# Traffic-Accident-Data-Analysis交通事故與科技執法專題

專題介紹：台灣交通事故死傷人數逐年攀升，交通部近年大幅修繕交通設施，2022年12月1日擴大實施科技執法取締，希望透過大數據分析了解科技執法對提升交通安全是否有幫助，並預測事故熱區在特定條件下發生事故機率。

## 資料收集：

1.六都天氣觀測資料(每小時)

區間：2018年1月~2023年5月

來源：中央氣象局250個有效觀測站

數量：約 1100萬筆，皆為csv檔，資料欄位一致且完整 利用爬蟲取得，程式參考Python資料夾下weatherCrawl.ipynb

2.六都交通事故資料

區間：2018年1月~2023年5月

來源：政府資料開放平台(交通部/警政署)

數量：約 280萬筆，皆為csv檔 資料欄位一致且完整，主要欄位皆無缺失值，可下載取得

3.六都科技執法地點資料

區間：截至2023年5月前最新公告

來源：地方政府交通警察大隊官網

數量：約 3300筆

六都檔案格式、欄位皆未統一，六都公告更新頻率不一致，大部分缺少精確地址/經緯度，因此藉由程式處理1007筆經緯度、人工處理2278筆

若有精確地址，使用Selenium自動化程式結合Google Map擷取經緯度，請參考Python資料夾下addressTrans.ipynb

## 資料清整

第一階段：導入資料庫

匯入MySQL+QGIS確認經緯度

程式請參考Python資料夾下天氣日資料清整和匯入importWeatherToMySQL.py

第二階段：篩選樣本及補缺失值

程式請參考MySQL資料夾下DataClean.sql

第三階段：資料統整及分析

資料統整：

所有資料(事故地點、科技執法CameraID、天氣測站StationID)Join成大表

將距離事故地點經緯度最近的天氣測站加入dataframe,以及距離300公尺內最近的科技執法加入dataframe

套件使用geopy的geodesic計算經緯度距離，程式請參考PythoncalculateLongitudeLatitudeDistance.ipynb

資料分析：

分析科技值法實施前後之事故月平均值

程式請參考MySQL資料夾下CompareAccident\_BF&AF2022.sql

Matplolib套件畫圖程式參考參考Python folder底下trafficAnalysis.ipynb

## 機器學習：

結合天氣，事故及科技執法總合併資料訓練模型，透過機器學習，根據相關特徵值預測重大事故發生機率。

機器學習之流程：訓練資料欄位＞特徵種類及模型選擇＞編碼＆標準化＞特徵工程＞模型評估指標＞資料偏態調整＞模型最佳參數調校＞模型選擇＞檢視模型表現。

綜合評估模型表現及時效，選擇CatBoost，程式參考MachineLearning資料夾。

## 程式總整理：

資料收集與清整過程總共需應用到的程式

Python資料夾

1.天氣爬蟲程式 weatherCrawl.ipynb

2.天氣資料csv導入資料庫 importWeatherToMySQL

3.經緯度距離計算calculateLongitudeLatitudeDistance.ipynb

4.Matplotlib資料分析 accidentAnalysis.ipynb

5.將csv轉換為JSON檔程式 convertCSVtoJSON.py

6.將資料庫中資料寫出來轉成CSV ExportCSVfromMySQL.py

## 視覺化及網頁

1.主程式：

使用Python flask輕量化的網頁框架搭配前台JavaScript和Leaflet套件來實做的，執行project.py

示意圖參考mapAnalysis.gif

2.地圖分析：

因總筆數有到280多萬筆，若一次在地圖上顯示，會使Loading速度降低，因此資料拆為六都年月，在使用者選取城市，可以先看到科技執法地點，使用者再輸入年、月、事故類型後，才到後台取得年月Json資料，傳回前台，顯示在地圖上。

3.資料分析：

最終清整結果導入Tableau製作儀表板，示意圖參考TableauDashboard.gif

4.即時事故熱點預測：

收集2018-2023/5交通部道安資訊網所公布的六都事故熱點，收集經緯度座標，預測這些事故熱區發生事故的機率。 將產生的模型結合天氣API取得我們所要的天氣特徵，包含溫度濕度降雨量，合併成一個Dataframe後餵進去模型，產生y預測值後，轉換成Json格式，再傳回前台，畫在地圖上。 示意圖參考predictionHotSpot.gif