeway.gov.tw/Publish.aspx?cnid=1652&p=37044>
- [10] G, Munkhjargal et al., "FishEye8K: A Benchmark and Dataset for Fisheye Camera Object Detection", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2024.
- [11] S. Wang et al., The 8th AI City Challenge, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshop, 2024.
- [12] E. Hu et al., "LoRA: Low-Rank Adaptation of Large Language Models", International Conference on Learning Representations (ICLR), 2022.

- [13] T. Brown et al., "Language Models are Few-Shot Learners" , Advances in Neural Information Processing Systems (NeurIPS), 2021.
- [14] A. Vaswani et al., "Attention Is All You Need", Advances in Neural Information Processing Systems (NeurIPS), 2017.
- [15] J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", North American Chapter of the Association for Computational Linguistics (NAACL), 2019.
- [16] J. Mueller et al., "AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data", 7th ICML Workshop on Automated Machine Learning (ICMLW), 2020.
- [17] J. Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [18] P. Dendorfer et al., A Benchmark for Single-Camera Multiple Target Tracking, International Journal of Computer Vision (IJCV), 2020.

附錄 A、事故排除時間預測模型之資料分析

由觀察 112 年 1-10 月 A1、2、3 事故資料在 1~10 月的資料中國道事故的資料筆數和簡訊資料集中的資料筆數，我們發現兩種資料集中所包含的交通事故並不完整，例如：4 月事故數明顯有缺漏，因此為了預防偏離現實狀況，我們建議不要利用此資料集另行結合現實世界的狀況做分析。詳細原因請看以下分析結果。

- 針對不同條國道

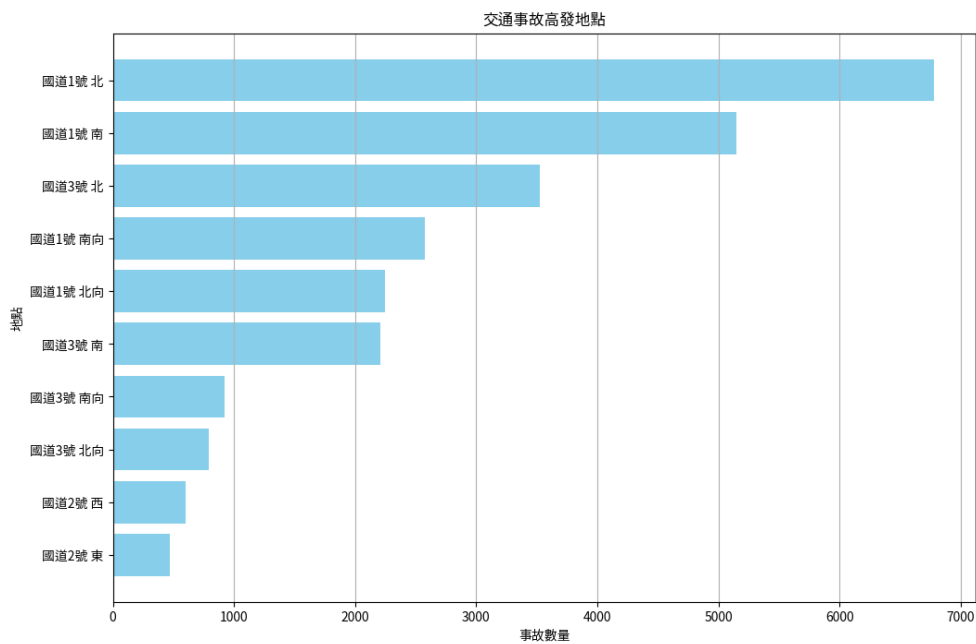


圖 18：交通事故高發地點圖

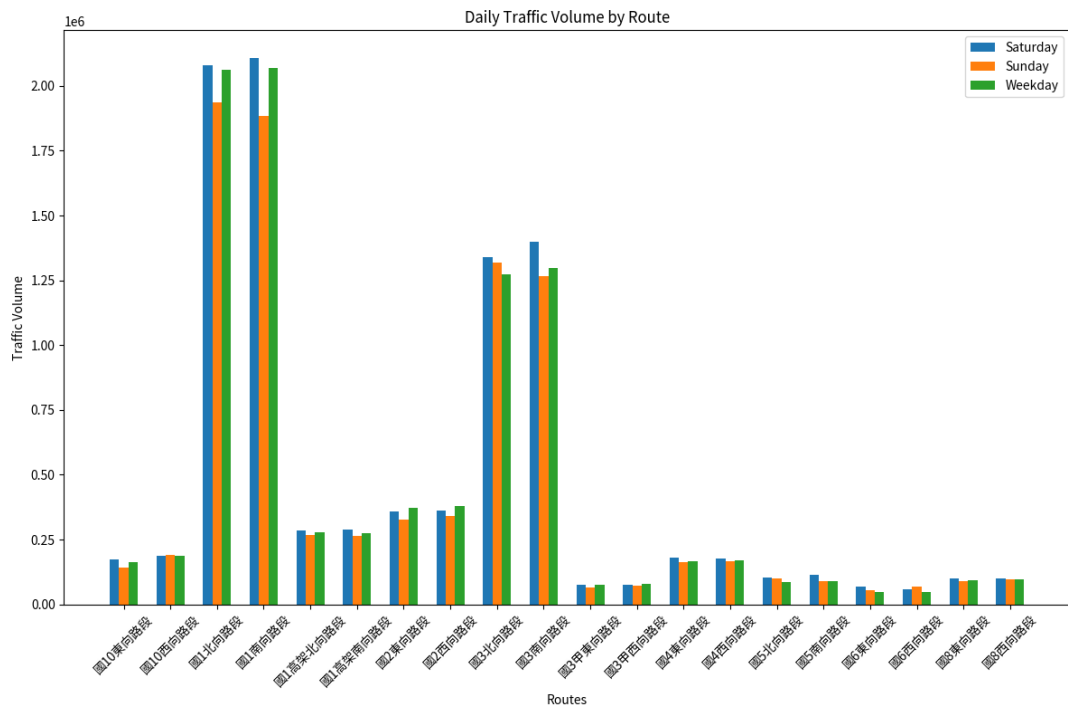


圖 19： 車流量與事故發生之長條圖

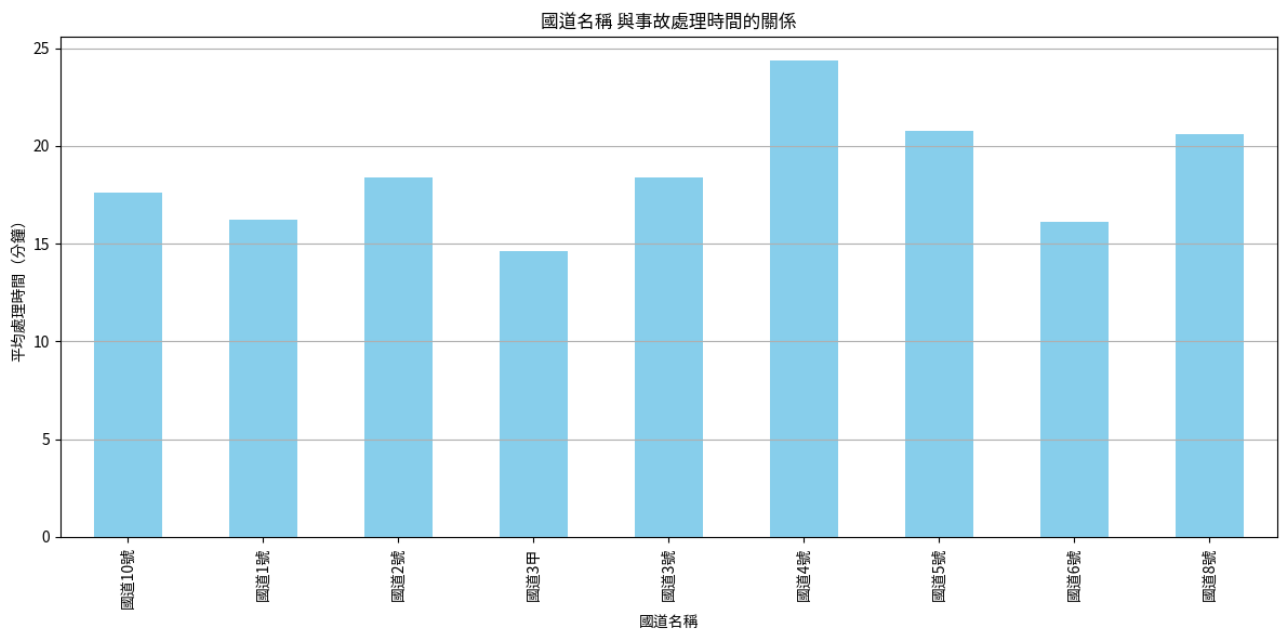


圖 20： 交通事故地點與平均處理時間圖

由圖 18 可看出事故主要發生在國道 1 號和國道 3 號，並且在多發生在北
上。為了比較車流量與事故的相關性，我們這邊利用「交通部 112 年日交通量
參考值」的資料，繪製了圖 19。從圖 19 的結果中，我們可以看到事故發生數

和車流量確實呈現正相關。另外為了確認事故處理時間、車流量以及不同國道是否有關係，我們另外繪製了圖 20，並且將不同國道的車流量與平均事故處理時間以及不同國道的事故數與平均事故處理時間做卡方獨立性檢定，且發現兩者間獨立，因此後續可以考慮將變數「國道名稱」刪除。

- 針對時段

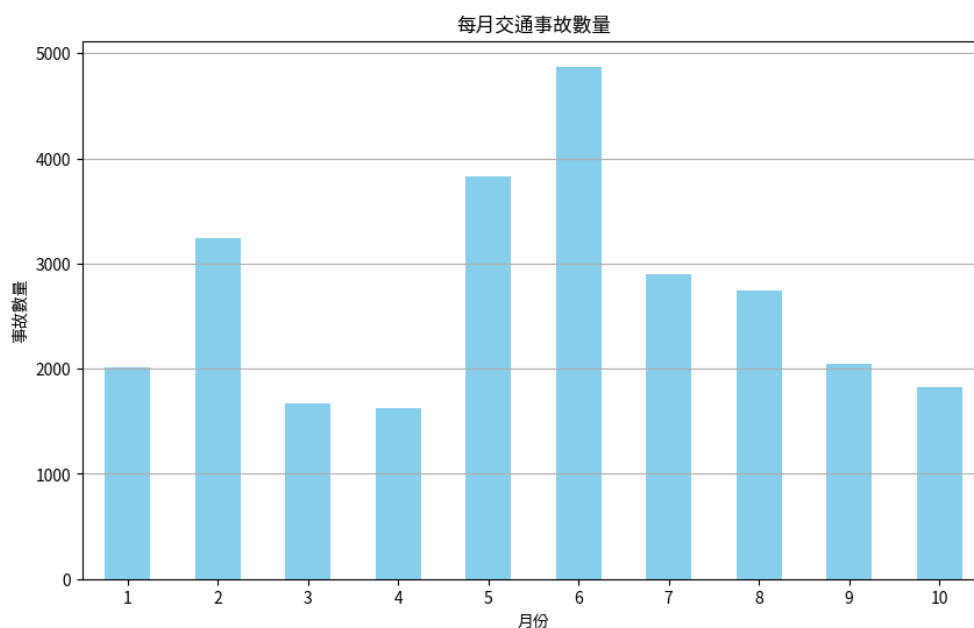


圖 21：112 年 1~10 月中不同月份的交通事故數量圖

從圖 21 中我們可以看出交通事故的高峰期 5、6 月，而低峰期則在 3、4 月，因此我們接著去比較了 4 月、6 月和全年的 24 小時交通事故圖，即圖 22，從中可知這兩個月和整年的分布並無明顯差異，說明了這兩個月分和全年的交通事故高發時段趨勢一致，都集中在上下班的高峰時段，因此我們並不能解釋為何兩個月分的故事數量有如此顯著的差異。為了進一步分析原因，我們藉由圖 23 做進一步分析。由圖 23 我們可以看到 4 月的事故高發時間為 1

號，也就是清明連假。而 6 月的事故高發時間為端午連假，另外每逢周五的也是事件發生的高峰期，因此後續在建模上將加入變數用以區分周末、平日和節日。

由於此處的數據有明顯不合理之處，例如：清明連假的事故發生率比 6 月的大部分天數(包含平日)都還要低。因此我們額外去確認了 112 年 1-10 月 A1、2、3 事故資料在 1~10 月的資料中，國道事故的資料筆數，發現 4 月有許多事故並未被正確的紀錄到交通事故簡訊通報資料集中，若是以此資料級的觀察結果作現實世界的分析，例如：因 4 月事故量較少，所以其他月份可參照四月時的警力佈署狀況等等，會有許多問題，因為事實上 4 月的事故數與其他月份並無明顯差異，因此此處的分析僅建議用於建模使用。

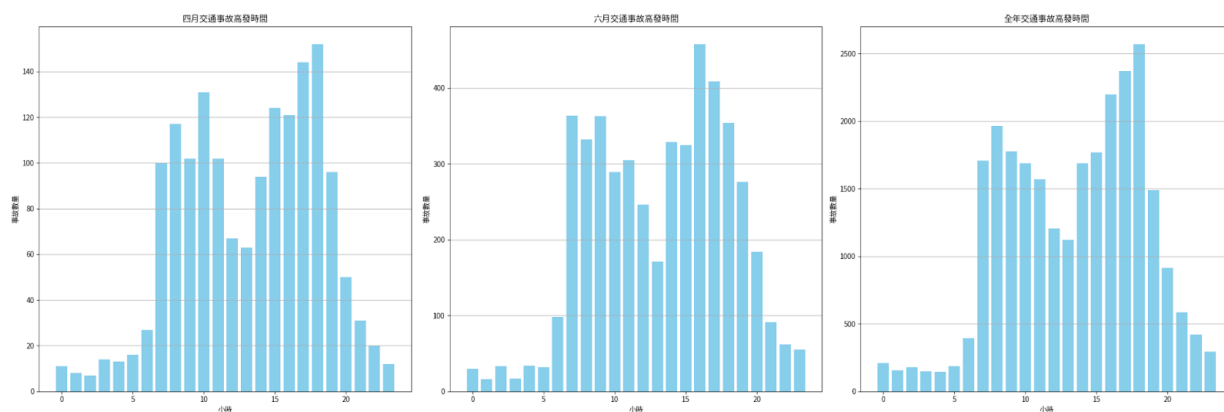


圖 22: 4 月、6 月和全年的 24 小時交通事故圖

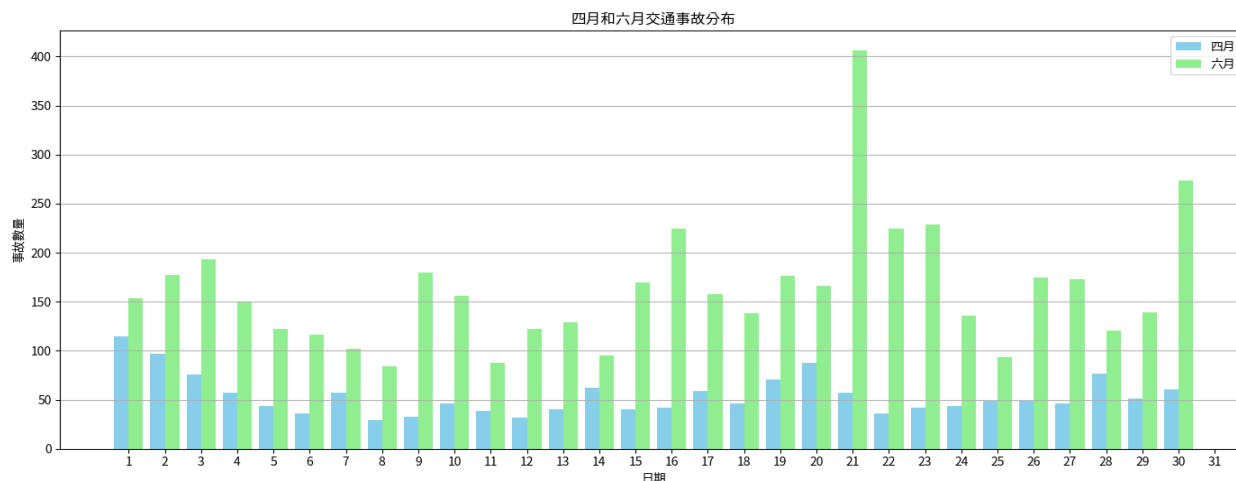


圖 23: 4 月、6 月的日交通事故圖

附錄 B、事故排除時間預測模型之資料前處理

- 處理變數：肇事車輛種類及數量

將通過建立變數「車輛 1」~「車輛 12」中的所有車輛種類的字典，將變數「車輛 1」~「車輛 12」轉由西建立的變數「大型車數量」、「小型車數量」、「其他」來做表示，最後將變數「車輛 1」~「車輛 12」以及「肇事車輛」刪除。此步驟的目的為將資料維度做縮減，由原先 13 個變數轉由僅使用 3 個變數表示，並解決變數「車輛 1」~「車輛 12」缺失值問題。

- 處理變數：年、月、日、時、分、事件發生、事件排除

通過刪除變數「事件排除」，避免模型在訓練時，通過「事件發生」、「事件排除」，得知事故發生的答案，導致結果過於樂觀。

將變數「年、月、日」合併轉為 date.time 格式，以利模型使用。將變數「事件發生」做清洗並建立防呆，以防人工輸入的資料出現不合格式的狀況。最後將「年、月、日、時、分」刪除。此目的主要為減少變數數量。

- 處理變數：事故類型

將種類名稱更改為數值表示，以利模型做訓練，並且因 A1 最嚴重，所以用 3 表示，而 A3 最不嚴重，所以用 1 表示。

- 處理變數：事故類型 將「事故類型」中的類別 A1、A2、A3 更改為數值表示，以利模型做訓練，並且因經計算，A1、A2、A3 和平均事故處理時間具有高度正相關，因此保留並將 A1 事故用 3 表示，而 A3 事故用 1 表示。
- 處理變數：簡訊內容 將簡訊內容中涉及事故排除時間的訊息拿掉，確保答案不會洩漏給模型，並將剩餘的內容利用 Bert 模型轉換為向量(Vector)的形式。
- 刪除變數：交控中心接獲通報，CCTV 監看現場，CMS 發布資訊，交控中心通報工務段，事故處理小組出發，事故處理小組抵達，事故處理小組完成 將缺失值超過 32%的變數刪除，此法除了是為了能符合常態假設，方便整理資料外，另外若缺失值過多會導致需要做許多的假設去填入缺失值，使資料不合真實狀況。
- 其餘變數則使用 AutoEDA 的技術讓模型自行去做學習及調整。

交通部高速公路局
113 年國道智慧交通管理創意競賽
投稿作品 3 年內是否公開獲獎切結書

附件 4

立切結書人 陳雅柔 等參加高速公路局舉辦之 113 年國道智慧交通管理創意競賽所投稿之作品

☒ 未曾於 3 年內公開獲獎

☐ 曾於 3 年內以 _____ (作品名)，公開參加 _____ (競賽)，獲得 _____ (獎項)，現投稿之作品，與之前獲獎作品有顯著之差異性，且已於報告中敘明其前後差異，若經主辦單位審核發現參賽作品有違反本比賽規則所規範事項者，本人同意被取消參賽資格；如已得獎，亦同意被追回已頒發之獎項，絕無異議。

立切結書人：(須全體成員簽章)

姓名及簽章 身分證字號

聯絡電話及戶籍地址

| | | |
|------------|-------------------|---------------------------------------|
| <u>陳雅柔</u> | <u>N226212148</u> | <u>0907094886 彰化縣大村鄉福興村山腳路106之47號</u> |
| <u>邱奕鈺</u> | <u>B123599217</u> | <u>0989270310 台中市西區廣民里12鄰大明街7之5號</u> |
| <u>林子謙</u> | <u>L125587428</u> | <u>0905380586 台中市大甲區庄美里1鄰光明路232號</u> |
| <u>陳碩鈞</u> | <u>O200585335</u> | <u>0989597876 新竹市八德路185號6樓</u> |
| <u>許光碩</u> | <u>V121681611</u> | <u>0972831806 台東縣關山鎮德高里東明21號</u> |
| <u>李宇銘</u> | <u>K122938110</u> | <u>0937828509 桃園市中壢區長春路267號2樓</u> |
| <u>余林昶</u> | <u>H125933357</u> | <u>0987791396 桃園市中壢區華祥一街三巷10號5樓</u> |
| _____ | _____ | _____ |
| _____ | _____ | _____ |
| _____ | _____ | _____ |

中 華 民 國 113 年 5 月 8 日