



# KAMERA急診資料挑戰賽

Wu, Kuan-Hshine Chen, Jun-Min

## 1. 介紹

- 過去這幾年，急診擁塞一直是許多國家面臨到的重大問題。無法預測的天災、人禍、人力資源等問題更加使得急診擁塞問題難以解決。而緊急醫療也是社會安全的重要基礎，要如何改善急診擁塞的問題，以達到最大人力資源效益，已成為目前社會急迫解決的議題。
- 透過使用大數據分析、工具，我們的團隊將分析及研究出一套適合預測台灣高屏地區急診擁塞程度的工具。
- 比賽資料：主辦單位提供2013.01.01 ~ 2013.12.31高屏地區11家醫院與各個時段班次的急診就診人數，目標是預測2014年第33週（2014.08.11 ~ 2014.08.17）之每間醫院個時段班次之急診就診人數。
  - ◆ Problem: 由於主辦單位並無提供任何有關測試集相關之資料，因此在預測方面也更加困難，對於就診人數的影響因素也不確定，所以需要額外去取得資料來看是否是影響就診人數的因素。
  - ◆ Solution: 找可能影響急診就診人數的環境因子。

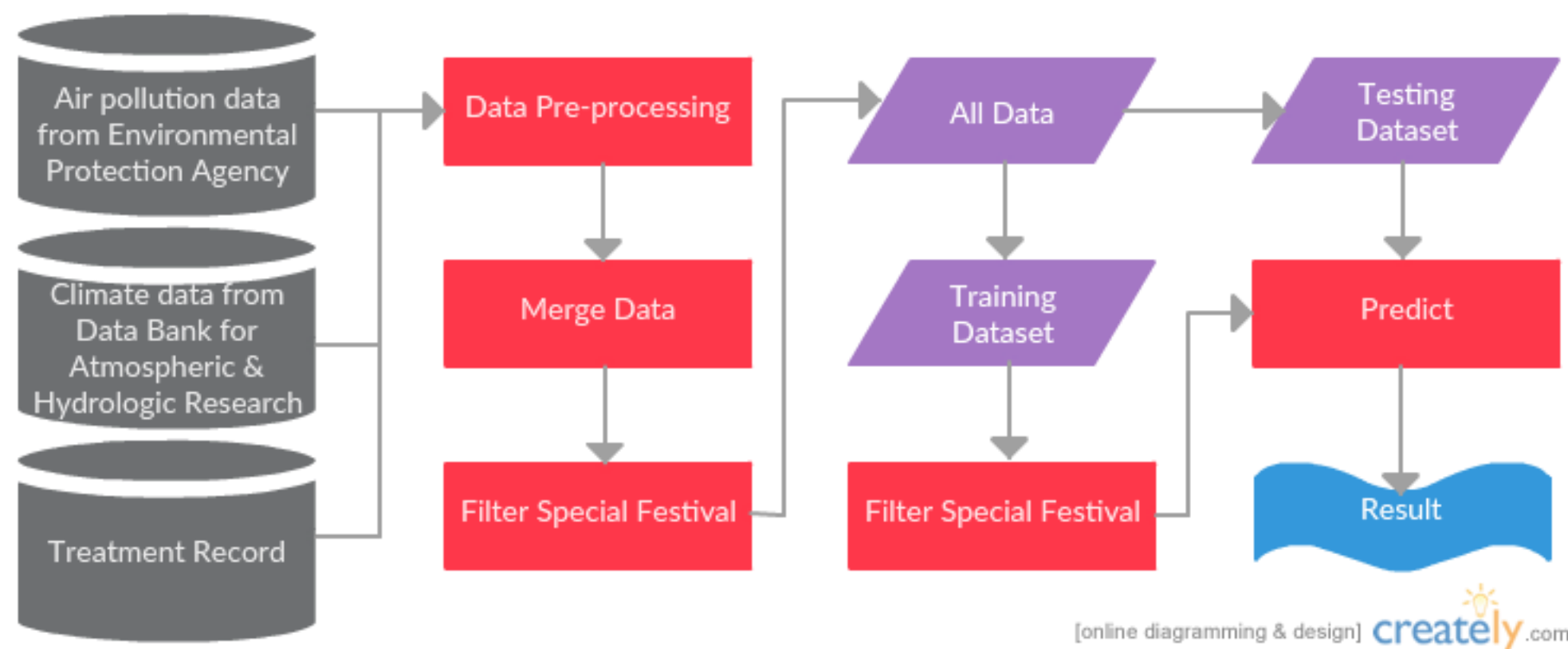


Fig1.workflow

## 2. 方法

- 資料處理：根據過往地研究顯示，天氣與空汙是會影響就診人數，因此我們先從環保局、大氣水文研究資料庫分別取得空汙資料以及天氣資料，再將醫院所提供的就診紀錄劃分為「環境因素」與「非環境因素」，借此濾掉非環境因素，再將環境因素資料與天氣及空汙資料依照日期做合併
  - 環境因素：由外在的環境因素而導致的看診人數，
  - 非環境因素：固定的就診人數，取決於各家醫院每個星期與每個時段之整年平均。
- 我們先將國曆各月份高屏地區就診人數統整分析得到圖2，發現在過年、連假以及颱風天的日子人數會特別低，因此我們在資料處理的部分有特別濾掉特殊節日、以及颱風天。（Total 為同一天每家醫院每個時段的總和）

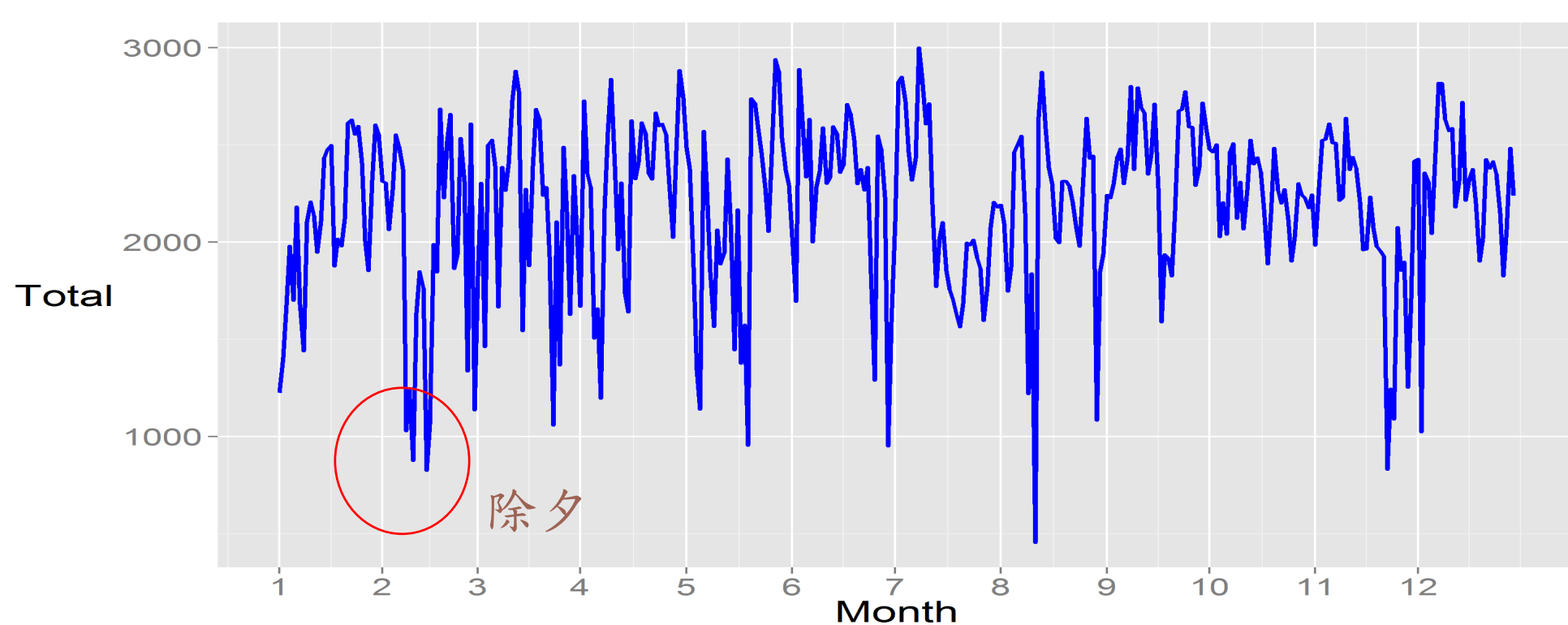


Fig2. 高屏地區醫院每日急診人數折線圖

- 我們將各星期各班次資料做整理成圖3，發現各星期及各班次人數起伏不定，所以我們將每個星期每個時段都獨立出來。
  - 觀察到禮拜六就診人數最少，禮拜一就診人數最多。
  - 凌晨四點到早上八點就診人數最低，中午十二點到下午四點就診人數最高。
- 因此我們將資料依據時段以及星期來區分。（Total 為同一天每家醫院的總和）

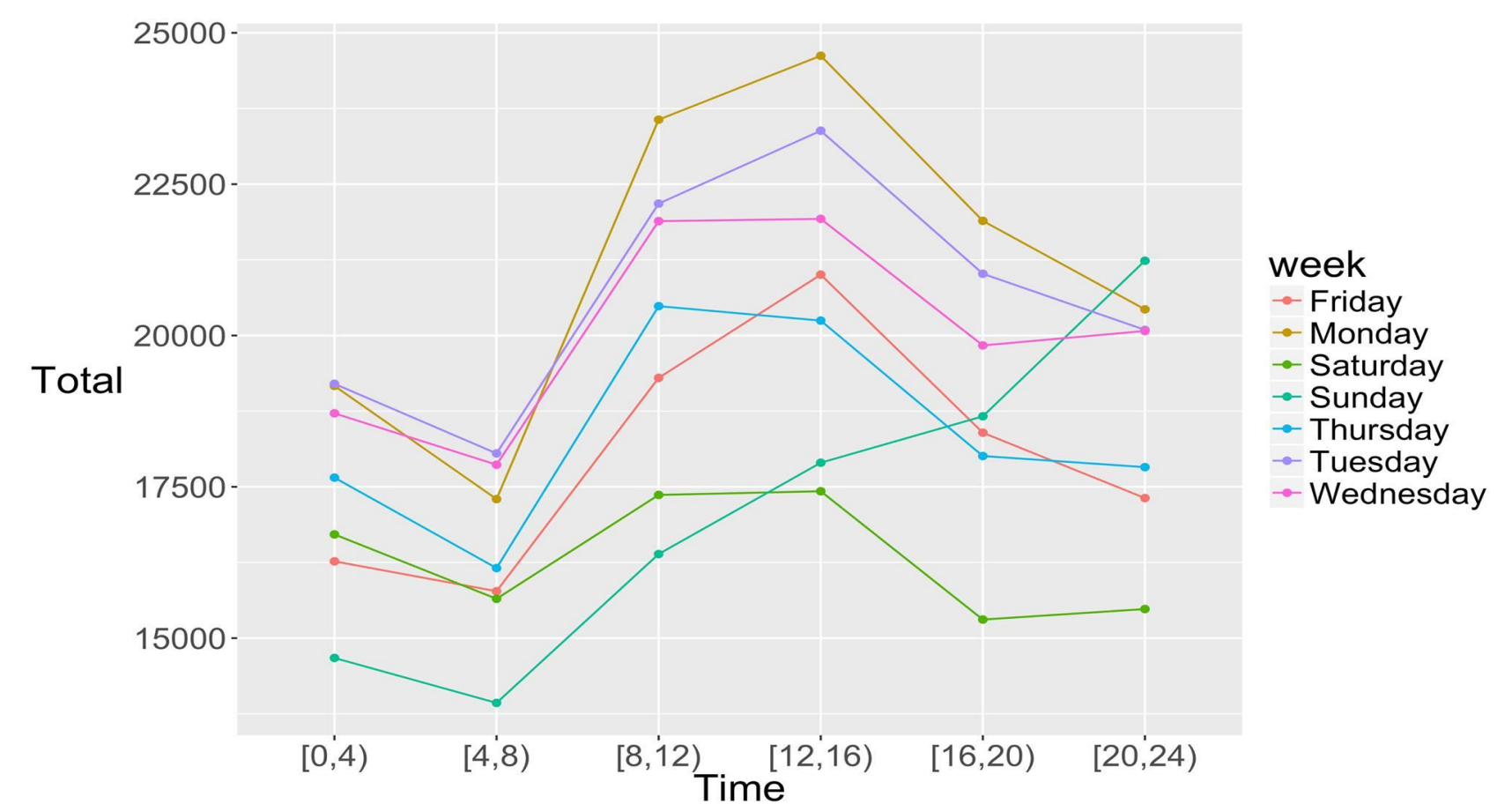


Fig3. 高屏地區醫院星期與時段之急診人數折線圖

- 將農曆各月份人數整理如圖4：我們觀察到農曆七月的時候人數最少，判定是台灣人對於鬼月的習俗問題而導致就診人數變低。因此我們在做分析時也將農曆七月的地方特別標記。（Total 為同一天每家醫院每個時段的總和）

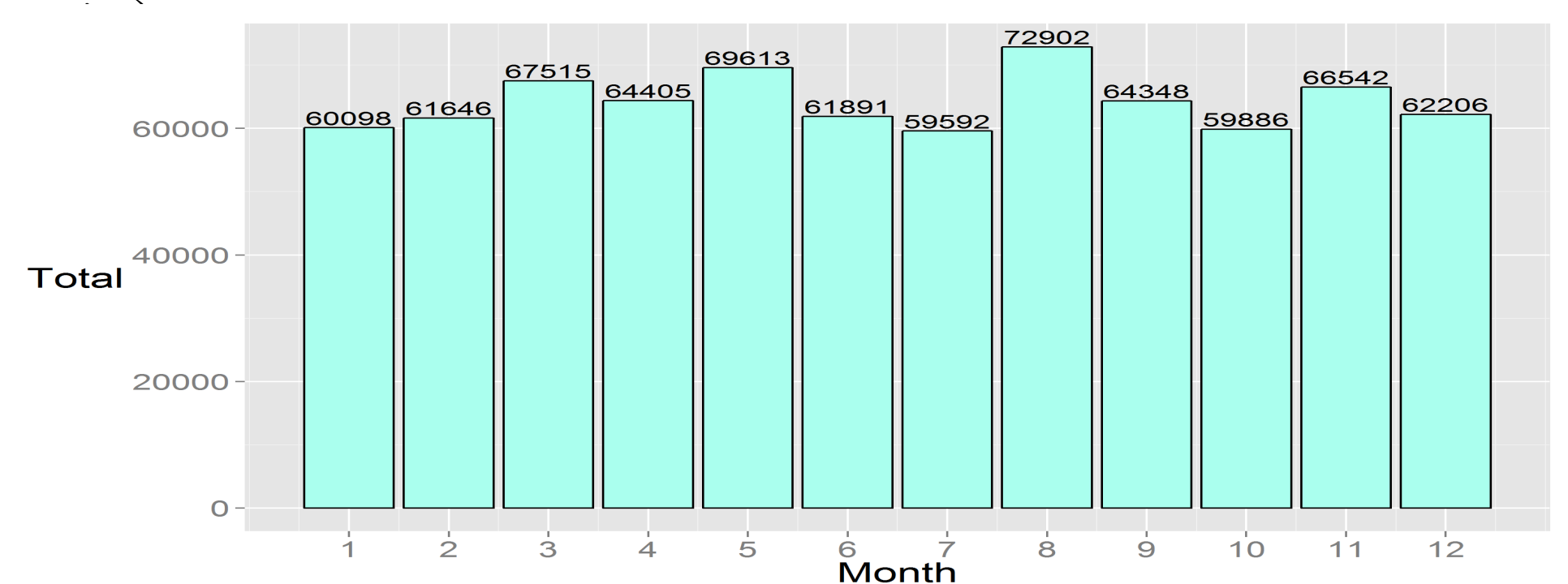


Fig4. 高屏地區醫院農曆月份與急診人數對應圖

- 模板建立：利用SVR（Support Vector Regression）工具建立Model找出人數與外在環境之間的關聯，將要預測的資料套到Model裡得到預測的結果
- 使用兩次SVR後預測出2014年第33週之就診的環境人數
  - Train\_1: 2013年整年但缺少8月的就診人數（減掉環境因素）、與氣候資料
  - Test\_1: 2013年八月的氣候資料
    - 利用SVM預測出2013年的人數，在做RMSE與實際住院人數比較，並且去找出最好的Model
    - 依據每家醫院、每個時段、每個星期找出最好的Model
  - Train\_2: 2013年整年但缺少8月的就診人數（減掉環境因素）、與氣候資料
  - Test\_2: 2014年第三十三週（8/11~8/17）的氣候資料
    - 利用SVR預測出2014年第33週的每家醫院、每個時段、每個星期的就診人數

## 3. 結果

- 根據做出來的結果我們可以透過空汙的六項指標：CO、SO<sub>2</sub>、PM<sub>10</sub>、PM<sub>2.5</sub>、NO、O<sub>3</sub>，和天氣的資料、以及醫院提供的資料來預測某時段該醫院的急診就診人數，也可得知環境因素與就診人數之趨勢有一定的關係。
- 藉由此研究，若未來透過與醫院合作，必定能夠大幅提升醫院之人資利用，以避免人資之浪費或缺失。
- 分數：97.231
- 評分標準：

$$C_M = \frac{\sum_{i=1}^S 2X_i Y_i}{\sum_{i=1}^S X_i^2 + \sum_{i=1}^S Y_i^2}$$

S表示類別總數

表示X群落(真實值)中，第i類別的個數(或相對次數)

表示Y群落(預測值)中，第i類別的個數(或相對次數)

## 4. Reference

- Effect of Seasonal and Monthly Variation in Weather and Air Pollution Factors on Stroke Incidence in Seoul, Korea
- <http://stroke.ahajournals.org/content/46/4/927.full.pdf>
- 台灣2013~2014年高屏地區空氣污染資料
- <http://taqm.epa.gov.tw/taqm/tw/YearlyDataDownload.aspx>
- 台灣2013~2014年高屏地區天氣資料
- <https://dbahr.narlabs.org.tw/>
- D C 資料挑戰賽
- <http://dc.dsp.im/index>