

智慧交通的未來

透過事故預測提升國道交通效率

隊伍名稱：梅林麵

## 目錄

|                        |    |
|------------------------|----|
| 封面.....                | 1  |
| 目錄.....                | 2  |
| 圖目錄.....               | 4  |
| (一)摘要.....             | 6  |
| (二)研究目的與研究範疇.....      | 6  |
| 2.1 研究目的.....          | 6  |
| 2.2 研究範疇.....          | 7  |
| (三)現況與問題盤點.....        | 7  |
| 3.1 現況.....            | 7  |
| (四)研究方法與步驟.....        | 8  |
| 4.1 資料來源.....          | 9  |
| 4.2 探索性資料分析.....       | 10 |
| 4.3 資料前處理.....         | 13 |
| 4.3.1 變數解釋.....        | 14 |
| 4.4 資料的拆分與分段建模.....    | 14 |
| 4.5 選擇最佳模型並代入驗證資料..... | 15 |
| (五)分析結果與討論.....        | 15 |
| 5.1 處理時間.....          | 15 |
| 5.1.1 訓練模型.....        | 15 |
| 5.1.2 驗證.....          | 17 |

|                   |    |
|-------------------|----|
| 5.2 回堵長度.....     | 19 |
| 5.2.1 訓練模型.....   | 19 |
| 5.2.2 驗證.....     | 20 |
| 5.3 討論.....       | 21 |
| (六)推廣與應用.....     | 22 |
| (七)參考文獻與參考資料..... | 24 |

## 圖目錄

|                                   |    |
|-----------------------------------|----|
| 圖3-1:處理時間盒狀圖.....                 | 8  |
| 圖3-2:處理時間之離群值.....                | 8  |
| 圖4-1:預測車禍處理時間之研究流程圖.....          | 9  |
| 圖4-2:預測車禍回堵長度之研究流程圖.....          | 10 |
| 圖4-3:處理時間盒狀圖與outlier分布情況.....     | 10 |
| 圖4-4:回堵長度分布情況.....                | 10 |
| 圖4-5:國道名稱出現的比例.....               | 11 |
| 圖4-6:事故類型的比例.....                 | 11 |
| 圖4-7:肇事車輛(數量)的分布情況.....           | 12 |
| 圖4-8:肇事車輛(車種)的分布情況.....           | 12 |
| 圖4-9:事故發生車道分布情況.....              | 13 |
| 圖4-10:國道一號樣本數熱圖.....              | 15 |
| 圖5-1:XGBoost.....                 | 16 |
| 圖5-2:處理時間對數轉換+XGBoost.....        | 16 |
| 圖5-3:分位數迴歸模型(中位數 左)(最佳分位數 右)..... | 16 |
| 圖5-4:分位數迴歸模型+XGBoost.....         | 16 |
| 圖5-5:變數重要性.....                   | 16 |
| 圖5-6:各模型RMSE與MAE之比較.....          | 17 |
| 圖5-7: XGBoost.....                | 17 |
| 圖5-8:處理時間對數轉換+XGBoost.....        | 17 |

|                             |    |
|-----------------------------|----|
| 圖5-9:分位數迴歸(中位數).....        | 17 |
| 圖5-10:分位數迴歸(最佳分位數).....     | 17 |
| 圖5-11:分位數迴歸+XGBoost.....    | 18 |
| 圖5-12:各模型RMSE與MAE之比較.....   | 18 |
| 圖5-13: XGBoost.....         | 18 |
| 圖5-14:處理時間對數轉換+XGBoost..... | 18 |
| 圖5-15:分位數迴歸(中位數).....       | 18 |
| 圖5-16:分位數迴歸(最佳分位數).....     | 18 |
| 圖5-17:分位數迴歸+XGBoost.....    | 19 |
| 圖5-18:各模型RMSE與MAE之比較.....   | 19 |
| 圖5-19: 零膨脹模型+XGBoost.....   | 20 |
| 圖5-20:分位數迴歸模型+XGBoost.....  | 20 |
| 圖5-21:各模型RMSE與MAE之比較.....   | 20 |
| 圖5-22:變數重要性.....            | 20 |
| 圖5-23:零膨脹模型+XGBoost.....    | 20 |
| 圖5-24:分位數迴歸模型+XGBoost.....  | 20 |
| 圖5-25:各模型RMSE與MAE之比較.....   | 21 |
| 圖5-26:零膨脹模型+XGBoost.....    | 21 |
| 圖5-27:分位數迴歸模型+XGBoost.....  | 21 |
| 圖5-28:各模型RMSE與MAE之比較.....   | 21 |

## (一)摘要

通過分析交通事故資料，我們利用 XGBoost 結合分位數迴歸建立了預測處理時間的統計模型，同時也使用 XGBoost 結合零膨脹模型來預測回堵長度。結果顯示，事故類型與處理時間之間的關係最為顯著，而回堵長度則與事故發生的車道最為相關，特別是發生在匝道的影響尤為顯著。我們希望通過建立即時預測系統，來減少交通堵塞，降低因塞車引發的次生車禍風險及怠速造成的空氣污染。此外，這些技術在環境永續方面具有應用潛力，例如提升國道大眾運輸工具的效能，以提升交通效率、減少環境影響，並促進社會的可持續發展。

## (二)研究目的與研究範疇

### 2.1 研究目的

高速公路車禍對社會、經濟、環境層面均產生深遠影響。在社會層面，車禍可能導致嚴重的人員傷亡，給受害者及其家庭帶來巨大心理創傷和經濟壓力。此外，車禍還會引發交通堵塞，特別是在高峰時段，進一步影響公共交通系統的運行效率，增加通勤時間，並浪費社會資源。在經濟層面，車禍引發的直接和間接成本不可忽視。直接成本包括醫療費用、車輛修理及保險理賠，間接成本則涉及生產力損失、企業運營中斷及公共基礎設施損壞。在環境層面，首先，車禍導致的交通堵塞會增加空氣污染。當車輛在堵塞中長時間怠速運行時，會排放大量的廢氣，包括一氧化碳、氮氧化物和揮發性有機化合物，這些污染物會加劇空氣質量的惡化。此外，堵塞延長了車輛行駛的時間，導致更多的二氧化碳排放，進一步加劇氣候變遷。

本研究旨在分析高速公路交通事件的處理時間及其影響因素。首先，通過數據分析找出容易發生車禍地點的特徵，協助政府決策是否需要調整該路段的速限或優化交通設施。其次，建立預測模型以估算車禍處理時間及回堵長度，期望在車禍發生時能夠迅速預測處理時間，並通過簡訊通知附近路段的駕駛員，以便他們決定是否更改路線，從而減少交通堵塞。此外，本研究還將對A1、A2、A3類交通事故進行時間分析、肇事原因分析及車種分析，協助相關部門針對性地改善交通狀況。同時，將研究結果與過往資料進行比較，以驗證當局政策的有效性。

## 2.2 研究範疇

本研究將使用112年1月至10月期間全台高速公路各路段的車禍數據及車禍通報的簡訊內容作為資料來源。研究對象涵蓋不同地點、不同時間及不同事故嚴重程度下的交通事件。為確保研究結果的針對性，本研究僅聚焦於高速公路上的交通事件，並不涉及市區道路或其他低速公路的事故情況。

### (三)現況與問題盤點

#### 3.1 現況

根據交通部高速公路局所發布的「交通部高速公路局111年國道事故檢討分析報告」，可以得知國道1號（含高架路段，總長432.5公里）及國道3號（總長432.9公里）作為主要交通動脈，在111年間的事件件數及交通量約占國道路網的9成。相比之下，其餘國道因交通量較低（僅占10.1%），因此事件件數也相對較少，占總數的10.07%(交通部高速公路局，2023)。

此外，事故發生的路段和時間分布存在顯著差異，事發路段大多集中在國道一號和三號，起始點至100公里的範圍內，時間主要集中在上午6:00至晚上9:00。由於時間變數具有明顯的週期性特徵，因此使用一般線性迴歸模型可能無法充分捕捉這些特性，效果可能不佳。其次，我們的分析目標是「車禍事件處理完成的時間」以及「車禍事件的回堵長度」，由下方(圖3-1)可以看出，在國道一號的17742筆資料中，處理時間大多集中在5分鐘至25分鐘之間，而離群值(outliers)共計879筆，占整體的約5%。值得注意的是，下方(圖3-2)顯示部分離群值超過200分鐘，如果預測目標是「平均處理時間」，這些離群值可能會對預測結果造成顯著影響。

吳承紘(2014)利用存活分析法中的Cox等比例危險模式，找出高速公路上事故排除時間的顯著影響變數，並構建兩者之關係。結果顯示：在國道1號與國道3號上，事故後占用越多車道、涉及越多車輛以及受傷人數越多，皆與事故排除時間呈現正相關；晚上相較於白天事故排除時間也較長。是否需工務段出勤協助排除事故只在國道1號上呈現顯著影響，且如需工務段協勤之事故排除時間較長。協助救護之管轄消防局與事故現場間之距離

則在兩條高速公路上皆顯著與事故排除時間呈現正相關，亦即救護距離會影響事故排除時間。

我們決定嘗試使用機器學習演算法中的 XGBoost (Chen & Guestrin, 2016) 搭配分位數迴歸 (Quantile Regression; Koenker & Bassett, 1978) 進行分析。由於資料中車禍事件造成的回堵長度大部分為 0 公里，即無回堵，儘管分位數迴歸模型也可以處理此類資料，我們決定將其與零膨脹模型 (Zero-Inflated Model; Lambert, 1992) 進行比較，並最終選擇表現較佳的模型來進行分析。

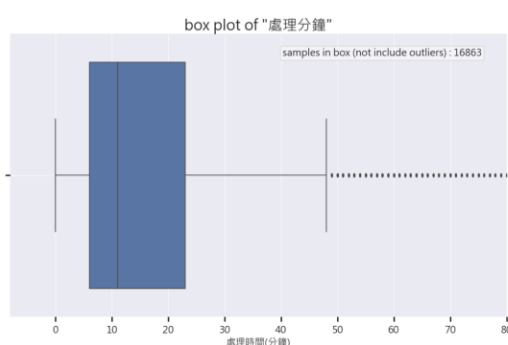


圖3-1:處理時間盒狀圖

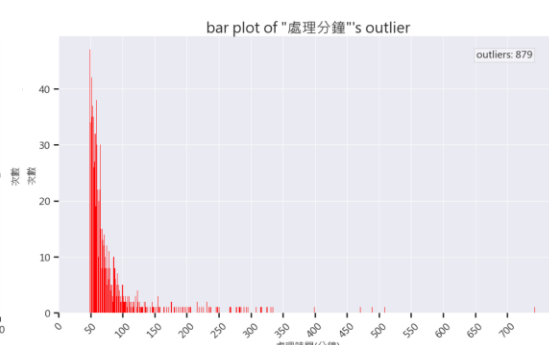


圖3-2:處理時間之離群值

XGBoost是一種基於梯度提升 (Gradient Boosting) 的強大機器學習演算法，由陳天奇開發。作為梯度提升樹 (Gradient Boosting Trees) 模型的高效實現，XGBoost的目的是提升模型的計算速度和效能。它通過結合多個弱學習器 (通常是決策樹) 來提高模型的準確性，並且擁有強大的參數調優功能和正則化支持，使其在處理過擬合問題上非常有效。

分位數迴歸模型的最大好處是它能夠提供對數據分布中不同位置 (或分位數) 的深入分析，而不僅僅是集中在平均數或中位數這樣的單一指標上。這種方法允許我們在解釋變數與響應變數之間的關係在數據的不同部分如何變化，這在處理具有異質性或極端值的數據集時尤其有用。此外，它還對異常值不敏感，因為它不僅僅依賴於平均值來描述數據，而是考慮到數據在不同分位數上的分布，這使得分析結果更具穩健性和洞察力。

零膨脹模型 (Zero-Inflated Model) 專門用來處理數據集中存在大量零值的情況。該模型假設數據由兩個過程生成：一個過程產生零值，另一



個過程根據某種分佈生成非零值。零膨脹模型能夠更好地捕捉數據中的零值過多的特性，適用於事件稀疏但仍有可能發生的情況，通過使用零膨脹模型，我們可以更準確地預測非零事件的發生情況，同時考慮數據中存在大量零值的現象。

國道交通量隨著路網的發展逐年增長，111年度的延車公里達到34,405百萬車公里，相較110年增加了6.2%。在111年度，A1類交通事故的件數、死亡及受傷人數較110年度有所減少；但A2、A3類事故則呈現增加的趨勢。整體而言，過去10年（102至111年）A1類事故件數大致穩定（約介於46至77件之間）；而A2及A3類事故件數則隨著延車公里的增加而增加，兩者之間呈現高度相關性。雖然我們的資料中並未包含延車公里數據，但可以看出交通堵塞對事故發生率的影響。如果能在事故發生的當下迅速預測處理時間和回堵長度，並採取措施避免短時間內發生多起車禍，不僅能維持交通順暢，更能有效降低事故發生率。

#### （四）研究方法與步驟

##### 4.1 資料來源

資料來源於交通部高速公路局所提供的「113年國道智慧交通管理創意競賽」資料集，涵蓋了112年1至10月及113年1至2月國道A1、A2、A3的交通事故數據，以及交通事故簡訊通報狀況資料。樣本總數為17,742件。交通事故資料的內容包括：事故類別、路線、方向、里程、發生時間、天候、車種、傷亡人數、車速、事故型態及肇因等。交通事故簡訊通報狀況資料則包含：路線、方向、里程、發生時間、排除時間、傷亡人數、佔用車道、車種、事故型態及回堵狀況等。

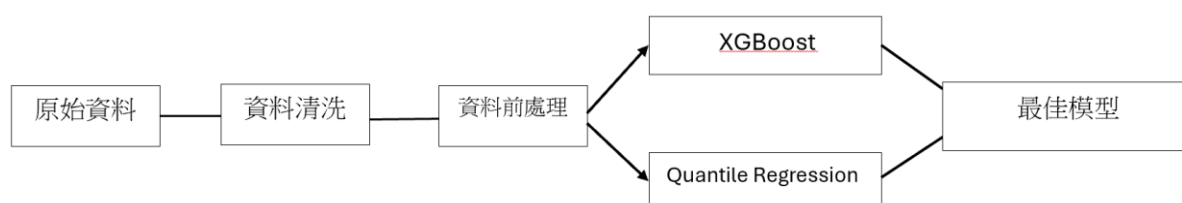


圖4-1:預測車禍處理時間之研究流程圖

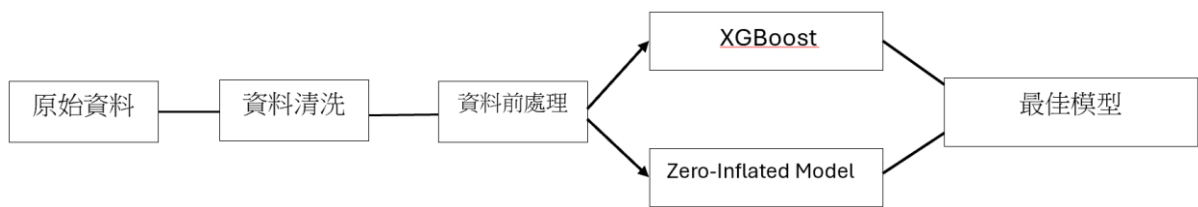


圖4-2:預測車禍回堵長度之研究流程圖

## 4.2 探索性資料分析

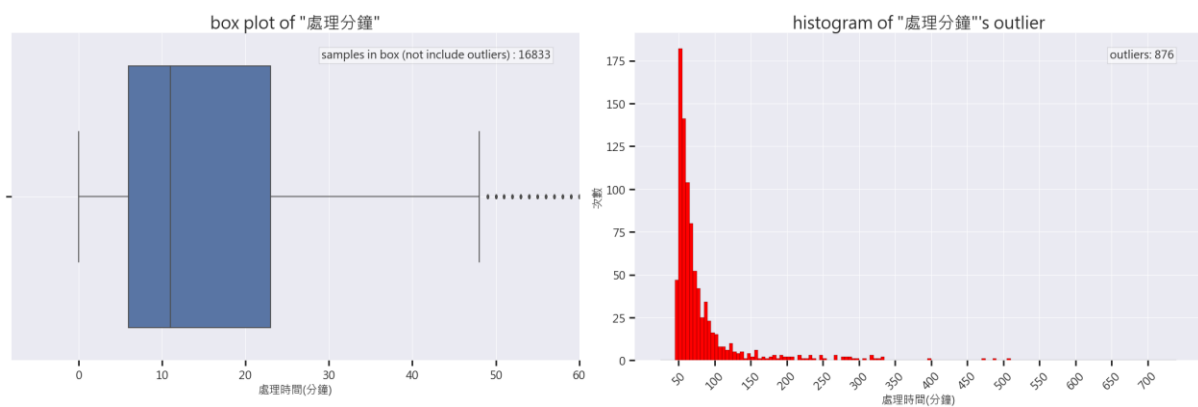


圖4-3:處理時間盒狀圖與outlier分布情況

處理時間大多集中在5分鐘至25分鐘之間，但有5%的數據屬於明顯較高的離群值，因此可以知道資料不適合用來估計平均值。

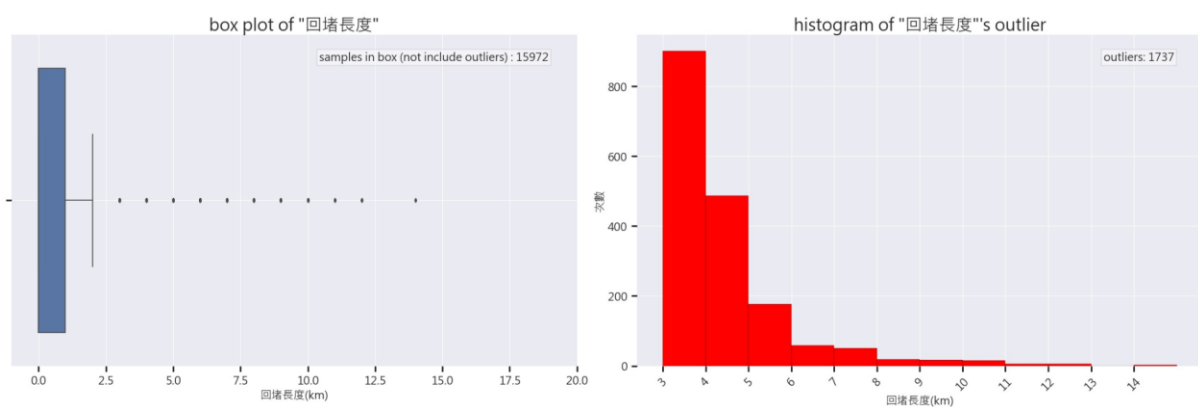


圖4-4:回堵長度分布情況

回堵長度大多集中在0公里至4公里，且分布情況明顯不符合常態分布，且無回堵的比例太高，也就是資料有太多0值存在，我們決定使用零膨脹模型。

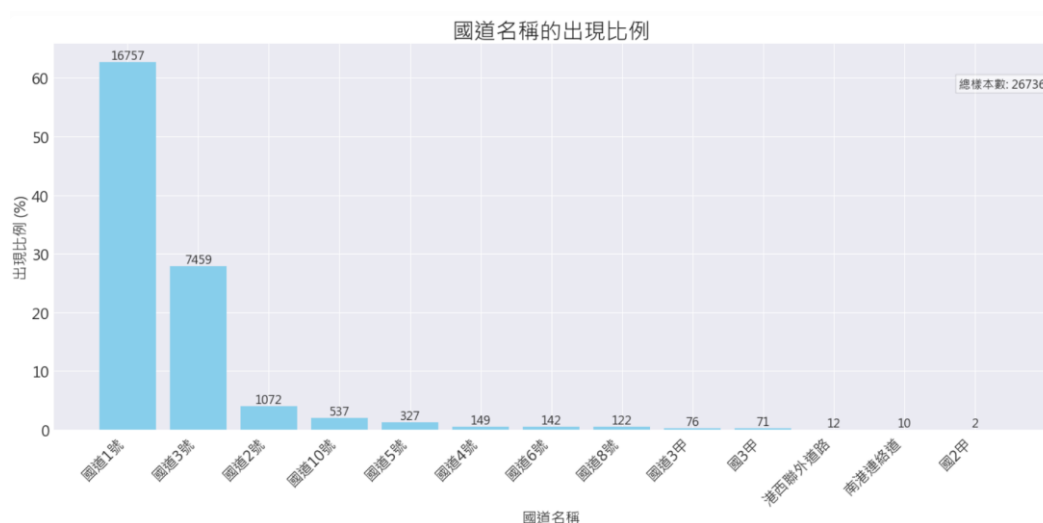


圖4-5:國道名稱出現的比例

從上圖4-3可以看到樣本集中在國道1號和國道3號。

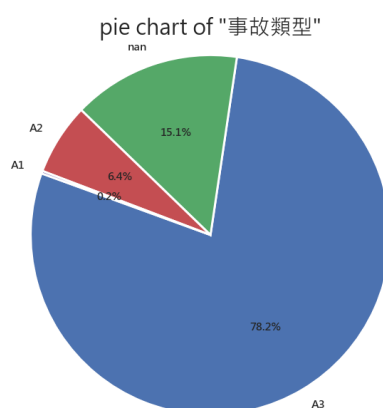


圖4-6:事故類型的比例

A1、A2、A3交通事件的定義如下：A1交通事故是指造成人員死亡的交通事故，通常是指在事故發生後24小時內有人因事故喪生的情況；A2交通事故是指造成人員受傷但無人死亡的交通事故，包括任何因事故需要接受治療的情況；A3交通事故則是指僅涉及財產損失而無人員傷亡的交通事故，即事故僅造成車輛或其他財物的損壞，未有任何人員受傷或死亡。從上圖可以看見事故類型多為A3。

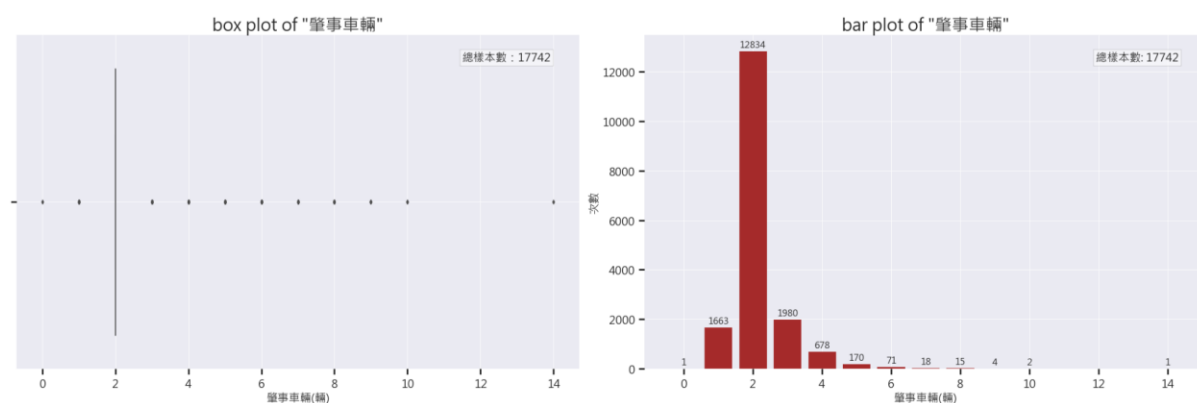


圖4-7: 肇事車輛(數量)的分布情況

肇事車輛為1輛時，表示為自撞事故，從上圖中可以看到肇事車輛多為2輛，顯示多數車禍是由兩輛不同的車輛所導致。

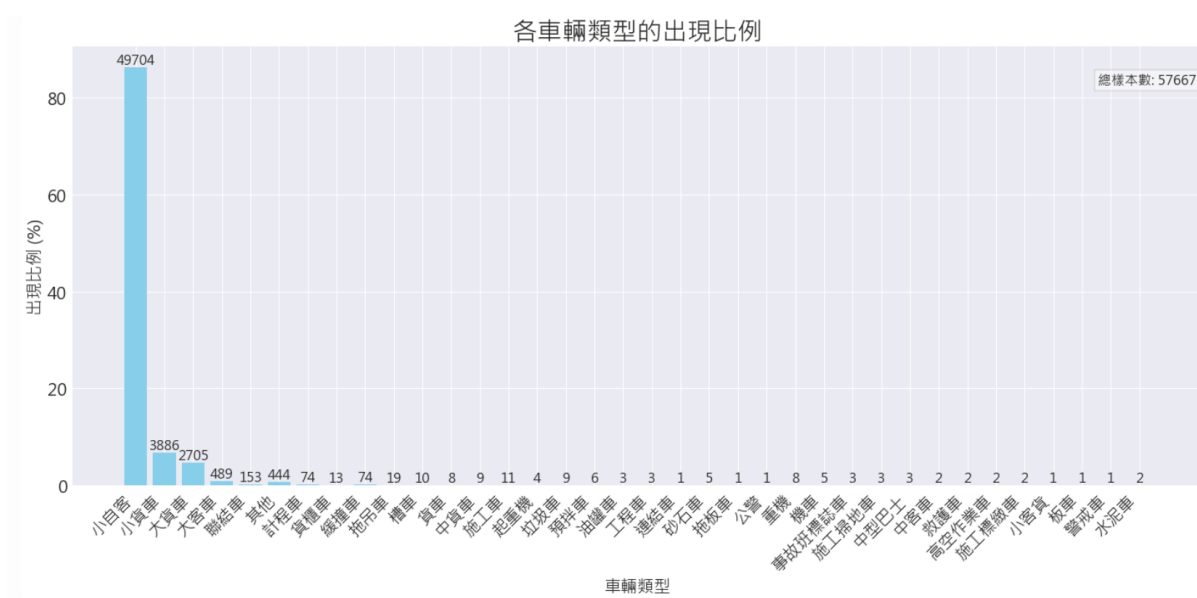


圖4-8: 肇事車輛(車種)的分布情況

從上圖可以看到大多數都是由自用小客車造成的車禍

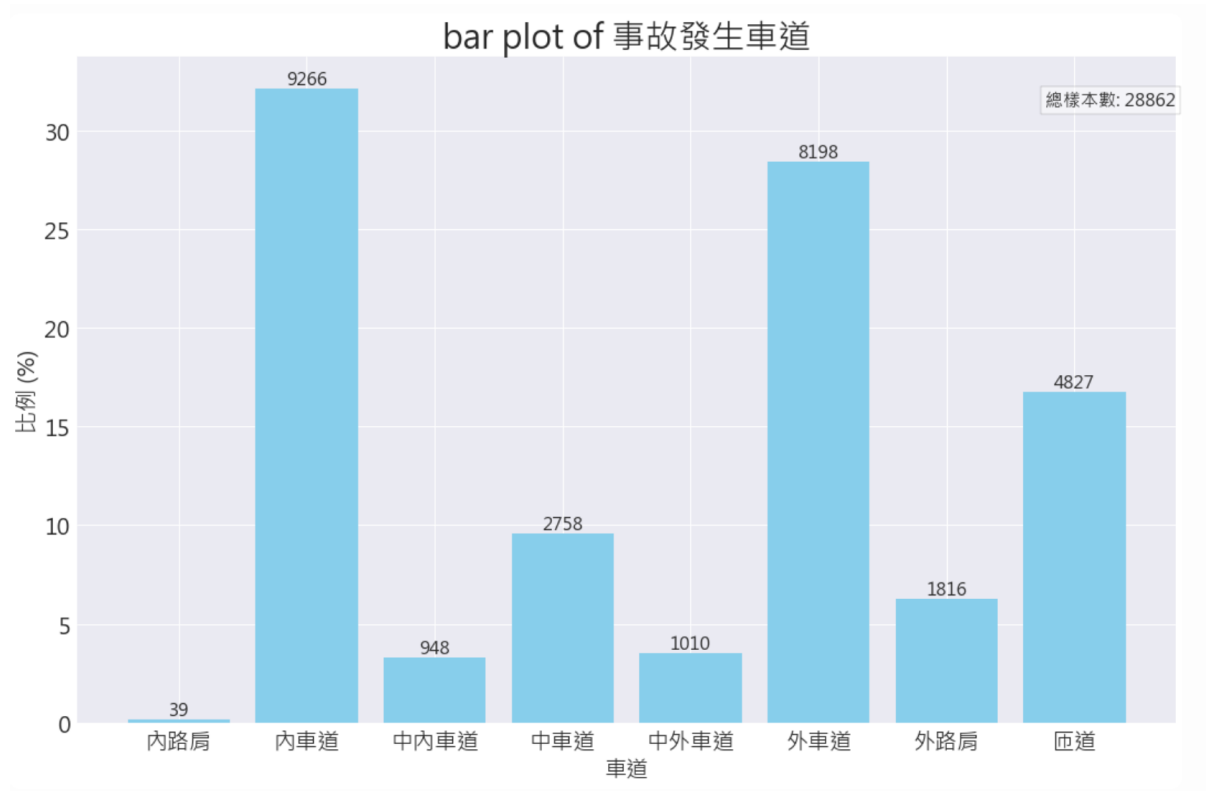


圖4-9: 事故發生車道分布情況

大多發生於內車道和外車道，初步猜測是內車道車速較快，且台灣人習慣長時間使用內車道，而外車道和匝道的部分可能是因為較頻繁的車道切換而導致事故發生。

#### 4.3 資料前處理

我們選用了以下變數來配適模型：國道名稱、里程、方向、事故類型、車道、肇事車輛、是否為假日、相近時間是否有事件。

##### 4.3.1 變數解釋

1. 處理時間：連續型變數，單位分鐘。
2. 回堵長度：連續型變數，單位公里。
3. 國道名稱：類別型變數，資料內共有13條不同的國道名稱。

4. 里程：連續型變數，不同國道長度不同，終點里程也就不同，單位公里。

5. 方向：類別型變數，分為東、西、南、北。

6. 事故類型：類別型變數，分為A1、A2、A3。

7. 車道：類別型變數，分為內路肩、內車道、中內車道、中車道、中外車道、外車道、路肩、匝道。

8. 肇事車輛：連續型變數，0輛至14輛。

9. 是否為假日：二元類別型變數。

10. 相近時間地點是否有事件：二元類別型變數，在A車禍發生至事件排除的時間區間內，發生地點後方10公里內是否有發生車禍，主要是想看車禍發生是否會導致另一起車禍發生的可能性增加，並造成處理時間和回堵長度增加。

#### 4.4 資料的拆分與分段建模

我們將資料的 80% 設為訓練集，20% 設置為測試集。另外，由於不同國道其長度不同，以及事故發生時間並不是一個容易分析的變數，我們將資料以國道、里程區間、事故發生時間區間做分段，里程區間一單位為50公里，事故發生時間區間一單位為3小時，下圖以國道一號做舉例，可以看到資料分布不平均，多數集中在0公里至50公里的區間，且時間為6:00至21:00，也就是車流量最大的時候。

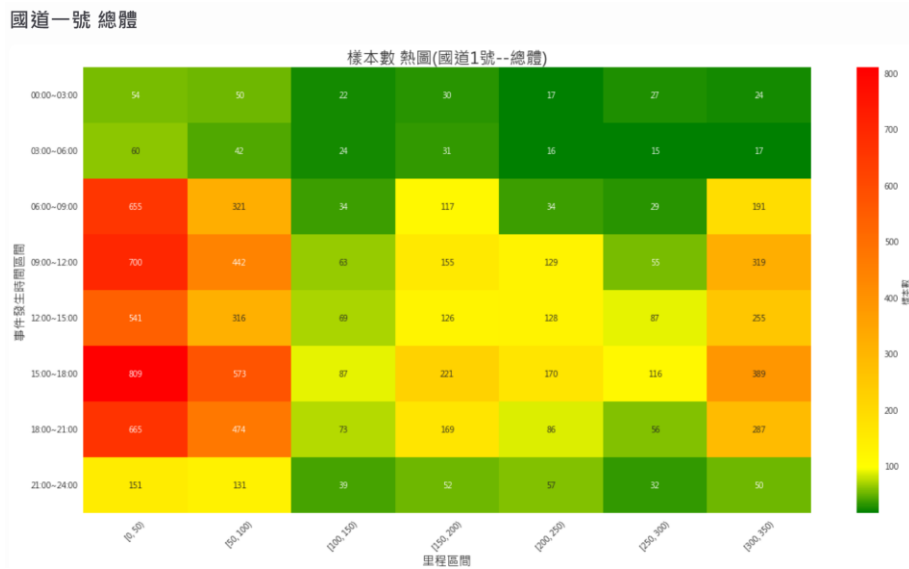


圖4-10:國道一號樣本數熱圖

#### 4.5 選擇最佳模型並代入驗證資料

以處理時間為例，我們將嘗試使用僅包含XGBoost的模型、僅包含分位數迴歸的模型，以及結合兩者的混合模型，並通過分析殘差來判斷最佳模型。接著，我們會將最佳模型應用於1至2月的數據，進行分析和預測效果的評估。

### (五)分析結果與討論

#### 5.1 處理時間:

##### 5.1.1 訓練模型:

以國道一號里程數0公里至50公里、時間範圍6:00至9:00的這組樣本為例，我們嘗試了三種模型：XGBoost、分位數迴歸模型、以及兩者結合的模型。同時，為了減少異常值對預測結果的影響，我們對處理時間進行了對數轉換。最終，透過分析殘差分佈，我們發現，將處理時間取對數並使用結合模型的預測效果最佳。結果顯示，模型的估計誤差在5分鐘以內。此外，預測的時間會因為里程與時間的不同去選擇最適合的Quantile，但大致上都是50%到60%位數。

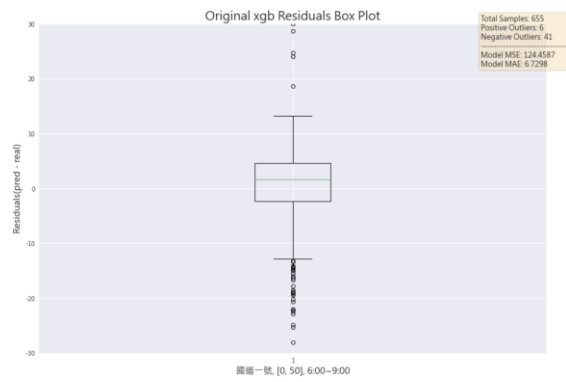


圖5-1:XGBoost

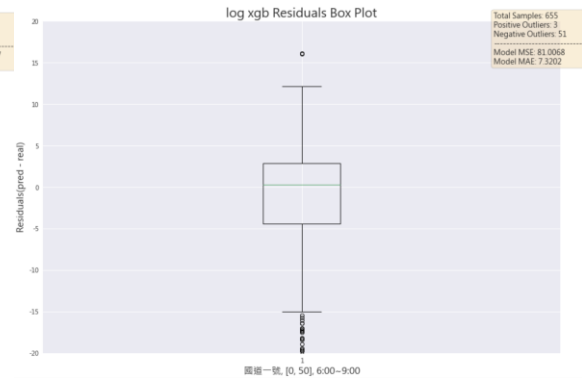


圖5-2:處理時間對數轉換+XGBoost

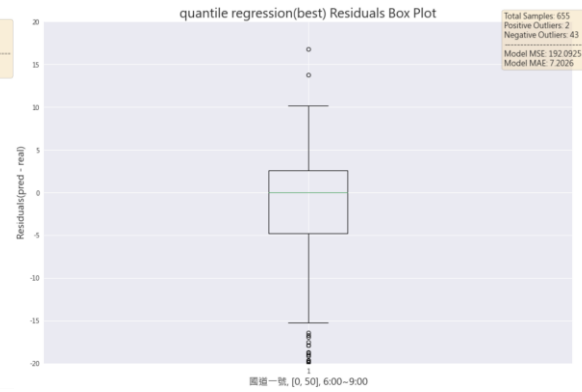
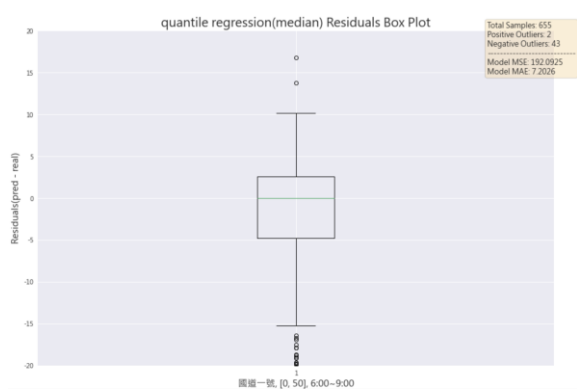


圖5-3:分位數迴歸模型(中位數 左)(最佳分位數 右)

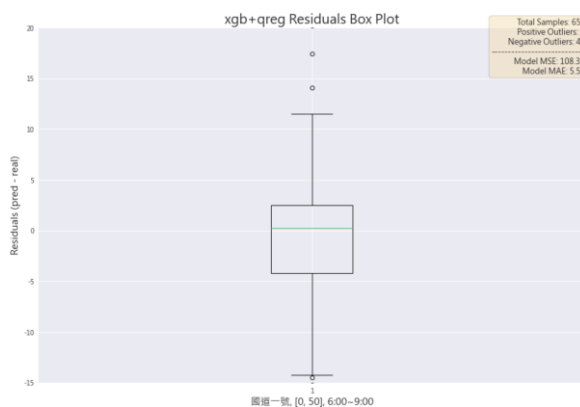


圖5-4:分位數迴歸模型+XGBoost

|    | Feature      | Importance |
|----|--------------|------------|
| 2  | 事故類型         | 0.378971   |
| 11 | 肇事車輛         | 0.213639   |
| 12 | 是否為假日        | 0.095842   |
| 14 | 事故車是否包含非小客車  | 0.080669   |
| 0  | 里程           | 0.057229   |
| 8  | 外車道          | 0.049481   |
| 9  | 外路肩          | 0.043949   |
| 7  | 中外車道         | 0.041342   |
| 5  | 中內車道         | 0.027722   |
| 6  | 中車道          | 0.011155   |
| 1  | 方向           | 0.000000   |
| 3  | 內路肩          | 0.000000   |
| 4  | 內車道          | 0.000000   |
| 10 | 匝道           | 0.000000   |
| 13 | 相近時間&地點是否有事件 | 0.000000   |

圖5-5:變數重要性



| 模型            | RMSE     | MAE    |
|---------------|----------|--------|
| XGBoost       | 11.15611 | 6.7298 |
| 對數轉換+XGBoost  | 9.000378 | 7.3202 |
| 分位數迴歸(中位數)    | 13.85974 | 7.2026 |
| 分位數迴歸(最佳)     | 13.85974 | 7.2026 |
| XGBoost+分位數迴歸 | 10.40913 | 5.57   |

圖5-6：各模型RMSE與MAE之比較

## 5.1.2 驗證

由113年國道一號1-2月的資料(0~50 6:00~9:00)帶入對應模型做驗證。可以看到各個模型的表現其實相差不大，誤差皆在5分鐘內，原因是驗證資料的outlier並不多，剛好都落在模型預測效果比較好的區間，倘若資料的極端值多一些，我們認為XGBoost+分位數迴歸模型仍會是最好的。其他組別的驗證結果皆附在附錄，如同4.4節所述，我們將國道一號和國道三號按照時間和里程分組，其餘的國道統一建立一個不分時間和里程的模型。

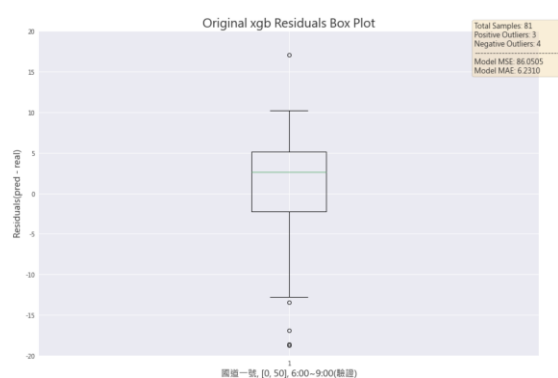


圖5-7:XGBoost

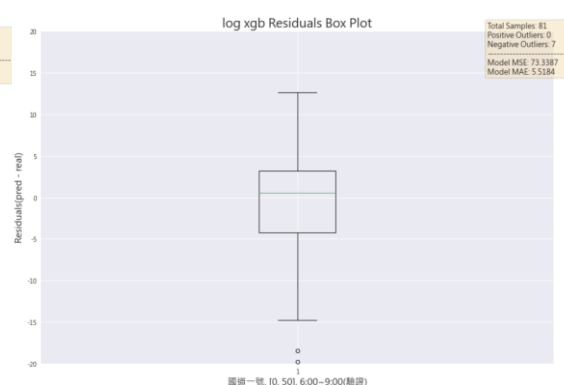


圖5-8:處理時間對數轉換+XGBoost

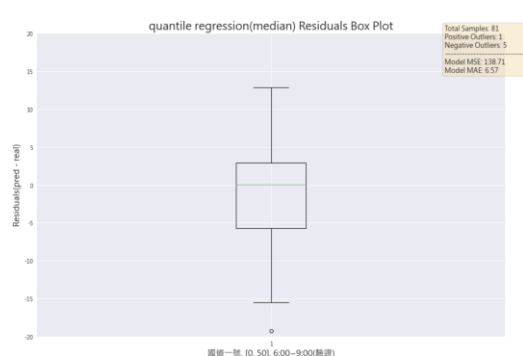


圖5-9:分位數迴歸(中位數)

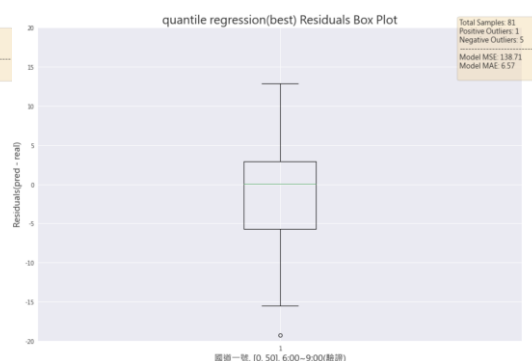


圖5-10:分位數迴歸(最佳分位數)

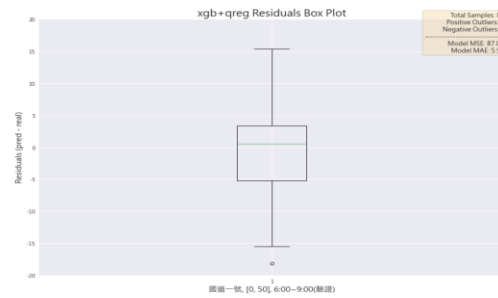


圖5-11:分位數迴歸+XGBoost

| 模型            | RMSE     | MAE    |
|---------------|----------|--------|
| XGBoost       | 11.15611 | 6.7298 |
| 對數轉換+XGBoost  | 9.000378 | 7.3202 |
| 分位數迴歸(中位數)    | 13.85974 | 7.2026 |
| 分位數迴歸(最佳)     | 13.85974 | 7.2026 |
| XGBoost+分位數迴歸 | 10.40913 | 5.57   |

圖5-12: 各模型RMSE與MAE之比較

下圖顯示了不分里程和時間的整體模型預測誤差。比較不分組的殘差與分組的殘差，可以看到不分組的殘差較大，但大多數誤差都在5分鐘左右。

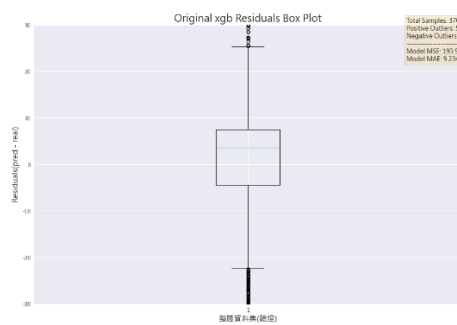


圖5-13:XGBoost

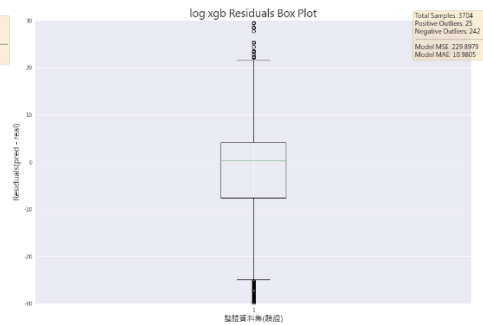


圖5-14:處理時間對數轉換+XGBoost

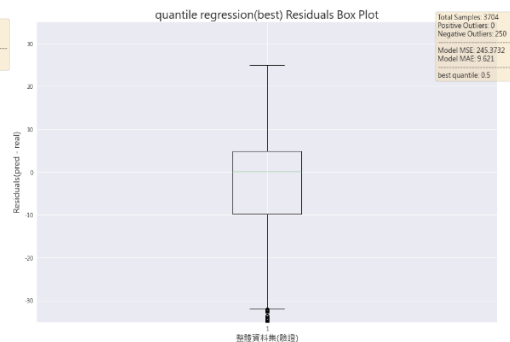
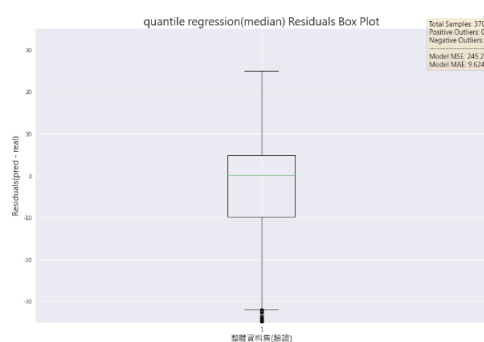


圖5-15:分位數迴歸(中位數)      圖5-16:分位數迴歸(最佳分位數)

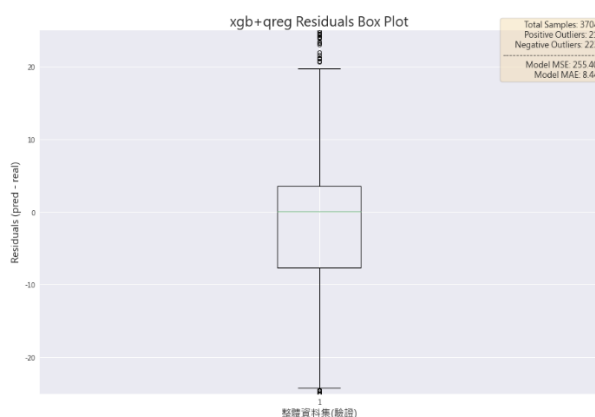


圖5-17:分位數迴歸+XGBoost

| 模型            | RMSE     | MAE     |
|---------------|----------|---------|
| XGBoost       | 13.92565 | 9.2349  |
| 對數轉換+XGBoost  | 15.16238 | 10.9805 |
| 分位數迴歸(中位數)    | 15.66191 | 9.6247  |
| 分位數迴歸(最佳)     | 15.66439 | 9.621   |
| XGBoost+分位數迴歸 | 15.98124 | 8.44    |

圖5-18: 各模型RMSE與MAE之比較

## 5.2 回堵長度

### 5.2.1 訓練模型

以國道一號里程數0公里至50公里、時間範圍6:00至9:00的這組樣本為例，我們將零膨脹模型與分位數迴歸模型進行了比較，最終選擇了表現較佳的零膨脹模型，可以看到零膨脹模型的outliers數量遠遠小於分位數迴歸模型，我們決定採用XGBoost結合零膨脹模型。結果顯示，該模型的多數估計誤差在2公里以內。

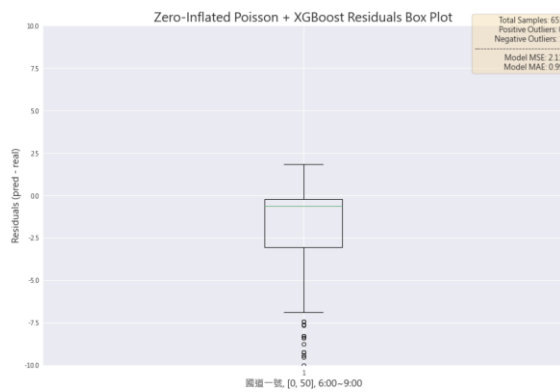


圖5-19: 零膨脹模型+XGBoost

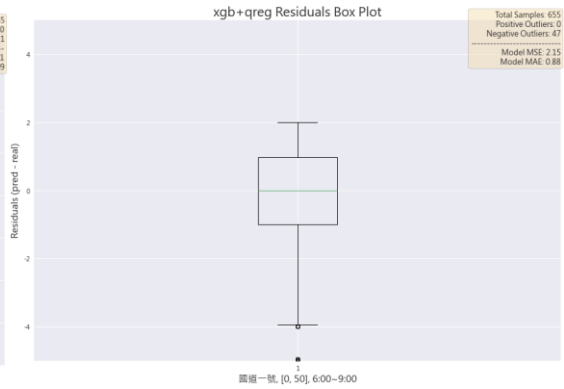


圖5-20: 分位數迴歸模型+XGBoost

|                       |          |      | Feature      | Importance (%) |
|-----------------------|----------|------|--------------|----------------|
| 模型                    | RMSE     | MAE  | 匝道           | 34.94          |
|                       |          |      | 內車道          | 8.43           |
| Zero-inflated+XGBoost | 1.452584 | 0.99 | 外車道          | 8.41           |
|                       |          |      | 中內車道         | 8.01           |
| 分位數迴歸+XGBoost         | 1.466288 | 0.88 | 中車道          | 6.98           |
|                       |          |      | 事故車是否包含非小客車  | 4.96           |
|                       |          |      | 是否為假日        | 4.83           |
|                       |          |      | 外路肩          | 4.79           |
|                       |          |      | 方向           | 4.02           |
|                       |          |      | 相近時間&地點是否有事件 | 3.34           |
|                       |          |      | 肇事車輛         | 3.11           |
|                       |          |      | 里程           | 3.10           |
|                       |          |      | 事故類型         | 3.04           |
|                       |          |      | 中外車道         | 2.04           |
|                       |          |      | 內路肩          | 0.01           |

圖5-21: 各模型RMSE與MAE之比較

圖5-22: 變數重要性

## 5.2.2 驗證

從下圖可以看到零膨脹模型+XGBoost較不容易高估回堵長度，且outlier只有一個。

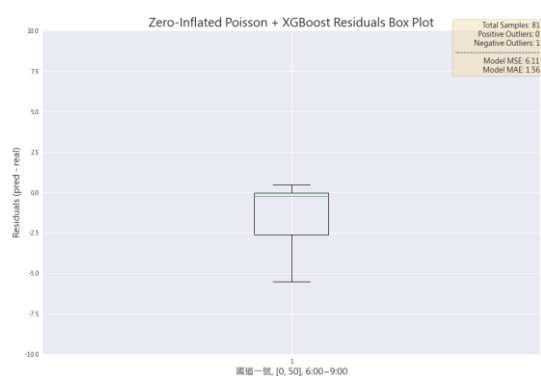


圖5-23: 零膨脹模型+XGBoost

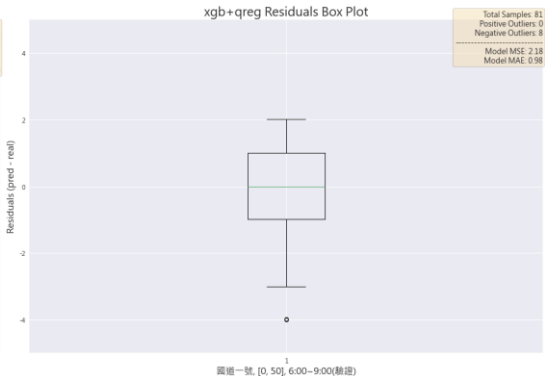


圖5-24: 分位數迴歸模型+XGBoost

| 模型                    | RMSE     | MAE  |
|-----------------------|----------|------|
| Zero-inflated+XGBoost | 2.471841 | 1.56 |
| 分位數迴歸+XGBoost         | 1.476482 | 0.98 |

圖5-25：各模型RMSE與MAE之比較

下圖顯示了不分里程和時間的整體模型預測誤差。比較不分組的殘差與分組的殘差，可以看到不分組的殘差較大，但大多數誤差都在1公里左右。從整體的模型也可以看到，不論是RMSE、MAE、outliers的數量，零膨脹模型結合XGBoost的模型表現都較佳。

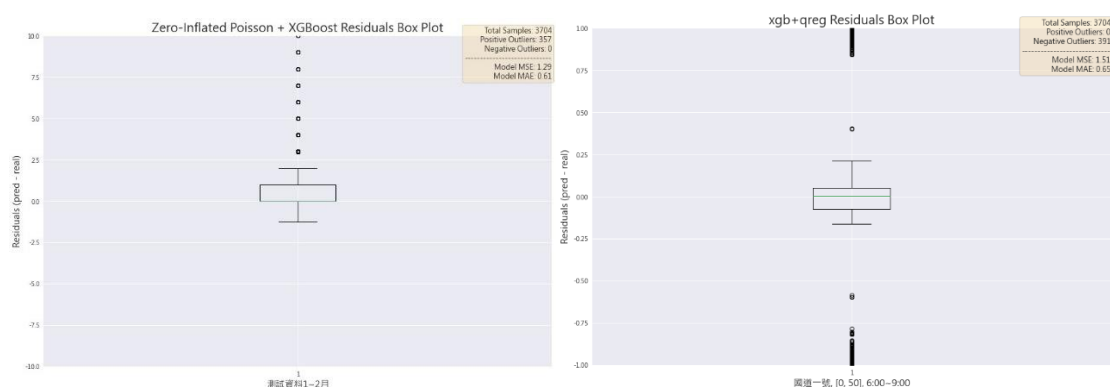


圖5-26：零膨脹模型+XGBoost 圖5-27：分位數迴歸模型+XGBoost

| 模型                    | RMSE     | MAE  |
|-----------------------|----------|------|
| Zero-inflated+XGBoost | 1.135782 | 0.61 |
| 分位數迴歸+XGBoost         | 1.228821 | 0.65 |

圖5-28：各模型RMSE與MAE之比較

最終我們決定使用分位數迴歸結合XGBoost來預測處理時間，使用零膨脹模型+XGBoost預測回堵長度。

### 5.3 討論

對於處理時間，從上圖中的變數重要性可以看到，首先，「事故類型」和「肇事車輛」影響最大，直觀上這兩個變數跟事件規模最相關，其次，「是否為假日」及「里程」和車流量有關，車流量越大，處理時間自然會越長，而「事故車是否包含非小客車」的關係到事件規模，但其影響並沒有「事故類型」和「肇事車輛」明顯，最後，車道的位置越靠外車道對處理時間的影響越明顯，外側車道通常是車輛進入和離開高速公路的主要通

道，承擔著更大的交通流量。當外側車道發生事故時，會影響更多的車輛通行，造成更大的交通壓力，從而延長事故處理的時間。

對於回堵長度，從上圖變數重要性中可以看到，影響回堵長度最主要的變數是車道，且匝道附近的車禍影響效果最顯著，匝道是車輛進出高速公路的主要通道，車流量較為集中。當車輛從匝道進入主線或離開主線時，會導致車流的變化，尤其是在高峰時段，容易造成車輛減速或停滯，增加回堵的風險。

我們認為，如果能夠再加入車流量相關的變數、各分段的最高速限，以及車禍地點與交流道距離等因素，將能夠進一步提升預測模型的準確性和效果。

## （六）推廣與應用

在應用方面，當車禍發生時，若能在通報時即確認相關資訊，便能透過模型預測出處理時間和回堵長度。此時，可以透過簡訊或APP通知該路段附近的駕駛者，讓他們根據預測結果，決定是否提前下交流道，或在上高速公路前更改路線。同時，還可提前啟動附近匝道的紅綠燈，調控車流量，縮減塞車範圍，從而避免因塞車引發的次生車禍，並降低怠速所造成的空氣污染和引擎耗損，達到環保的效果。

此外，目前資料中缺乏車流量相關的變數，僅憑里程與時間並無法準確計算車禍發生的機率。若能納入車流量變數進行預測，應用範圍將更為廣泛。例如，可以在車禍機率較高的路段適當降低速限，提醒駕駛人保持更寬裕的行車距離，並加強執行在高車流量時段禁止惡意超車和頻繁變換車道的規定。

在現今社會，環保意識日益提升，越來越多人選擇搭乘大眾運輸工具。然而，每當公車在高速公路上因車禍而導致大幅延誤時，不僅車上乘客的行程受影響，還可能波及他們所屬的公司、學校或政府部門，進而影響整個社會的運作。我們認為，結合車禍處理時間的預測與發生機率，能在此發揮重要作用。首先，公車在規劃國道路線時，可以根據車禍機率與處理時間計算更準確的行駛時間；其次，每當車禍發生時，公車站的誤點預告也能更為精確，讓乘客決定是否改搭其他班次，從而降低公車的不便利性，提升乘客的搭乘意願。保險公司亦可利用此模型設計針對公車的「旅遊不

便險」，依據乘客的搭乘路線計算保費和賠償金額，鼓勵更多人使用大眾運輸工具，形成良性循環。

此外，隨著Uber在台灣的普及，該服務的收費模式是根據路線預先確定金額，不因行車時間改變而調整。然而，若行駛國道途中發生車禍，便可能對駕駛者造成嚴重損失。我們建議在計算費用時應考慮上述模型的結果。此外，自動駕駛車輛也可事先設定，在車禍機率較高的路段保持更大的行車距離和較低的行駛速度，以降低風險。

當前技術的進步使得車禍處理時間的預測變得可行，並為多種應用場景提供了潛力。通過預測模型，我們能夠更精確地掌握車禍的處理時間和影響範圍，並據此做出相應的交通調控措施，能夠為大眾運輸、保險服務、自動駕駛等領域提供有力支持，從而促進整體交通環境的優化與可持續發展。這不僅提升了道路交通的效率，也對環保和社會運作產生積極影響。

## (七)參考文獻與參考資料

1. 交通部高速公路局。(2022)。111年國道事故檢討分析報告。交通部：高速公路局
2. 林煒宸(2024)分位數結合機器學習與神經網絡的新穎分析方法。國立陽明交通大學公共衛生研究所碩士論文
3. 吳承紘(2014)。高速公路事故排除時間影響因子之研究。國立交通大學運輸與物流管理學系碩士論文
4. So, B. (2024). Enhanced gradient boosting for zero-inflated insurance claims and comparative analysis of CatBoost, XGBoost, and LightGBM. *arXiv:2307.07771v3*.
5. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). Laurens Sluijterman, Frank
6. Kreuwel, Eric Cator, Tom Heskes(2024)Composite Quantile Regression With XGBoost Using the Novel Arctan Pinball Loss.*arXiv:2406.02293v1*.
7. Völz, B., Mielenz, H., Siegwart, R., & Nieto, J. (2016). Predicting pedestrian crossing using quantile regression forests. *2016 IEEE Intelligent Vehicles Symposium (IV)*.
8. Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1), 1-14.