中華民國第 62 屆中小學科學展覽會

作品說明書

高級中等學校組　地球與行星科學科

佳作

051907

人工智慧進行汙染因子的模式及時序分析

學校名稱：國立臺南女子高級中學

作者：	指導老師：
高二 洪雅玲	洪士薰
高二 黃楷茵	游明珠

關鍵詞：人工智慧、時間序列分析、懸浮微粒

摘要

本研究蒐集松山、新竹、沙鹿、台南、前鎮、台東六個北中南東空品測站的空氣汙染和氣象因子資料，並用人工智慧：決策樹、邏輯回歸、支持向量機、K－平均演算法、ARIMA分析 PM2.5、PM10、O_3 的時空分布模式及汙染預測模型，研究顯示 PM2.5、PM10 污染好發特性相似，冬季汙染皆較嚴重，並與 CO、NOx、NO　及風速具相似時序節律；以汙染因子預測 PM2.5、PM10 皆有好的效果，其中 CO 為決定性因子，氣象因子的預測效果較汙染因子差，且各測站預測模式不盡相同。　O_3 汙染多為偶發時序節律不明顯，通常不伴隨 PM2.5、PM10 汙染，又因 O_3 源自 NO 光化反應，汙染及氣象因子預測皆比 PM2.5、PM10 好。

壹、前言

一、研究動機

隨著都市化與工業的進步，空氣汙染的議題日趨重要。工業日新月異雖帶來富裕與經濟的成長，但是也對環境帶來不少的汙染，常使台灣的天空蒙上灰濛濛的一片。無形的健康殺手「細懸浮微粒」充斥在空氣中，對人體的危害從頭到腳（Forum of International Respiratory Societies，2018）。可深入肺泡，並可能抵達細支氣管壁，干擾肺內的氣體交換，若長期暴露於懸浮微粒，可引發心血管、呼吸道、腦部疾病以及增加肺癌的危險（康健雜誌，2017）；此外，細懸浮微粒對環境也有極大的影響，如造成可見度降低、酸雨量高、生物多樣性減少等（葉惠中，2013）。

在現今科技中，仍無法精確對空氣汙染進行預測，而汙染來源的擴散因很難有效的標定所以也難以用動力學模型來預測懸浮微粒的時空變化。但若能夠根據各測站氣象因子與化學污染物觀測值建立懸浮微粒隨時序變化的模式並透過機器學習的方法建立各種氣象因子與化學污染物與 PM2.5、PM10 及 O_3 的關聯性，勢必能對健康以及汙染防制會有正面的影響。於是本研究從時間、地點等方面切入研究，蒐集北部、竹苗、中部、雲嘉南、高屏、花東不同空品區，共六個觀測站，且分別具不同人文產業及特色，台北－松山（航空業）、新竹－新竹（高科技）、台中－沙鹿（重工業）、臺南－臺南（研究者所身處的城市）、高雄－前鎮（重工業：工業測站）、台東－台東（多一、二級產業），針對六測站的空品資料進行分析，討論不同地區的汙染物時間序列、時空分布、及地理人文間彼此的關係。希望能對污染的預測變化等多一些以人工智慧的方式的探討。

二、研究目的

1. 將行政育院環保署資料各監測站之每小時資料統整並進行懸浮微粒污染程度分析，了解 PM2.5、PM10、O_3 含量的時空變化。
2. 以人工智慧方法 ARIMA、DT、LR、K－means、SVM 進行不同因子對汙染物質 PM2.5、PM10、O_3 的關係分析。
3. 針對明顯汙染影響因子進行再預測，找出 PM2.5、PM10 空氣汙染特徵的化學汙染物及天氣因子。
4. 以 K－means 聚類的方式分析各觀測站 PM2.5、PM10 及 O_3 的空氣污染型態，並比較測站彼此間的關聯性。

三、文獻回顧

（一）懸浮粒子

【1】【2】大氣中懸浮微粒的來源相當複雜，依來源不同可將大氣懸浮微粒區分為原生性氣膠（Primary Aerosol）與衍生性氣膠 （Secondary Aerosol）（參見圖 1－1）。原生性氣膠之定義為該成份主要由物理破碎或原生性污染排放直接產生，如海鹽飛沫、地殼元素、燃燒排放等。衍生性氣膠的來源則為原生性氣膠經由大氣化學反應作用所產生，非直接排放而得，其主要成份包含硫酸鹽、硝酸鹽、銨鹽與有機氣膠等。再依其粒徑大小可分為三類（參見圖 1－2）：

1. 粒徑大於 2 微米:由機械力量直接產生者，如燃燒、路面揚塵等，屬於原生性微粒。
2. 粒徑介 0.1～2 微米之間:由大氣中的氣體經化學反應轉變而成，經均勻核化及凝結成長所形成的液滴，主要包含氨氮硫酸鹽及氨氮硝酸鹽，屬於衍生性微粒。
3. 粒徑小於 0.1 微米:由熱蒸汽經冷凝成原生性微粒，再經凝聚及鏈結聚合所形成者，可藉由風，傳輸到遠距離的地方，更可以深入人體肺部。

衍生性微粒組成:

- 硝酸鹽：主要為氮氧化物（NO 與 NO_2 合稱為 NO_x，主要是燃料在高溫燃燒時空氣中過量的氮和氧反應產生）氧化所生成。
- 硫酸鹽：主要為硫氧化物（SO_2 與 SO_3 合稱為 SO_x，其主要來源為化石燃料的燃燒產生）。
- 銨鹽：銨鹽的前驅物為氨氣（NH_3），其來源主要為動植物殘體或排泄物以及其它微生物的活動所產生。

- 有機氣膠（Organic aerosol）：大氣有機氣體（reactive organic gas, ROG）經過光化作用產生半揮性有機物，當氣體傳輸緩慢時，有機氣膠含量會因此增加。

大多數離子在冬天的濃度較夏季高，但硫酸鹽與硝酸鹽的二次離子在夏天與冬天濃度都高。硫酸鹽與硝酸鹽的形成分別受到溫度與 NH_4^+ 的影響。溫度、相對濕度、降雨頻率、以及空氣密度可能是影響氣膠分布的主要因素,而燃煤、生物質燃燒、地殼、工業、交通排放為二次轉換為細氣膠的主要來源。

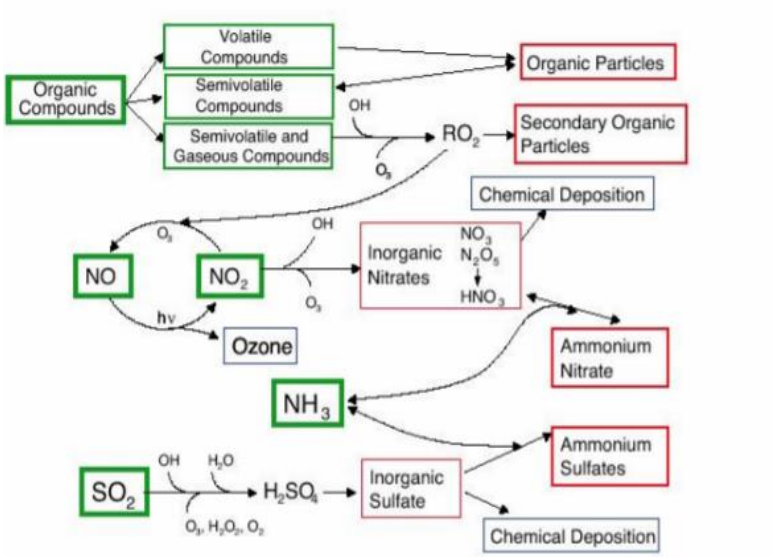


圖1－1 大氣氣膠形成路徑示意圖

（來源：Particulate Matter Science for Policy Makers－A NARSTO Assessment, 2003）

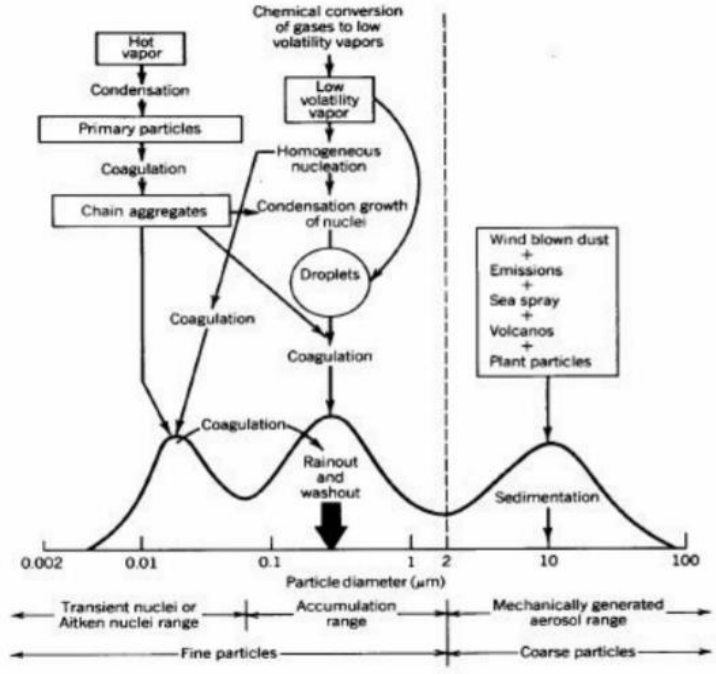


圖1－2 大氣懸浮微粒主要來源及粒徑分佈

（來源：行政院國家科學委員會補助專題研究計畫 2018）

（二）汙染物簡介【1】【3】【4】【5】

- 一氧化碳（CO）

  性質：無色、無味的氣體。

  來源：化石燃料或汽機車汽油燃燒不完全產生。

  對人體影響：容易與紅血球結合，造成缺氧，可能會引發：頭痛、噁心、嘔吐，嚴重甚至可能會有生命危險。

- 氮氧化物（NO_x）

  性質：一氧化氮（NO）無色、無味，微溶於水；二氧化氮（NO_2）是赤褐色氣體，具刺激味道，易溶於水形成酸雨。

  來源：自然情況，N_2在閃電高溫作用下，會與 O_2反應形成 NO（$N_2 + O_2 \rightarrow 2\ NO$），NO 進一步與 O_2反應後可形成 NO_2（$2\ NO + O_2 \rightarrow 2\ NO_2$）。另外，汽機車排氣、化石燃料燃燒過程中高溫造成氮或氮化物的氧化，也是氮氧化物大量形成的原因。

  對人體影響：短暫灼傷、痙攣、喉間組織及上呼吸道腫脹、體內組織缺氧、肺水腫，嚴重則會死亡。

- 二氧化硫（ SO_2）

  性質：具刺激臭味之無色氣體。溶於水後會形成具有腐蝕性的亞硫酸，也是引起酸雨的物質之一。

  來源：自然情況下，二氧化硫是火山爆發時噴出，人為來源則有燃燒石油及煤礦等途徑（$S_8 + 8\ O_2 \rightarrow 8\ SO_2$）。

  對人體影響：而若接觸到高濃度的 SO_2 可能會導致鼻子、喉嚨、肺部的灼傷，造成呼吸困難以及嚴重的呼吸道阻塞。

- 臭氧（O_3）

  性質：具強氧化性的溫室氣體，由氮氧化物（NO_x）與揮發性有機物（VOCs）二類前驅物共同經由大氣光化學反應所產生，NO_x 排放集中於市區及工業區，而 VOCs 排放除市區與工業區外，常溫下是有臭味的淡藍色氣體，微溶於水，更易溶於碳氟化合物或四氯化碳等極性溶劑，90% 的 O_3存於距地 15 到 50 公里之間的區域，可形成臭氧層幫助吸收紫外線。

  來源：自然情況下，O_2 經紫外光照射而分解為氧原子，而氧原子再與另一氧分子結合生成臭氧。

  $O_2 \rightarrow O + O$（紫外光催化）

$O_2 + O \rightarrow O_3$

人為活動下，臭氧並非直接排放，汽機車高溫引擎內氧氣和氮氣發生反應，產生 NO

$N_2 + O_2 \rightarrow 2\ NO$

NO 在空氣中再與 O_2 形成二氧化氮

$2\ NO + O_2 \rightarrow 2\ NO_2$

NO_2 在強光下產生分解產生 O

$NO_2 \rightarrow NO + O$

氧原子很容易與氧氣形成臭氧。

$O + O_2 \rightarrow O_3$

對人體影響：對呼吸系統具刺激性，能引起咳嗽、氣喘、頭痛及肺部傷害，若植物長期暴露在高濃度臭氧，則會使植物降低其生產力。

- 細懸浮微粒（PM2.5）

  性質：指粒徑在 2.5 微米以下的粒子，質量小，容易攜帶重金屬及微生物，能在大氣中停留，並可被風擴散到遠距離。成分包含有機碳、元素碳、硝酸鹽、硫酸鹽、銨鹽（多為 NO_X、SO_2、CO 等氣體經化學反應衍生而來）。受氣象因子、氣候狀況、地理位置、當地活動或遠距離傳輸影響，衍生性來源遠大於原生性，管制困難。

  來源分為:

  - 自然排放：火山爆發、地殼岩石、海鹽飛沫
  - 人為排放：燃燒、裸露揚塵、火力發電廠、石化廠、煉鋼廠、交通運輸工具的　廢氣排放、營建工程、不完全燃燒後所產生的重金屬、多環芳香烴，或是原生性汙染物經化學反應作用衍生而來。

  對人體影響：可穿透肺泡，進入血管隨著血液循環，引發哮喘、心血管病和支氣管炎等疾病。若是其再夾帶著重金屬或微生物，又會對人體造成更嚴重的影響。

- 懸浮微粒（PM10）

  性質：粒徑在 10 微米以下之粒子，包含 PM2.5 的含量。

  來源：道路揚塵、車輛排放廢氣、露天燃燒等，或是原生性汙染物經化學反應衍生而來。

  對人體影響：侵害呼吸系統誘發哮喘及心血管疾病，甚至會沉積在肺部。

（三）台灣位置與氣候

【1】臺灣位於西太平洋地區屬海島季風氣候，冬季受到西伯利亞大陸冷高壓影響，盛行東北風，夏季則是由印度低壓和太平洋高壓所影響，主要為西南季風。通過中國大陸之東北季風影響，造成 PM2.5 濃度上升，而夏季因為季風不經過亞洲大陸，空氣品質較不受影響（Chang et al.，2007）。依據台灣空氣品質監測站資料顯示，台灣地區最主要空氣污染物為懸浮微粒與臭氧，為台灣地區空氣品質惡化主要禍首，於懸浮微粒方面除宜蘭、花東地區外，其它空品區懸浮微粒監測值普遍超過法規標準值，尤以高屏地區最為嚴重。由歷年研究顯示，南部地區懸浮微粒污染現象，主要受細微粒（PM2.5）濃度影響鉅深（Lin,2002；廖2006）。

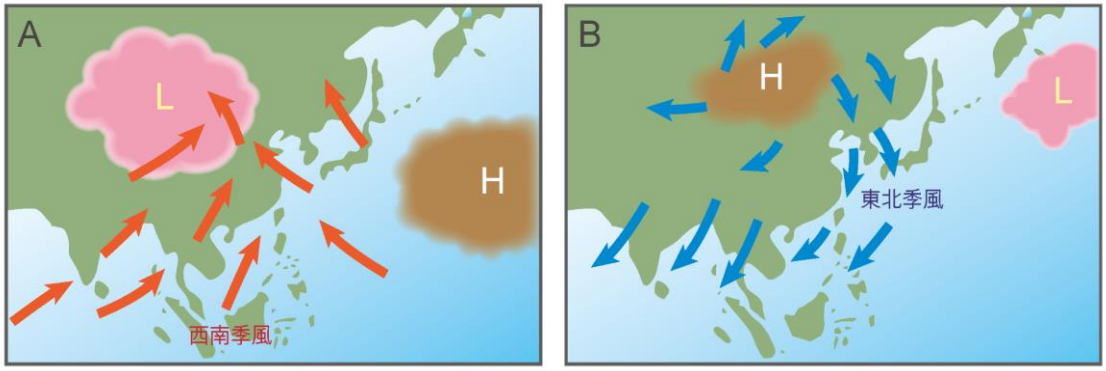


圖1－3 季風形成與路徑示意圖
（來源：中央氣象局數位科普網）

（四）研究選取測站說明

- 一般測站：松山（15）、新竹（24）、沙鹿（29）、臺南（46）、台東（62）
- 工業測站：前鎮（57）

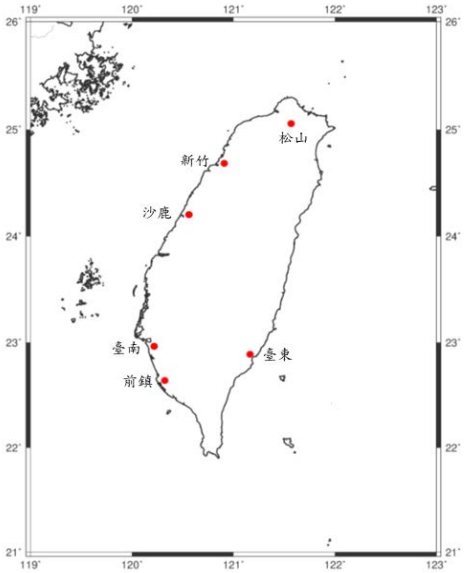


圖1－4　各測站在台灣分布位置

圖1－5－1松山測站周遭地形圖

圖1－5－2新竹測站周遭地形圖

圖1－5－3沙鹿測站周遭地形圖

圖1－5－4臺南測站周遭地形圖

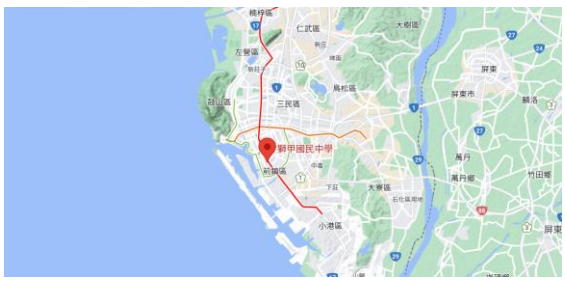


圖1－5－5前鎮測站周遭地形圖

圖1－5－6臺東測站周遭地形圖

表1－1各測站地形、氣候、產業活動說明（來源：維基百科）

測站（編號、詳細位置）	地形	氣候	產業活動
松山（15） 松山區松山國小	盆地	類型:副熱帶季風 溫度：盆地周圍受高山阻擋，不易使熱氣排出，市內的氣溫通常較周圍地區高。 降雨：山地與丘陵容易形成地形雨。	交通建設：松山國際機場
新竹（24） 東區東門國小	丘陵、平原	類型:副熱帶季風 風：四周地形高，中央沖積平原由東南向西北呈喇叭狀開啟類似「風口」，東北季風時，容易產生輻合效應，使風勢增強。	科技業：新竹科學園區 交通建設：新竹車站
沙鹿（29） 沙鹿區北勢國中	台地、平原	類型:副熱帶季風 降雨：集中在夏季，冬季少雨，東北季風盛行期間雨量較少，西南季風期間的雨量較豐沛。	重工業：台中火力發電廠、中龍鋼鐵、臺中國際機場及中科園區，鄰近中部科學園區

台南（46） 中西區中山國中	平原	類型:副熱帶與熱帶季風氣候過渡帶 日照：多 雨：全年少雨，西南季風盛行及對流作用，午後易生局部性對流雨，但太平洋高壓強時亦會出現長達整個月的夏旱。	觀光景點：古蹟及廟宇，如赤崁樓、孔子廟、台灣首廟天壇
前鎮（57） 前鎮區獅甲國中	平原	類型:熱帶季風氣候 溫度：終年高溫 降雨：夏季受西南季風影響，雨量多，冬季因東北季風被中央山脈阻擋，氣候乾燥。	交通建設：前鎮漁港、部分高雄港、高雄國際機場與旗津渡輪站 工業園區：高雄加工出口區、高雄軟體園區、中鋼集團
臺東（62） 台東市台東縣政府	平原	類型:熱帶季風氣候 降雨：受溫暖的海洋氣流調節，全年雨量充沛， 溫度：而在冬季南下的東北季風因受到中央山脈和海岸山脈阻隔，年溫差小，氣候溫和。	農、漁業和服務業主 交通建設：台東車站、富岡漁港、豐年機場

貳、研究設備及器材

一、電腦數部含網路

二、應用軟體

（一） Excel：數據演算、趨勢變化比較、繪製圖表

（二） Rapidminer：數據演算、人工智慧

（三） Word：文字記錄、報告彙整

三、觀測資料來源：

行政院環境保護署空氣品質監網

參、研究過程或方法

因為 PM2.5、PM10 及O_3的生成來源各有其因子，有部份易受天氣因素影響，部份則不易。加上雖有觀測資料卻不易追蹤，分析上很難建立完整的動力學模型，但若只從簡單的統計量及迴歸直線又很難有足夠有力的分析。所以本文從人工智慧的機器學習數學方法來分析。以下是我們的流程圖:

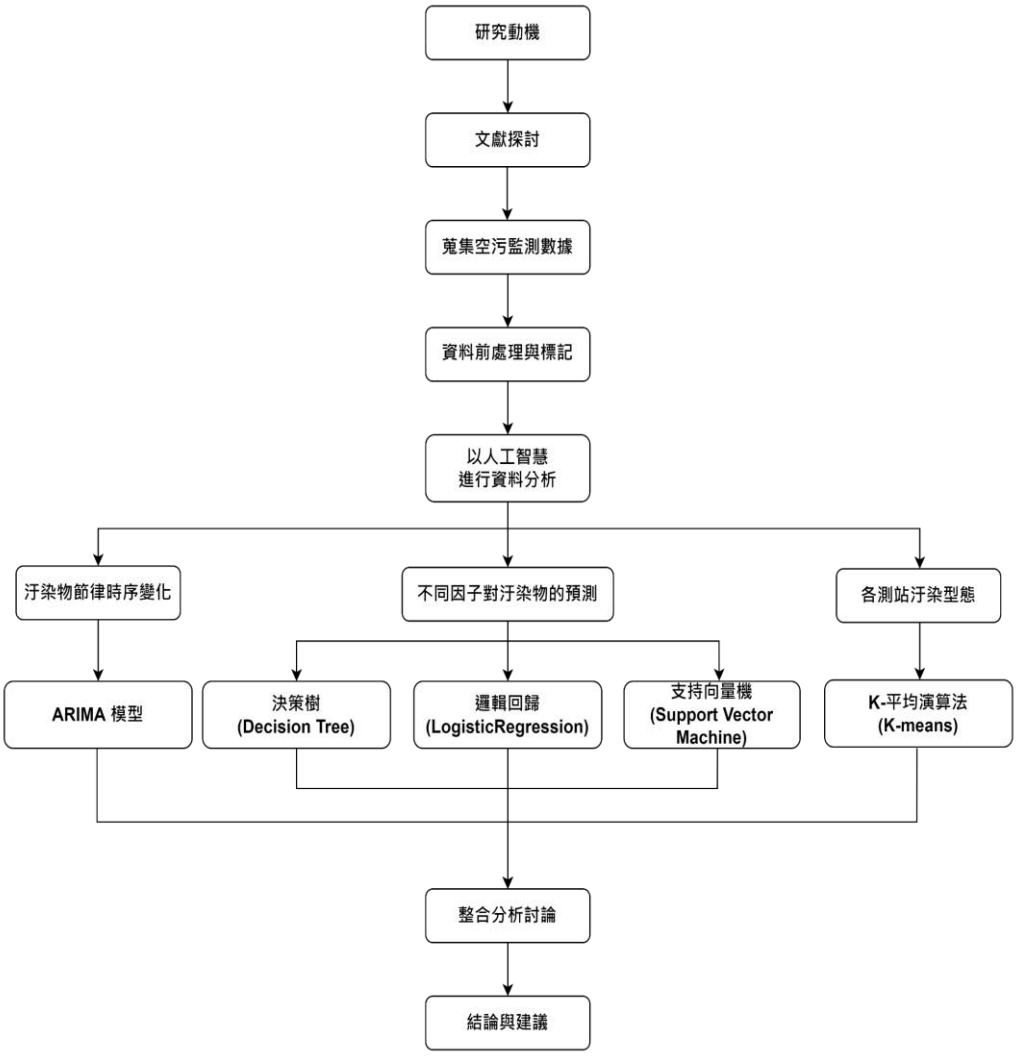


圖3-1 研究流程圖

1.資料前處理與標記：

（1）各測站原始資料為 CO、NO、 NO_2、Nx、SO_2、PM2.5、PM10、O_3 及風向、風速、相對濕度、溫度每一小時的觀測值，因考量預測以日為單位難以看出汙染因子與氣象因子隨時間變化的影響，而以小時為單位太細易受干擾，因此先進行 6 小時為一週期，每日分成 3～9、9～15、15～21、21～3 四個時段，對每一時段的最大、最小及平均值整理。

表3－1－1 空汙資料與氣象因子單位列表

觀測項目	PM2.5	PM10	O_3	CO	NO	NO2	Nx	SO2	風向	風速	溫度	相對濕度
單位	$\mu g/m^3$	$\mu g/m^3$	ppb	ppm	ppb	ppb	ppb	ppb	°	m/s	˚C	%

（2）對於 PM2.5、PM10 及 O_3 資料高於平均兩個標準差標記為污染，此標記範圍介於環　境保護署的空氣品質汙染指標對敏感族群不健康及對全部族群不健康。

表3－1－2 空氣品質指標（AQI）（來源：行政院環保署空氣品質汙染指標）

空氣品質指標 (AQI)							
AQI 指標	O_3 (ppm) 8 小時平均值	O_3 (ppm) 小時平均值 (1)	$PM_{2.5}$ (μg/m³) 24 小時平均值	PM_{10} (μg/m³) 24 小時平均值	CO (ppm) 8 小時平均值	SO_2 (ppb) 小時平均值	NO_2 (ppb) 小時平均值
良好 0～50	0.000 - 0.054	-	0.0 - 15.4	0-50	0 - 4.4	0-20	0-30
普通 51～100	0.055 - 0.070	-	15.5 - 35.4	51-100	4.5 - 9.4	21-75	31-100
對敏感族群不健康 101～150	0.071 - 0.085	0.125 - 0.164	35.5 - 54.4	101-254	9.5 - 12.4	76-185	101-360
對所有族群不健康 151～200	0.086 - 0.105	0.165 - 0.204	54.5 - 150.4	255-354	12.5 - 15.4	186-304 (3)	361-649
非常不健康 201～300	0.106 - 0.200	0.205 - 0.404	150.5 - 250.4	355-424	15.5 - 30.4	305-604 (3)	650-1249
危害 301～400	(2)	0.405 - 0.504	250.5 - 350.4	425 - 504	30.5 - 40.4	605-804 (3)	1250-1649
危害 401～500	(2)	0.505 - 0.604	350.5 - 500.4	505-604	40.5 - 50.4	805-1004 (3)	1650-2049

2. 透過決策樹、邏輯迴歸、SVM 預測：

分別將資料分成以下列方式處理分析:

（1）以空氣污染因子: CO、NO、NO_2、Nx、SO_2 的六小時均量及六小時最大值進行預測分類。

（2）以天氣因子：風向、風速、相對濕度、溫度的六小時均量及六小時最大值進行預測。並計算前述（1）方式所得的所得的六小時均量及六小時最大值，大於 2 個標準差者標計為污染時段（以 6 小時為單位）

以（1）、（2）預測統計，並比較此兩種方法的差異。

表 3－1－2 與表 3－2－1中對於 PM2.5、PM10 及 O_3 三項空品資料，大於平均兩個標準差的值也正好介於行政院環保署空氣品質汙染指標對敏感族群不健康與對全部族群不健康。

決策樹、邏輯迴歸、SVM 的預測方法：進行運算時，因高於平均兩個標準差才標記為污染，導致原始資料比例上，空氣汙染時段數據量與空氣品質良好時段數據量比例太過懸殊，將導致人工智慧的方法上會過度偏重依賴空氣品質良好時段的數據做處理以獲取其特徵，所以，我們在進行運算上，將空氣品質良好時段進行隨機抽樣數據，使得樣本分布呈現：「空氣汙染時段資料數量」：「空氣品質良好資料數量」約略為 1：1.5 的方式取樣，再進行人工智慧運算。

以前鎮觀測站（57）2016 年 1 月 1 日至 2021 年 6 月資料為例，以兩種取樣，第一種以全部資料進行人工智慧運算結果如表 3－2－2（a）。第二種則從空氣品質良好時段隨機抽樣 15% 資料加上全部的空氣汙染時段，來進行人工智慧運算結果如表 3—2—2（b）。兩者皆有 83% 左右的準確率，但第一種方法因資料比重的原因，演算時太偏重空氣品質良好時段特徵。所以判斷空氣汙染時段的精確率及召回率均偏低。但經過方法二適當取樣後，再進行人

工智慧運算的結果則不會有太偏重於空氣品質良好天數資料更平衡了只取空氣汙染天數資料的偏差的偏差。

表3－2－1 PM2.5、PM10 及 O_3 空氣品質平均值及標準差（2016 年 1 月 1 日至 2021年 6 月 30 日)

測站	沙鹿	前鎮	松山	新竹	臺南	臺東
汙染源	PM2.5					
平均	22.782	30.173	20.419	20.685	29.636	11.015
標準差	15.161	19.587	11.26	12.661	18.433	6.563
最大值	137	445	134	104	140	147
2個標準差	53.104	69.347	42.939	46.007	66.502	24.141
3個標準差	68.265	88.934	54.199	58.668	84.935	30.704
汙染源	PM10					
平均	44.877	60.969	38.249	41.242	58.173	29.285
標準差	26.264	34.948	21.37	22.835	35.652	24.068
最大值	253	489	249	460	396	1000
2個標準差	97.405	130.87	80.989	86.912	129.48	77.421
3個標準差	123.669	165.81	102.36	109.75	165.13	101.49
汙染源	O3					
平均	40.52	37.635	36.85	39.161	41.441	30.529
標準差	18.495	22.22	20.118	17.668	21.399	12.017
最大值	131	135	168	142	143	92
2個標準差	77.51	82.075	77.086	74.497	84.239	54.563
3個標準差	96.005	104.3	97.204	92.165	105.64	66.58

表3－2－2（a） 全部資料進行 Decision Tree 運算結果

accuracy: 83.60%

	true false	true true	class precision
pred. false	1570	308	83.60%
pred. true	0	0	0.00%
class recall	100.00%	0.00%	

\

表3－2－2（b）空氣品質良好時段隨機抽樣 15% 資料加上全部的空氣汙染時段進行 Decision Tree 運算結果

accuracy: 83.45%

	true false	true true	class precision
pred. false	258	19	93.14%
pred. true	51	95	65.07%
class recall	83.50%	83.33%	

（a）決策樹：是以資訊分配準確分類的熵與否來進行演算，以最大資訊增益比。

在樣本目標中根據前面標記的有空氣污染與空氣品質良好來進行樹形結構的分類，以最大資訊增益方式演算 全體訓練樣本集合為 D，其空氣污染分類的熵為

$H(D) = -(p_0 log_2 p_0 + p_1 log_2 p_1)$

樣本特徵值（x, v）為樣本 x 在指定特徵（例如2（1）或2（2））的值為 v。

對於特徵 A 選定值 a，可以將所有樣本分成特徵值大於 a 的集合 S a 及其他集合 S a~

$$H(D, A) = \frac{|S_a|}{|D|} \cdot H(S_a) + \frac{|\widetilde{S}_a|}{|D|} \cdot H(\widetilde{S}_a)$$

增益值公式為

$$g(D, a) = H(D) - H(D, A)$$

另考慮特徵 A 選定值 a 時的熵為

$$H_A(a) = -\frac{|S_a|}{|D|} \cdot log_2\left(\frac{|S_a|}{|D|}\right) - \frac{|\widetilde{S}_a|}{|D|} \cdot log_2\left(\frac{|\widetilde{S}_a|}{|D|}\right)$$

特徵A 值 a 時的資訊增益比為

$$G(D, A, a) = \frac{g(D, a)}{H_A(a)}$$

計算對於每個特徵由起使值開始進行梯度向量的疊代計算最小資訊增益比時的 a 值作為選擇分枝的判準以生成決策樹。

以 RapidMiner Studio 9.10.001 版本 執行運算結果如下圖

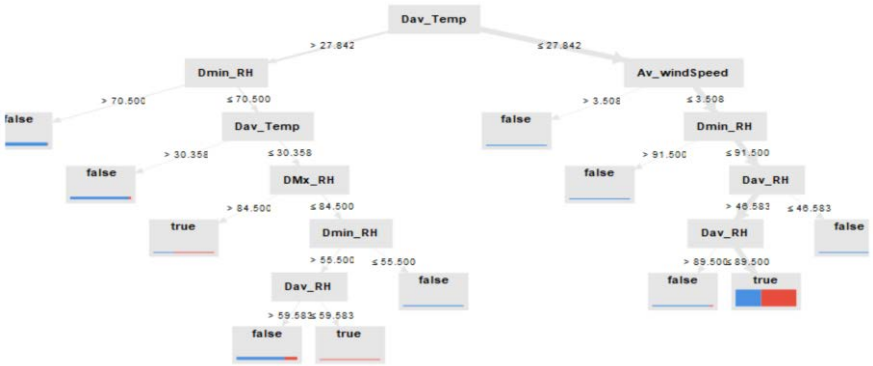


圖3－2－3 決策樹的資訊分類 (前鎮觀測站 2016 年 1 月 1 日至 2021 年 6 月資料空氣品質良好時段隨機抽樣 15% 資料加上全部的空氣汙染時段進行 Decision Tree 運算結果)

（b）邏輯迴歸

對於變量$X = （x_1, x_2, x_3, \cdots, x_n）$，找到權數$w_0, w_1, w_2, \cdots, w_n$，令

$g(X) = w_0 + w_1x_1 + w_2x_2 + \cdots + w_nx_n$，對於標記的事件，例如汙染的發生以應變量$y = 1$表示。則條件機率設為$P(X) = \frac{1}{1+e^{-g(x)}}$。 若$p = P(X)$，$\frac{p}{1-p}$稱為發生比。

求得使 $ln\ ln\ \frac{p}{1-p}$ 最大的權數$w_0, w_1, w_2, \cdots, w_n$即是。所以是一種多變量的二元分類模型。利用各相關因子找出標記的事件發生與否的二元分類。例如以松山觀測站PM2.5執行程式的結果如下：

Logistic Regression Model　Model Metrics Type: BinomialGLM
MSE: 0.112079225　RMSE: 0.33478236　R^2: 0.50669324　AUC: 0.91585314
pr_auc: 0.8558165　log loss: 0.34975612　mean_per_class_error: 0.16345733
default threshold: 0.40648317337036133
CM: Confusion Matrix (Row labels: Actual class; Column labels: Predicted class):

	false	true	Error	Rate
false	399	58	0.1269	58 / 457
true	49	196	0.2000	49 / 245
Totals	448	254	0.1524	107 / 702

Gains/Lift Table (Avg response rate: 34.90 %, avg score: 40.17 %)：

Attribute	Coefficient	Std. Coefficient	Std. Error	z-Value	p-Value
Dav_SO2	0.176	0.531	0.050	3.490	0.000
Dav_CO	12.875	3.438	1.274	10.105	0
Dav_O3	0.068	1.255	0.019	3.621	0.000
DMx_O3	0.005	0.116	0.011	0.405	0.685
Dav_NO2	-0.149	-1.537	0.090	-1.663	0.096
DMx_NO2	0.014	0.180	0.046	0.306	0.760
Dav_NO	-0.156	-1.532	0.112	-1.393	0.164
DMx_NO	-0.021	-0.331	0.055	-0.381	0.704
Dav_Nx	0.060	1.072	0.084	0.713	0.476
DMx_Nx	0.021	0.497	0.042	0.486	0.627
DMx_RH	-0.003	-0.035	0.013	-0.262	0.794
Intercept	-8.485	-1.096	1.316	-6.447	0.000

圖3-2-4 邏輯迴歸結果

（c）SVM：

採用一般的係數乘積。訓練樣本為(x_i , y_i)， $1 \le i \le N$， x_i 是每個樣本的資料向量，y_i 是汙染與否為 0 或 1$w^T \cdot x$表示 w 與 x_i 內積，其中 w 是待決定的權重， ξi 是每一樣本的容許誤差成本。

$minimize \frac{1}{2}||w||^2 + C\sum_i \quad \xi i$

$subtect\ to\ y_i(w^T x_i$-b)-1+$\xi i \ge 0 \forall i$

$\xi i \ge 0 \forall i$

以梯度疊代的方式計算。

決策樹（Decision Tree）、邏輯回歸（Logistic Regression）及支持向量機（SVM）分析三種方法的預測，我們較著眼於天氣汙染與良好的特徵是什麼? 但因原始資料中空氣品質良好資料遠較空氣汙染多，所以演算結果空氣品質良好的精確率及召回率都比較偏高。這意味著，若只以原始資料進行人工智慧的預測，則所得結果會較偏重於空氣品質良好的特徵，但我們更關心的是空氣汙染的特徵是什麼？因此，我們比較人工智慧運算的結果時採取空氣汙染的F1－score 值（亦被稱做F-measure） F1－score 定義為 $\frac{2}{\frac{1}{p}+\frac{1}{r}}$ 。

這是一種量測方法的精確度常用的指標，經常用來判斷演算法的精確度。目前在辨識、偵測相關的演算法中經常會分別提到精確率 p（precision）和召回率 r（recall），F1－score 能同時考慮這兩個數值，平衡的反映這個演算法的精確度。

表3－2－5 六觀測站 PM2.5 以空氣污染物 CO,NO, NO_2, Nx, SO_2 的六小時均量及六小時最大值進行預測結果之混淆矩陣（縱軸：N 為預測為陰性、P 為預測為陽性；橫軸： F 為真實為假、T 為真實為真）

performance	PM2.5											
Decision Tree	測站	前鎮		沙鹿		松山		新竹		臺南		臺東	
		F	T	F	T	F	T	F	T	F	T	F	T
	N	133	11	211	7	130	38	140	12	186	12	142	40
	P	84	99	95	144	41	125	62	133	60	133	26	62
	精確率	0.541		0.6025		0.753		0.6821		0.6891		0.7045	
	召回率	0.9		0.9536		0.7669		0.9172		0.9172		0.6078	
	準確率	0.7095		0.7768		0.7635		0.7867		0.8159		0.7556	
	F1 SCORE	0.6758		0.7385		0.7599		0.7824		0.787		0.6526	

performance	PM2.5											
Logistec Regration	測站	前鎮		沙鹿		松山		新竹		臺南		臺東	
		F	T	F	T	F	T	F	T	F	T	F	T
	N	169	36	268	47	283	30	276	41	228	40	167	29
	P	25	74	38	104	22	133	31	194	34	105	27	73
	精確率	0.7475		0.7324		0.8581		0.8622		0.7554		0.73	
	召回率	0.6727		0.6887		0.816		0.8255		0.7241		0.7157	
	準確率	0.7993		0.814		0.8889		0.8672		0.8182		0.8108	
	F1 SCORE	0.7081		0.7099		0.8365		0.8435		0.7394		0.7228	

performance	PM2.5											
svm	測站	前鎮		沙鹿		松山		新竹		臺南		臺東	
		F	T	F	T	F	T	F	T	F	T	F	T
	N	193	35	250	47	182	33	247	51	207	25	168	32
	P	31	75	25	104	31	130	29	94	43	120	26	70
	精確率	0.7075		0.8062		0.8075		0.7642		0.7362		0.7292	
	召回率	0.6818		0.6887		0.7975		0.6483		0.8276		0.6863	
	準確率	0.8024		0.831		0.8298		0.81		0.8278		0.8041	
	F1 SCORE	0.6944		0.7429		0.8025		0.7015		0.7792		0.7071	

在決策樹、邏輯迴歸、SVM 人工智慧模型預測中，選擇預測結果良好的模型（AUC 值在 0.88 以上）分析空氣汙染特徵的前驅汙染物及天氣 p 值小於 0.025 的前三個因子對於 PM2.5、PM10的預測 。

表3－2－6 邏輯迴歸中松山測站 PM2.5 以前驅污染物的預測結果

Attribute	Coefficient	Std. Coefficient	Std. Error	z-Value	p-Value
Dav_SO2	0.284	0.859	0.057	4.973	0.000
Dav_CO	15.659	4.276	1.537	10.185	0
Dav_O3	0.066	1.247	0.019	3.532	0.000
DMx_O3	-0.004	-0.109	0.012	-0.336	0.737
Dav_NO2	-0.382	-3.981	0.099	-3.860	0.000
DMx_NO2	0.124	1.570	0.043	2.844	0.004
Dav_NO	-0.176	-1.570	0.116	-1.518	0.129
DMx_NO	-0.029	-0.425	0.048	-0.603	0.547
Dav_Nx	0.242	4.237	0.093	2.604	0.009
DMx_Nx	-0.080	-1.899	0.036	-2.183	0.029
Intercept	-9.027	-0.626	0.747	-12.092	0

表3－2－7 以前驅污染物的決定因子進行對松山測站 PM2.5再預測的結果

performance Decision Tree	PM2.5 測站	松山 avCO		avSO2		avO3	
		F	T	F	T	F	T
	N	228	61	236	81	263	128
	P	46	103	38	83	11	36
	精確率	0.691275		0.685950		0.765957	
	召回率	0.628048		0.506097		0.219512	
	準確率	0.755707		0.728310		0.682648	
	F1 SCORE	0.658146		0.582456		0.341232	

performance Logistec Regration	PM2.5 測站	松山	avCO	avSO2		avO3	
		F	T	F	T	F	T
	N	238	72	241	109	255	122
	P	36	92	33	55	19	42
	精確率	0.71875		0.625		0.688524	
	召回率	0.560975		0.335365		0.256097	
	準確率	0.753424		0.675799		0.678082	
	F1 SCORE	0.630136		0.436507		0.373333	

performance svm	PM2.5 測站	松山	avCO	avSO2		avO3	
		F	T	F	T	F	T
	N	235	55	230	112	274	164
	P	38	109	14	52	0	0
	精確率	0.741496		0.787878		#DIV/0!	
	召回率	0.664634		0.317073		0	
	準確率	0.787185		0.691176		0.625570	
	F1 SCORE	0.700964		0.452173		#DIV/0!	

3. K－means分類：

PM2.5、PM10 及 O_3 的 6 小時數值進行分類，選定分類效益最大時的 K 值，再依 K 類中心與環境保護署的空氣品質不良及嚴重汙染的標準值，進行分組。

K－means 是一種聚類（Cluster）的方式，原始每一筆資料的經緯度分別為 x、y 坐標，記為 $\boldsymbol{x_j}$，一開始選定 K 個初始值資料 $\mu_1, \mu_2, \cdots, \mu_k$，計算

$J=\| \boldsymbol{x_j} - \mu_i \|^2$，i=1,2,⋯,k，若 J 值最小，則 $\boldsymbol{x_j} \in S_t$，如此原始每一筆資料可分成 $S_1, S_2, \cdots, S_k$類， 再計算 $J=\sum_{i=1}^{k}\sum_{x_j \in S_i}\|x_j-\mu_i\|^2$

的值，對初始值資料 $\mu_1, \mu_2, \cdots, \mu_k$ 分別做變動並進行疊代求 J 的最小值，此時所得的 $\mu_1^*, \mu_2^*, \cdots, \mu_K^*$，即是 K 個聚類中心。再將原來資料依與聚類中心的距離歸類，即可將資料分類。

平均中心距離= $\sqrt{\frac{1}{n}\sum_{i=1}^{k}\sum_{x_j \in S_i}\|x_j-\mu_i^*\|^2}$ ，其中，K 越大，K－means 的平均中心距離當然會越小，但當當 K 達一個值後，平均中心距離下降的情形會趨緩。當然 K 越大，K－means 的平均中心距離當然會越小，當 K 達一個值後，平均中心距離下降的情形會趨緩。以下圖(圖3－3－1)為例，當 K＝12 時，平均中心距離斜率下降趨緩，分類效果已較難改進，所以就選

定 K＝12，將空氣品質先分成 12 類，再依 PM2.5、PM10 及 O_3 資料高於平均兩個或 3 個標準差標記為污染及高度污染將 12 類分成若干組。

表 3－3－2，選定 12 個聚類中心，再根據高於平均兩個標準差標對此 12 個聚類中心分成 4組，如下表 3－3－4。單純從 3 項不同的空品指標，分成良好與不良應會有 8 種可能，但就 K－means 的結果來看只有四大類組，可見 PM2.5、PM10 及 O_3 這三項汙染並非是獨立的，其中有一定的關聯性存在。

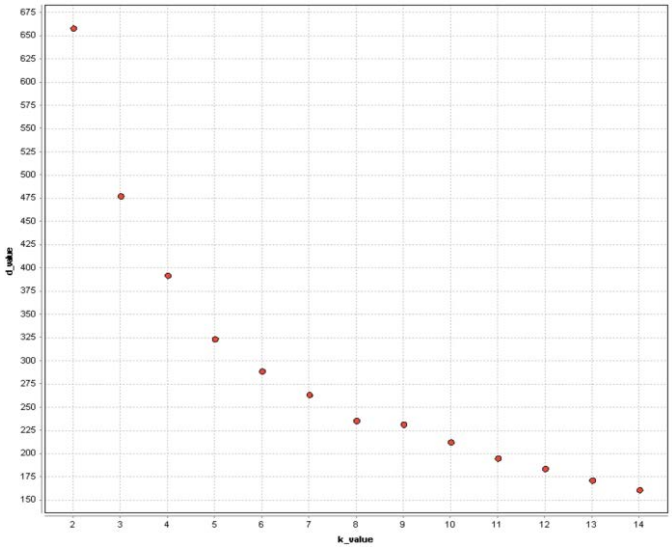


圖3－3－1 K－means K值（x 軸） 與平均中心距離（y 軸）

表3－3－2 Cluster Model （57 K=12）

前鎮												
分類	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4	cluster_5	cluster_6	cluster_7	cluster_8	cluster_9	cluster_10	cluster_11
PM10	93.74	71.41	109.78	122.50	38.16	160.00	281.64	65.04	45.51	61.07	22.25	91.05
O3	52.52	73.53	77.68	32.31	27.19	47.83	55.94	43.70	58.72	18.03	17.57	21.19
PM2.5	46.39	37.40	60.05	71.83	17.12	43.50	71.50	34.90	24.10	31.31	8.47	47.54

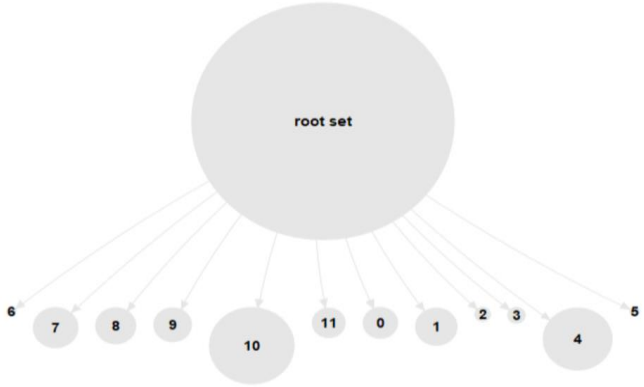


圖3－3－2 Cluster Model 圓形圖示

表 3－3－4 Cluster Model再分組

組別	PM10	O3	PM2.5	包含的聚類中心
1	良好	良好	良好	cluster_0,1,4,7,8,9,10,11
2	不良	良好	良好	cluster_5
3	不良	良好	不良	cluster_3,6
4	不良	不良	不良	cluster_2

4. ARIMA 模型（Autoregressive Integrated Moving Average model），差分整合移動平均自迴歸模型，又稱整合移動平均自回歸模型，時間序列預測分析方法之一。 ARIMA（p，d， q）中，AR 是"自回歸"，p 為自回歸項數；MA 為"移動平均"，q 為滑動平均項數，d 為使之成為平穩序列所做的差分階數。

ARIMA（p，d，q）模型是 ARMA（p，q）模型的擴展。 ARIMA（p，d，q）模型可以表示為：

$$（1 - \sum_{i=1}^{p} \phi_i L^i）（1 - L）^d X_t = （1 + \sum_{i=1}^{q} \theta_i L^i）\varepsilon_t$$

其中 L 是遲滯算子。即 $L^k x_t = x_{t-k}$

或者也可以整理成下列形式

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

μ 是 X_t 平均數，ϕ_i 是自回歸參數，θ_i 是移動平均參數，ε_t 是誤差量。一般至少要 有 50 個歷史資料才能有好的預測。

肆、研究結果

一、基本統計：

根據表 3－1－2 行政院環保署空氣品質汙染指標，以 PM2.5、PM10、O_3 三種污染源對敏感族群不健康的的標準（PM2.5 35.5 ug／M^3, PM10 101 ug／M^3, O_3 0.071 ppm ）。計算六個觀測站，逐月統計自 2016 年 1 月 1 日至 2020 年 12 月 31 日以 6 小時為單位的超出標準的污染次數。結果如表 4－1－1。數據呈現:

（1）PM2.5 汙染多發生在 1～4 月；PM10 汙染多發生在 1～5 月； O_3 多發生在 3～5 月及 9～10 月。

（2）PM2.5 汙染時段數＞PM10 汙染時段數＞O_3 汙染時段數。

（3）對各測站而言PM2.5汙染時段數：前鎮＞台南＞沙鹿＞新竹＞松山，台東則幾乎沒有；PM10汙染時段數：前鎮＞沙鹿＞新竹＞松山，台南及臺東較少；O_3 汙染時段數：台南＞前鎮＞沙鹿，新竹、松山及台東則很少。

（4）2020 年汙染時段數明顯變少，應該是受 COVID－19 疫情影響，工業及經濟活動減少所致。

表 4－1－1 觀測站前鎮、台南、沙鹿、新竹、松山、台東依年、月統計汙染時段數量

PM2.5	35.5											
15	1	2	3	4	5	6	7	8	9	10	11	12
2016	13	14	16	34	9	1	3	3	2	7	4	8
2017	9	13	18	28	6	3	1	1	7	6	3	6
2018	1	11	10	5	4	0	1	0	2	1	3	1
2019	0	4	11	5	1	0	0	0	0	5	0	0
2020	7	3	6	1	0	0	0	0	0	0	1	2
total	30	45	61	73	20	4	5	4	11	19	11	17
百分比	10.0%	15.0%	20.3%	24.3%	6.7%	1.3%	1.7%	1.3%	3.7%	6.3%	3.7%	5.7%
24	1	2	3	4	5	6	7	8	9	10	11	12
2016	21	13	23	34	3	0	0	0	2	12	13	11
2017	16	21	22	9	1	0	0	0	4	3	6	13
2018	4	25	23	27	0	0	0	0	5	3	12	19
2019	22	16	12	0	3	0	0	0	0	10	4	0
2020	8	7	2	0	1	0	2	1	2	0	2	3
total	71	82	82	70	8	0	2	1	13	28	37	46
百分比	16.1%	18.6%	18.6%	15.9%	1.8%	0.0%	0.5%	0.2%	3.0%	6.4%	8.4%	10.5%
29	1	2	3	4	5	6	7	8	9	10	11	12
2016	11	25	53	40	5	0	1	7	3	36	29	26
2017	17	19	36	38	12	0	1	3	12	10	8	18
2018	10	32	37	25	1	3	5	1	6	3	20	14
2019	18	19	11	4	0	0	0	3	2	22	11	2
2020	15	13	7	6	1	0	0	1	0	0	3	2
total	71	108	144	113	19	3	7	15	23	71	71	62
百分比	10.0%	15.3%	20.4%	16.0%	2.7%	0.4%	1.0%	2.1%	3.3%	10.0%	10.0%	8.8%

46	1	2	3	4	5	6	7	8	9	10	11	12
2016	52	44	80	34	18	0	0	12	14	37	47	66
2017	65	58	67	25	12	0	2	1	6	17	31	40
2018	38	58	46	20	4	1	0	0	3	32	38	29
2019	57	30	33	27	14	1	0	0	5	13	19	12
2020	35	30	17	2	0	0	0	0	0	3	5	16
total	247	220	243	108	48	2	2	13	28	102	140	163
百分比	18.8%	16.7%	18.5%	8.2%	3.6%	0.2%	0.2%	1.0%	2.1%	7.8%	10.6%	12.4%
57	1	2	3	4	5	6	7	8	9	10	11	12
2016	36	45	47	17	6	0	1	0	10	23	36	75
2017	91	64	86	26	24	0	1	5	15	52	79	87
2018	73	71	56	11	0	0	0	0	2	32	54	42
2019	76	28	28	14	6	0	0	0	6	10	18	24
2020	25	14	16	7	0	0	0	0	0	6	11	35
total	301	222	233	75	36	0	2	5	33	123	198	263
百分比	20.2%	14.9%	15.6%	5.0%	2.4%	0.0%	0.1%	0.3%	2.2%	8.2%	13.3%	17.6%
62	1	2	3	4	5	6	7	8	9	10	11	12
2016	0	2	0	1	0	0	0	0	1	0	0	2
2017	0	0	1	0	0	0	0	0	0	0	0	3
2018	0	0	1	0	0	0	0	0	0	0	0	0
2019	0	0	0	0	0	0	0	0	0	0	0	0
2020	0	0	0	0	0	0	0	0	0	0	0	0
total	0	2	2	1	0	0	0	0	1	0	0	5
百分比	0.0%	18.2%	18.2%	9.1%	0.0%	0.0%	0.0%	0.0%	9.1%	0.0%	0.0%	45.5%

PM10	101											
15	1	2	3	4	5	6	7	8	9	10	11	12
2016	2	9	2	4	0	0	0	0	0	0	0	0
2017	0	0	0	0	2	0	0	0	0	0	0	0
2018	0	5	3	6	0	0	0	0	0	0	0	0
2019	0	0	2	0	0	0	0	0	0	1	0	0
2020	0	0	0	0	0	0	0	0	0	0	1	0
total	2	14	7	10	2	0	0	0	0	1	1	0
百分比	5.4%	37.8%	18.9%	27.0%	5.4%	0.0%	0.0%	0.0%	0.0%	2.7%	2.7%	0.0%
24	1	2	3	4	5	6	7	8	9	10	11	12
2016	1	4	2	3	0	0	0	0	0	2	0	2
2017	1	4	6	0	3	0	0	0	0	0	1	4
2018	2	8	2	5	0	0	0	0	0	0	0	2
2019	2	1	3	0	0	0	0	0	0	1	0	0
2020	0	0	0	0	0	0	0	0	0	0	0	0
total	6	17	13	8	3	0	0	0	0	3	1	8
百分比	10.2%	28.8%	22.0%	13.6%	5.1%	0.0%	0.0%	0.0%	0.0%	5.1%	1.7%	13.6%
29	1	2	3	4	5	6	7	8	9	10	11	12
2016	1	7	7	9	0	0	0	0	0	13	9	2
2017	5	4	7	0	2	0	1	0	0	0	1	4
2018	2	6	4	6	0	0	0	0	0	1	4	1
2019	0	0	2	0	0	0	0	0	0	2	0	0
2020	1	2	0	0	0	0	0	0	0	0	0	0
total	9	19	20	15	2	0	1	0	0	16	14	7
百分比	8.7%	18.4%	19.4%	14.6%	1.9%	0.0%	1.0%	0.0%	0.0%	15.5%	13.6%	6.8%

46	1	2	3	4	5	6	7	8	9	10	11	12
2016	2	9	2	4	0	0	0	0	0	0	0	0
2017	0	0	0	0	2	0	0	0	0	0	0	0
2018	0	5	3	6	0	0	0	0	0	0	0	0
2019	0	0	2	0	0	0	0	0	0	1	0	0
2020	0	0	0	0	0	0	0	0	0	0	1	0
total	2	14	7	10	2	0	0	0	0	1	1	0
百分比	5.4%	37.8%	18.9%	27.0%	5.4%	0.0%	0.0%	0.0%	0.0%	2.7%	2.7%	0.0%
57	1	2	3	4	5	6	7	8	9	10	11	12
2016	5	6	3	0	1	0	0	0	0	4	18	31
2017	43	35	24	0	0	0	0	0	0	16	26	32
2018	25	19	9	3	0	0	0	0	3	14	11	21
2019	21	1	2	0	0	0	0	0	1	5	6	2
2020	2	0	0	0	0	0	0	0	1	0	0	1
total	96	61	38	3	1	0	0	1	4	39	62	86
百分比	24.6%	15.6%	9.7%	0.8%	0.3%	0.0%	0.0%	0.3%	1.0%	10.0%	15.9%	22.0%
62	1	2	3	4	5	6	7	8	9	10	11	12
2016	0	0	0	0	0	0	4	0	2	2	3	1
2017	0	0	0	0	0	0	1	1	0	4	3	0
2018	0	0	0	0	0	0	0	0	1	1	0	0
2019	0	0	0	0	0	0	0	0	1	0	0	0
2020	0	0	0	0	0	0	0	0	0	0	0	0
total	0	0	0	0	0	0	5	1	4	7	6	1
百分比	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	20.8%	4.2%	16.7%	29.2%	25.0%	4.2%

O3	71											
15	1	2	3	4	5	6	7	8	9	10	11	12
2016	0	0	2	4	1	2	2	2	1	2	0	0
2017	0	0	1	4	0	0	2	0	4	1	1	0
2018	0	0	1	3	3	5	0	0	1	1	0	0
2019	0	0	0	2	0	1	1	0	0	1	0	0
2020	0	0	0	6	3	0	0	0	4	0	0	0
total	0	0	4	19	7	8	5	2	10	5	1	0
百分比	0.0%	0.0%	6.6%	31.1%	11.5%	13.1%	8.2%	3.3%	16.4%	8.2%	1.6%	0.0%
24	1	2	3	4	5	6	7	8	9	10	11	12
2016	0	1	5	4	2	0	1	5	0	3	4	0
2017	0	0	1	4	1	0	0	1	5	1	0	1
2018	0	0	2	5	0	2	0	0	3	4	0	1
2019	0	0	2	2	2	2	0	2	2	3	2	0
2020	0	0	1	6	2	0	3	2	4	1	2	0
total	0	1	11	21	7	4	4	10	14	12	8	2
百分比	0.0%	1.1%	11.7%	22.3%	7.4%	4.3%	4.3%	10.6%	14.9%	12.8%	8.5%	2.1%
29	1	2	3	4	5	6	7	8	9	10	11	12
2016	0	1	4	5	3	0	1	6	1	5	2	0
2017	0	0	5	9	5	0	0	1	5	2	0	0
2018	0	1	2	7	2	4	3	0	5	9	2	1
2019	1	0	0	3	4	2	1	1	4	8	2	0
2020	0	1	2	4	1	0	0	0	5	1	0	0
total	1	3	13	28	15	6	5	8	20	25	6	1
百分比	0.8%	2.3%	9.9%	21.4%	11.5%	4.6%	3.8%	6.1%	15.3%	19.1%	4.6%	0.8%

46	1	2	3	4	5	6	7	8	9	10	11	12
2016	0	0	6	5	4	0	0	5	6	6	6	4
2017	0	2	11	12	7	0	1	0	5	10	2	1
2018	0	2	7	8	3	5	0	0	1	20	2	2
2019	2	3	7	6	7	4	0	0	6	14	7	0
2020	1	1	4	10	1	0	0	0	11	10	1	0
total	3	8	35	41	22	9	1	5	29	60	18	7
百分比	1.3%	3.4%	14.7%	17.2%	9.2%	3.8%	0.4%	2.1%	12.2%	25.2%	7.6%	2.9%
57	1	2	3	4	5	6	7	8	9	10	11	12
2016	1	1	6	5	4	0	0	1	4	6	11	4
2017	1	0	6	6	2	0	0	0	1	12	4	0
2018	0	0	2	1	2	1	0	0	3	15	6	3
2019	2	1	4	2	2	0	0	0	8	15	8	0
2020	2	1	2	8	0	0	0	0	8	2	1	1
total	6	3	20	22	10	1	0	1	24	50	30	8
百分比	3.4%	1.7%	11.4%	12.6%	5.7%	0.6%	0.0%	0.6%	13.7%	28.6%	17.1%	4.6%
62	1	2	3	4	5	6	7	8	9	10	11	12
2016	0	0	0	0	0	0	0	0	0	0	0	0
2017	0	0	0	0	0	0	0	0	0	0	0	0
2018	0	0	0	0	0	0	0	0	0	0	0	0
2019	0	0	0	0	0	0	0	0	0	0	0	0
2020	0	0	0	0	1	0	0	0	0	0	0	0
total	0	0	0	0	1	0	0	0	0	0	0	0
百分比	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

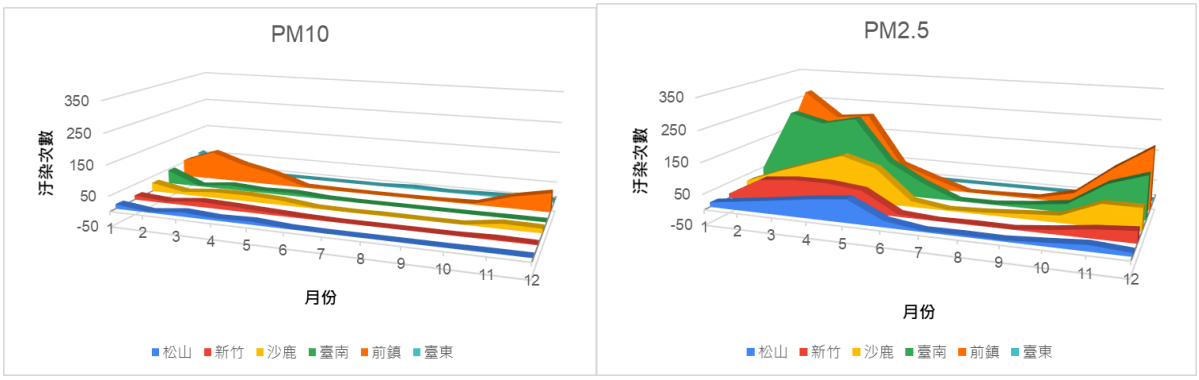


圖4－1－2 PM2.5 及 PM10 2016~2021年每月統計汙染時段數量

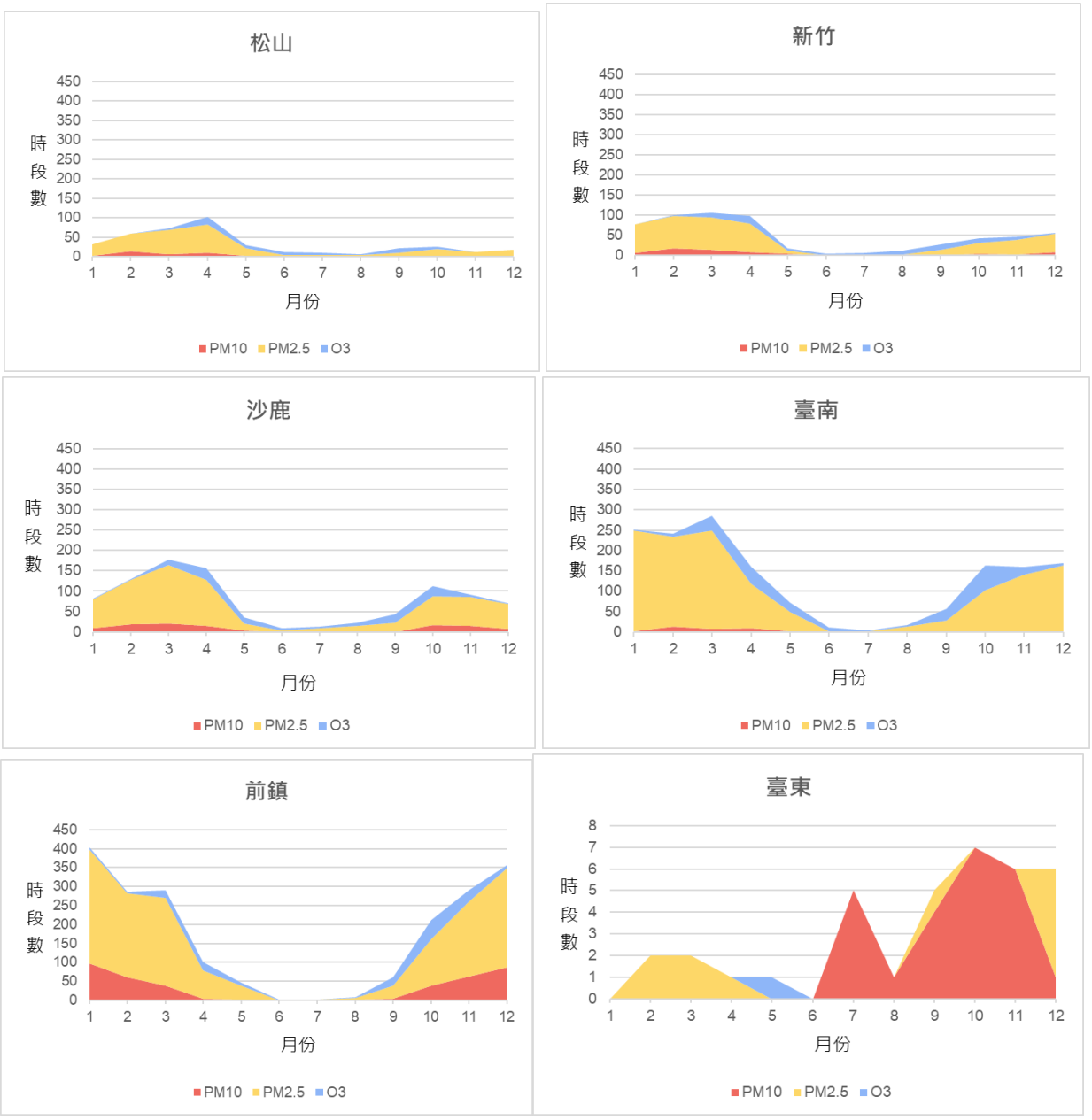


圖4－1－3各別觀測站自 2016 年～2020 年 PM2.5、PM10 及 O_3 每月累計汙染時段數量

二、以時序分析 ARIMA 的方式分析各觀測站 PM2.5、PM10 及 O_3 的空氣污染型態：

PM2.5、PM10 及 O_3 污染物的觀測量有季節的時序節律，若以 ARIMA 進行時序分析及預測比較。不同的觀測站結果，以預測的均方根誤差（RMSE Root Mean squart error）來衡量比較如表。

表4－2－1各測站 ARIMA 分析之 RMSE 值

	松山	新竹	沙鹿	台南	前鎮	台東
PM2	7.074	7.591	8.869	10.502	10.064	極大
PM10	12.203	13.457	15.648	18.906	17.382	極大
O3	極大	極大	極大	極大	極大	極大

以前鎮為例：

圖4－2－2（縱軸:汙染時段數 橫軸:時間）

左 觀測站前鎮的 PM2.5（紅）、PM10（藍）、O_3（黃）汙染物的觀測量的時序變化

右 觀測站前鎮的 PM2.5（紅）、PM10（深藍）與 CO（綠）、SO_2（淺藍）污染物的觀測量的時序變化

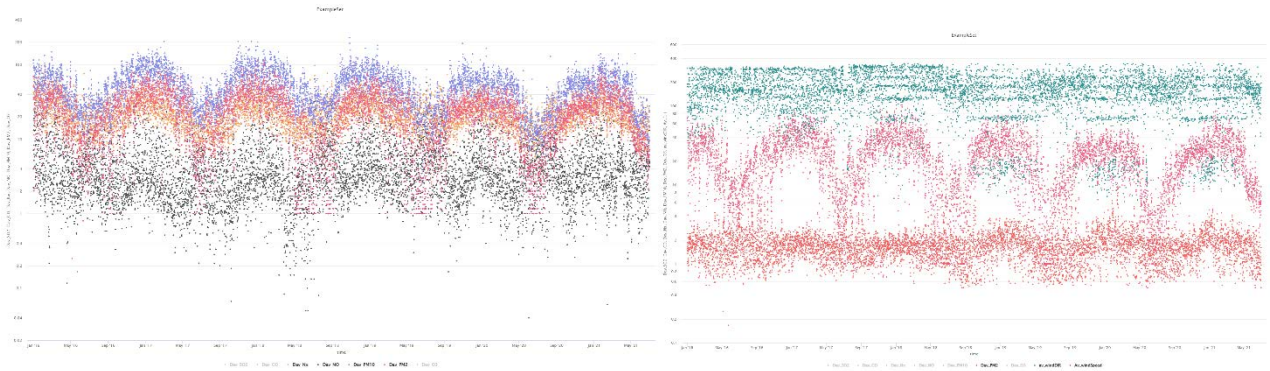


圖4－2－3（縱軸:汙染時段數 橫軸:時間）

左 觀測站前鎮的 PM2.5（紅）、PM10（藍） 與 NO（黑）、Nx（橘） 污染物的觀測量的時序變化

右 觀測站前鎮的 PM2.5（暗紅）、PM10（藍） 與風速（鮮紅）、風向（綠）的觀測量的時序變化

（1） 顯然從 ARIMA 的方法來看，O_3 的時序節律並不明顯。

（2） 松山、新竹、沙鹿 在 PM2.5 及 PM10 的 RMSE 值都優於 台南、前鎮。

（3） 台東在 PM2.5、PM10、O_3 的 RMSE 值皆為極大。

（4） PM2.5 的 RMSE 值都優於 PM10。

三、決策樹、邏輯迴歸、SVM 是完全不同的人工智慧方法，我們分別以

（1）單純以空氣污染物：CO、NO、NO_2、Nx、SO_2 進行預測分類。

（2）單純以天氣因子： 風向、風速、相對濕度、溫度、雨量進行分類預測。

將資料隨機分成 60% 訓練集與 40% 的測試集，從訓練集所得模型套用在測試集上的結果如下列表 4 - 3：採取空氣汙染的 F1 - score 值以兼顧空氣汙染特徵的精確率及召回率。

表4 - 3 決策樹、邏輯迴歸、SVM 預測所得 F1 - score 值

performance	PM2							performance(2)	PM2						
Decision	測站	前鎮	沙鹿	松山	新竹	臺南	臺東	Decision	測站	前鎮	沙鹿	松山	新竹	臺南	臺東
Tree	F1 SCORE	0.715328	0.601503	0.690607	0.744444	0.764705	0.475524	Tree	F1 SCORE	0.673202	0.669856	0.479452	0.177514	0.630806	0.465909
performance	PM2							performance	PM2						
Logistec	測站	前鎮	沙鹿	松山	新竹	臺南	臺東	Logistec	測站	前鎮	沙鹿	松山	新竹	臺南	臺東
Regration	F1 SCORE	0.669902	0.639097	0.798742	0.681992	0.744680	0.681081	Regration	F1 SCORE	0.619718	0.501960	0.511475	0.628352	0.565055	0.563218
performance	PM2							performance(2)	PM2						
svm	測站	前鎮	沙鹿	松山	新竹	臺南	臺東	svm	測站	前鎮	沙鹿	松山	新竹	臺南	臺東
	F1 SCORE	0.694444	0.640287	0.771704	0.713178	0.761565	0.707070		F1 SCORE	0.626865	0.622377	0.371900	0.606060	0.559386	0.552147
performance	PM10							performance(2)	PM10						
Decision	測站	前鎮	沙鹿	松山	新竹	臺南	臺東	Decision	測站	前鎮	沙鹿	松山	新竹	臺南	臺東
Tree	F1 SCORE	0.695652	0.724358	0.492307	0.511013	0.664723	0.612903	Tree	F1 SCORE	0.609090	0.8	0.2625	0.663316	0.668693	0.591836
performance	PM10							performance	PM10						
Logistec	測站	前鎮	沙鹿	松山	新竹	臺南	臺東	Logistec	測站	前鎮	沙鹿	松山	新竹	臺南	臺東
Regration	F1 SCORE	0.663101	0.715953	0.692307	0.653543	0.580357	0.666666	Regration	F1 SCORE	0.592964	0.530303	0.465346	0.518518	0.587155	0.759259
performance	PM10							performance(2)	PM10						
svm	測站	前鎮	沙鹿	松山	新竹	臺南	臺東	svm	測站	前鎮	沙鹿	松山	新竹	臺南	臺東
	F1 SCORE	0.714975	0.734848	0.666666	0.636015	0.611872	0.576576		F1 SCORE	0.580310	0.584615	0.577586	0.596363	0.602739	0.660377
performance	O3							performance(2)	O3						
Decision	測站	前鎮	沙鹿	松山	新竹	臺南	臺東	Decision	測站	前鎮	沙鹿	松山	新竹	臺南	臺東
Tree	F1 SCORE	0.994764	0.995594	0.992125	1	1	1	Tree	F1 SCORE	0.714859	0.740157	0.82	0.700421	0.671328	0.407079
performance	O3							performance(2)	O3						
Logistec	測站	前鎮	沙鹿	松山	新竹	臺南	臺東	Logistec	測站	前鎮	沙鹿	松山	新竹	臺南	臺東
Regration	F1 SCORE	0.994818	0.996655	0.988235	0.968036	0.980769	0.972602	Regration	F1 SCORE	0.563218	0.716723	0.823076	0.666666	0.675925	0.464
performance	O3							performance(2)	O3						
svm	測站	前鎮	沙鹿	松山	新竹	臺南	臺東	svm	測站	前鎮	沙鹿	松山	新竹	臺南	臺東
	F1 SCORE	0.955223	0.965811	0.980544	0.968609	0.949308	0.980132		F1 SCORE	0.62	0.776371	0.770370	0.607843	0.700934	0.389830

1. 決策樹、邏輯迴歸、SVM 三種模型，在選定的第（1）或（2）類資料上，F1 - score 值接近，其中決策樹有時會出現偏差的情況（例：臺東在 PM2.5 的 performance）。

2. 各測站的分析結果（F1-score大於 0.6 判斷預測結果為良好的標準；兩個數值落差 0.15 以上為有明顯落差）：

1.前鎮在 PM2.5 上空氣污染物預測結果與氣象因子預測結果 F1 - score 值相近。但 PM10 的預測結果有落差。

2.松山以空氣汙染物預測 PM2.5、PM10 的結果為良好，而氣象因子的預測結果很糟且兩者有明顯落差。

3.新竹、沙鹿以空氣汙染因子在 PM2.5 及 PM10 二種污染物的預測結果優於氣象因子，而在O_3預測方面兩者有較明顯落差。

4.臺南以空氣汙染物預測 PM2.5 的結果優於氣象因子；PM10的預測結果則為空氣汙染物和氣象因子相當。

5.臺東以氣象因子預測 PM10 的結果優於空氣汙染物；以空氣汙染物預測 PM2.5 的結果則優於氣象因子。

3. 不論哪一個觀測站，若以空氣汙染物因子進行 O_3 預測，預測結果非常好 F1－score 值皆介於0.95～1。若以氣象因子進行 O_3 預測則有時很糟 F1－score 值 0.3898，兩者有明顯落差。但以氣象因子進行 O_3 預測的效果較進行 PM2.5、PM10 來得好。

四、在決策樹、邏輯迴歸、SVM 人工智慧模型預測中，選擇預測結果良好的模型（AUC 值在 0.88 以上）分析 PM2.5、PM10 空氣汙染特徵的化學汙染物及天氣因子為何，並挑選出 p 值小於 0.025 者，做為決定性因子的判斷依據，結果如下：

1. PM2.5 方面

（1）觀測站松山以空氣汙染因子進行邏輯迴歸 MSE：0.1023、RMSE：0.3198 。 AUC：0.9330。平均一氧化碳 Dav_CO 有最佳的 p 值 0 (z 值 10.185)，其次為平均二氧化硫 Dav_SO_2 ，p 值為 0.000(z 值 4.973) 以及平均臭氧 Dav_O_3 ，p 值為 0.000(z 值 3.532)。

Attribute	Coefficient	Std. Coefficient	Std. Error	z-Value	p-Value
Dav_SO2	0.284	0.859	0.057	4.973	0.000
Dav_CO	15.659	4.276	1.537	10.185	0
Dav_O3	0.066	1.247	0.019	3.532	0.000
DMx_O3	-0.004	-0.109	0.012	-0.336	0.737
Dav_NO2	-0.382	-3.981	0.099	-3.860	0.000
DMx_NO2	0.124	1.570	0.043	2.844	0.004
Dav_NO	-0.176	-1.570	0.116	-1.518	0.129
DMx_NO	-0.029	-0.425	0.048	-0.603	0.547
Dav_Nx	0.242	4.237	0.093	2.604	0.009
DMx_Nx	-0.080	-1.899	0.036	-2.183	0.029
Intercept	-9.027	-0.626	0.747	-12.092	0

圖 4－4－1 松山以前驅汙染物因子第（1）類資料進行邏輯回歸預測結果

（2）觀測站新竹以空氣汙染因子進行邏輯迴歸，MSE：0.1123、RMSE：0.3350、 AUC：0.9168 。平均一氧化碳 Dav_CO 有最佳的 p 值 0（z 值 9.299），其次為平均臭氧 Dav_O_3，p 值為 0.000（z 值 3.704），再其次為最大氮化物 DMx_Nx，p 值為 0.122（z 值 1.546）。

（3）臺南站以空氣汙染因子進行邏輯迴歸 MSE：0.1170、RMSE：0.3420。 AUC：0.9125。平均一氧化碳 Dav_CO 有最佳的 p 值 0 (z 值 5.941)，其次為平均二氧化硫 Dav_SO_2 ，p 值為 0.003(z 值 3.009)，最後是平均臭氧 Dav_O_3 ，p 值為 0.015(z 值 2.433)。

（4）前鎮站以空氣汙染因子進行邏輯迴歸 MSE：0.1351、RMSE：0.3676。 AUC：0.8819。平均一氧化碳 Dav_CO 有最佳的 p 值 0 (z 值 5.996)。

從上述結果可知在以汙染因子預測松山、新竹、臺南、前鎮的模型中，平均一氧化碳 Dav_CO 皆有最小的 p 值（0.000）。因此我們從新只以 Dav_CO 預測，得到的混淆矩陣值與原先的比較如下表。僅只以平均一氧化碳 Dav_CO 即有良好的預測結果。

表 4－4－3 以 Dav_CO 再次預測新竹的 PM2.5

av_CO	Decision Tree		Logistec Regration		svm		第(1)類資料					
測站	新竹		新竹		新竹		新竹		新竹		新竹	
	F	T	F	T	F	T	F	T	F	T	F	T
N	193	8	253	50	257	56	140	12	276	41	247	51
P	83	137	23	95	19	89	62	133	31	194	29	94
精確率	0.6227		0.8051		0.8241		0.6821		0.8622		0.7642	
召回率	0.9448		0.6552		0.6138		0.9172		0.8255		0.6483	
準確率	0.7838		0.8266		0.8219		0.7867		0.8672		0.81	
F1 SCORE	0.7507		0.7224		0.7036		0.7824		0.8435		0.7015	

2. PM10方面

（1）松山以空氣汙染因子進行邏輯迴歸，MSE：0.1295、RMSE：0.3598、AUC： 0.8836 。平均一氧化碳 Dav_CO 有最佳的 p 值 0.000（z 值 7.344）以及平均臭氧 Dav_O_3 ，p 值為 0.000（z 值 5.321），其次為最大臭氧DMx_O_3 ，p 值為 0.005（z 值 -2.788）。

（2）臺東站以氣象因子進行邏輯迴歸，MSE：0.0980 、RMSE：0.3131、AUC：0.8849。最大降雨量 DMx_RAINFALL 有最佳的 p 值 0.002（z 值 3.076）。

在進行決定因子再分析的過程中，我們發現以汙染因子 O_3 預測效果最差；以氣象因子風速、雨量的預測效果會很差（有時F1-score 小於 0.1），推測是因為這些因子的變動幅度大。

五、以 K－means 聚類的方式分析各觀測站 PM2.5 、 PM10 及 O_3 的空氣污染型態：

根據 PM2.5、PM10 及 O_3 高於平均兩個標準差標對聚類中心分組，來觀察各觀測站的空氣污染型態：

表4－5 K-means 聚類結果（紅字部分為數值大於或等於兩個標準差者；紅色框為數值大於或等於三個標準差者）

松山												
分類	cluster0	cluster1	cluster2	cluster3	cluster4	cluster5	cluster6	cluster7	cluster8	cluster9	cluster10	cluster11
PM10	20.11	92.39	45.77	35.29	63.57	160.33	30.64	15.64	75.52	62.87	36.65	46.62
O3	14.38	36.59	78.25	54.46	52.23	48. 19	35.32	36.70	92.54	18.67	15.24	37.10
PM2.5	11.44	46.50	27.32	19.14	32.90	54.02	15.54	10.15	44.23	35.43	19.15	23.64
新竹												
PM10	22.22	39.65	70.30	18.02	116.94	260.40	36.27	86.66	55.80	46.72	31.53	60.38
O3	16.88	18.54	24.61	37.69	41.08	68.64	57.07	57.12	77.95	38.11	38.32	49.77
PM2.5	10.47	20.76	40.05	9.45	59.67	55.00	18.66	43.52	29.79	22.33	14.58	29.62
沙鹿												
PM10	20.12	102.81	42.25	146.47	36.10	17.83	56.74	69.52	92.16	73.12	33.37	51.86
O3	18.12	37.33	19.54	49.23	57.76	37.45	74.24	24.58	74.14	53.32	36.90	44.24
PM2.5	9.74	54.01	21.96	76.63	17.54	8.67	29.34	38.48	50.29	36.00	16.32	25.66
台南												
PM10	60.35	20.15	185.34	75.26	55.94	37.00	39.10	92.64	119.06	311.94	129.30	86.94
O3	27.30	23.73	56.46	49.48	71.62	23.25	47.89	28.82	52.73	58.28	49.05	77.76
PM2.5	32.47	10.34	41.13	36.04	30.76	17.97	21.12	56.71	40.22	45.50	86.38	48.96
前鎮												
PM10	93.74	71.41	109.78	122.50	38.16	160.00	281.64	65.04	45.51	61.07	22.25	91.05
O3	52.52	73.53	77.68	32.31	27.19	47.83	55.94	43.70	58.72	18.03	17.57	21.19
PM2.5	46.39	37.40	60.05	71.83	17.12	43.50	71.50	34.90	24.10	31.31	8.47	47.54
台東												
PM10	53.53	34.36	244.96	143.71	26.17	81.49	42.72	29.58	19.81	666.67	12.72	17.54
O3	47.52	44.02	40.59	36.02	30.46	37.95	28.42	17.08	43.45	24.33	29.91	15.24
PM2.5	21.67	15.28	23.23	22.17	10.79	17.56	14.95	10.29	8.66	29.00	6.77	6.54

1.松山空氣污染可能的情形是 PM2.5 、PM10 同時或 PM2.5 、O_3 同時或僅 O_3 污染。

2.新竹與沙鹿空氣污染可能的情形是 PM2.5、PM10 同時或僅 O_3 污染。

3.台南空氣污染可能的情形是僅 PM10 或僅 PM2.5 污染。

4.台東空氣污染可能的情形是 PM2.5、PM10 同時或僅 PM10 污染。

5.前鎮空氣污染可能的情形是 PM2.5、PM10 同時或僅 PM10 污染或僅PM2.5 污染。

6.這些觀測站的空氣污染型態 O_3 污染時多半不會伴隨 PM2.5、PM10 的污染。

伍、討論

一、人工智慧方法特點比較：

我們選擇了五種人工智慧方式進行空氣汙染分析與預測方式，對於汙染物的相關性、時序分布的節律、預測等有不同的優勢預測，表 5-1 為五種人工方式的特色比較，針對不同的污染模式測站特性等因此針對不同的污染模式、測站特性等，可參考不同的方式進去分析。

表 5-1 人工智慧分法比較

人工智慧方法	用途	特點
差分整合移動平均自迴歸模型	1.得出各類汙染物季節的時序節律 2.相互比較各類汙染物時序的節律	透過圖表方式清楚得知汙染物時段數時序的變化
決策樹	1.以兩大類因子(氣象因子、汙染物因子)對PM2.5、PM10和O3進行預測 2.比較三種方法的結果，都預測良好的結果留下 3.挑出汙染物的決定因子，進行再一次的預測	1.決策樹預測結果與其他兩者偏差大 2.在以SVM 預測的模型中，預測的效果通常較決策樹、邏輯迴歸佳
邏輯迴歸		
支持向量機		
K－平均演算法	1.求出各測站各類汙染物的分布型態	

二、人工智慧分析結果：

（一）汙染物時段數時序變化探討：

基本統計中可發現 PM2.5 汙染時段數高於 PM10、O_3 且 RMSE 值最小，可能是 PM10 質量較大，多會在幾個小時之內因為地心引力而落地，而 PM2.5 卻會長時間在空氣中懸浮，也導致其時序節律與風速較吻合。對各個測站而言，O_3 汙染時段數，西南部多於東北部測站，可能是因為臭氧來源的光化反應會受濕度、溫度、風的影響，較高的溫度及日照量有助於臭氧的形成（陳鴻烈,羅惠芬），而臺南、高雄所屬南部日照較長、溫度普遍較北部高。PM2.5、PM10 汙染時段數量多的月份恰好與東北季風好發季節吻合，推測東北季風會加速境外汙染物的擴散至北部地區，西南部則因受地形阻擋較少影響。但在圖 4－1－3 中松山測站冬季汙染物上升幅度小並結合結果三（表4－3）松山測站污染因子的預測比氣象因子佳，推測是因為汙染物來源在地多於境外的緣故。

綜合污染物的觀測量與 ARIMA 的結果，台東觀測站 PM2.5、PM10 及 O_3 與各觀測站的 O_3 汙染時段數，ARIMA 預測所得 RMSE 值都極大，推測是因為臺東建設較少且多為一、二級產業，汙染排放較其他觀測站少，其污染物生成多為偶發性、單一事件。季節節律方面， P

M2.5 和 PM10 的時序節律極度相似，可能是 PM2.5 和 PM10 來源組成類似且 PM10 有包含 PM2.5。在氣象因子中，風速節律明顯，風向節律不明顯，可見風速對於汙染物時段數的影響大，推測是因為汙染物隨風飄散是需要足夠的能量才能帶動粒子移動，光是風向無法進行。在空氣污染物中，CO、Nx 和 NO 與 PM2.5 和 PM10 具有相似的節律，可推知這些汙染物與 PM2.5 和 PM10 關聯大，而 SO_2 節律較不明顯，關聯性較低。綜合結果四，在許多測站平均一氧化碳多為 PM2.5 和 PM10 的決定因子，可見一氧化碳的排放與 PM2.5 和 PM10 汙染物有極大的關聯。

（二）空氣汙染因子、氣象因子預測結果：

以空氣汙染因子對每個測站的 PM2.5、PM10 進行預測，都有一定的效果，F1-score 多大於 0.6（此亦是我們判斷預測結果為良好的標準），推測與 PM2.5、PM10 本身的組成相關：PM2.5、PM10 的組成有部分來自一氧化氮、氮化物衍生的硝酸鹽、一氧化碳衍生的碳酸鹽，或是二氧化硫衍生的硫酸鹽等，另外，觀察邏輯迴歸預測結果良好的模型，發現在以空氣汙染因子預測良好的模型中，一氧化碳多是決定性因子，表示 PM2.5、PM10 與空氣汙染物有很大的關連。

以空氣汙染因子對 O_3 進行預測，F1-score 都大於 0.95，推測與 O_3 的光化反應形成有關：NO 經光化反應後，會產生 O_3 ，造成 O_3 與空氣汙染物 NO 的關聯性極大，而以氣象因子進行 O_3 預測，效果雖較以空氣汙染因子來的差，但仍有一定的效果（除了台東外其餘測站 F1-score 多大於 0.6），推測是因為 O_3 光化反應會受到溫度、濕度、日照量等氣象因子的影響，其中日照量充足的南部地區，O_3 汙染比北部嚴重（台南＞前鎮＞沙鹿，新竹、松山、台東幾乎沒有）。

反觀以氣象因子進行 PM2.5、PM10 的預測結果，推測與各測站的地理條件、產業有關，擁有高密度的人口、車輛以及盆地地形的松山、具發達的高科技產業與眾多人口的新竹、有重工業且位盆地地形的沙鹿，以氣象因子預測 PM2.5、PM10 的結果差，可能是因為汙染物會在往往會積聚在盆地地形中，難以隨風擴散（莊銘棟，2008），造成空氣汙染物對預測結果的影響蓋過了氣象因子對預測結果的影響。以氣象因子和空氣汙染因子預測效果相當的台南和高雄，其周遭環境同樣是重工業、古蹟等地所在地，會產生大量汙染，透過風自外地帶來大量汙染物，且台南、高雄平原地形開闊，汙染物不易累積，空氣汙染因子的預測較不會去影響到氣象因子的預測，故預測效果相當。

（三）PM2.5、PM10、O_3 汙染模式：

結果五（表４－5）中，單一測站來看，3 項不同的空品指標分成良好與不良應有 8 種可能，但分類結果最多只有四大類組，可見 PM2.5、PM10 及 O_3 這三項汙染有一定關聯性存在。且空氣污染型態 O_3 污染時多半不會伴隨 PM2.5、PM10 的污染；PM2.5、PM10 幾乎同時發生。

三、各測站分析：

1. 松山測站：由結果二（表 4-2-1）顯示，ARIMA 預測 PM2.5、PM10 的 RMSE 值都很小，可見兩汙染物具時序節律性，結果三（表 4-3）顯示，以空氣汙染因子預測 PM2.5、PM10 的效果明顯較氣象因子佳。結果四中，以空氣污染因子預測方面，PM2.5 的決定性因子是 CO，其次為 SO_2 ；PM10 的決定性因子是 CO及 O_3 。我們推測是因測站地處台北盆地，且當地的服務業繁盛，人潮、車潮多（見表 1），連帶造成空氣汙染物排放量多，CO 排放量也多，其衍生物組成 PM2.5 的成分，再加上 PM2.5 和有機質往往會積聚在盆地地形中，難以隨風擴散（莊銘棟，2008），可推測松山的 PM2.5、PM10 多來自本地排放或隨風而來的汙染物受地形影響停留。
2. 新竹測站：在結果二（表 4-2-1）發現，ARIMA 預測 PM2.5、PM10 的 RMSE 值都很小，可見兩汙染物具時序節律性。結果三（表 4-3）顯示，以空氣汙染因子預測 PM2.5、PM10 效果優於氣象因子。而在結果四中，在空氣汙染因子預測方面，PM2.5 的決定性因子為 CO，其次為 O_3，並只用其單一測量即有良好的預測結果。我們推測是因測站具有許多本地產業的汙染物，加上新竹當地的服務業也很繁盛（見表 1），造成 CO 排放量也多。故推測的 PM2.5、PM10 多來自本地排放或是自外地擴散而來在本地聚積。
3. 沙鹿測站：從結果二（表 4-2-1）得知，ARIMA 預測 PM2.5、PM10 的效果較松山 、新竹測站差；結果三（表 4-3）中，以空氣汙染因子預測 PM2.5、PM10 效果優於氣象因子，再綜合結果一中沙鹿測站的汙染時段數較多，且沙鹿為重工業重地，有火力發電廠等二級工業場所，另有中科園區高科技產業進駐（見表１），重工業燃燒煤礦不完全產生的 CO 形成衍生性微粒，組成 PM2.5 和 PM10 的成分，推知沙鹿的汙染受本地排放的影響極大。
4. 台南測站：結果三（表 4-3）中，以空氣汙染因子預測 PM2.5 的效果優於氣象因子；PM10的預測結果則為空氣汙染因子和氣象因子相當。再綜合結果一中汙染時段數較多。推測是因為台南有許多古蹟，每到假日就會聚集觀光人潮、車潮，再加上眾多廟宇燒香排放的煙，使得本地 PM2.5 及 PM10 排放量多，且因所處地形平坦，無山

脈阻擋風，PM2.5 及 PM10 容易自同位於市中心的鄰近區域擴散而來，使得預測結果不易只受單一因子影響。

5. 前鎮測站：在結果三（表 4-3）中，前鎮在 PM2.5 上空氣污染因子與氣象因子預測效果相近，但 PM10 的預測效果則有落差，綜合結果一的數據，前鎮測站的 PM2.5 及 PM10 的汙染時段總數皆是所有測站中最多，以及前鎮是重工業重地，有中鋼集團、加工出口區等，並包含部分高雄港，海上船隻及陸上運輸車來往排放的廢氣（見表一），造成本地的汙染排放嚴重，另外，前鎮的鄰近區域同樣有重工業進駐，且前鎮測站地形平坦，使汙染物容易受風的影響擴散而來，故以氣象因子和空氣汙染物來預測皆會有一定效果。而 PM10 以空氣汙染物預測明顯比氣象因子好，同上所述，是因為處於重工業重地，但 PM10 不易被風吹送，故以氣象因子預測效果不好。另外，在結果四的數據顯示，和 PM2.5 關聯性最大的是 CO，推測同沙鹿般是因為重工業燃燒煤礦不完全產生的 CO 形成衍生性微粒，組成了 PM2.5 、PM10 的成分。
6. 台東測站：結果三（表 4-3）呈現出，臺東以氣象因子預測 PM10 的結果優於空氣因子，可推測汙染來源較多都為外來汙染，且當地產業多為汙染較輕微的第一級產業（見表一），外來因子影響大；以空氣汙染因子預測 PM2.5 的效果則優於氣象因子。結果一中，臺東汙染時段數較不具季節差異性，且其數量比其他測站少許多，綜合結果二（表 4-2-1）ARIMA 運算得出的 RMSE 值皆為無限大，得出台東的汙染情況較其他測站輕微且多為偶發。

四、各測站比較：

由結果一顯示，各測站 2020 年汙染時段明顯變少，應該是受 COVID-19 疫情影響，工業及經濟活動減少所致。同為重工業重地的沙鹿和前鎮，以及人文產業繁盛的臺南，汙染時段數量多。結果二（表 4-2-1）中，松山、新竹的 RMSE 值相近。結果三中（見表 4-3），松山、新竹、沙鹿預測模式相近。結果四中，PM2.5 方面，松山、新竹、臺南、前鎮的關鍵因子都是 CO；PM10 方面，松山的關鍵因子也是 CO 。結果五中，汙染型態多是 PM2.5、PM10 同時或僅 O_3 污染；各站的氣候特性皆不同，但各站中的 PM2.5 與 PM10 及 O_3 都有很大的相關性，此表示區域的氣候影響比地區的事件更加顯著。

表5-2 研究結果統整

測站名稱/研究結果	一	二	三	四	五
松山	PM2.5、PM10 有季節差異性，O_3 則無	PM2.5、PM10 的時序RMSE值小，O_3 極大。	以空氣汙染因子預測 PM2.5、PM10 效果較以氣象因子佳。	PM2.5：汙染因子中的平均一氧化碳是決定性因子 PM10：汙染因子中的平均一氧化碳是決定性因子	1.僅 O_3 污染 2.PM2.5 、 PM10 同時汙染 3.PM2.5 、 O_3 同時汙染
新竹	PM2.5、PM10 有季節差異性，O_3 則無	PM2.5、PM10 的時序RMSE值小，O_3 極大。	以空氣汙染因子預測 PM2.5、PM10 效果較以氣象因子佳；預測 O_3 則有明顯落差。	PM2.5：汙染因子中的平均一氧化碳是決定性因子	1.僅 O_3 污染 2.PM2.5、PM10 同時汙染
沙鹿	PM2.5、PM10 有季節差異性，O_3 則無	PM2.5、PM10 的時序RMSE值中等，O_3 極大。	以空氣汙染因子預測 PM2.5、PM10 效果較以氣象因子佳；預測 O_3 則有明顯落差。	PM2.5：汙染因子中的平均一氧化碳是決定性因子	1.僅 O_3 污染 2.PM2.5、PM10 同時汙染
台南	PM2.5、PM10 有季節差異性，O_3 則無，總汙染次數次多。	PM2.5、PM10 的時序RMSE值中等，O_3 極大。	以空氣汙染因子預測 PM2.5 與以氣象因子預測結果皆良好；以空氣汙染因子及氣象因子預測、 PM10 效果相當。	PM2.5：汙染因子中的平均一氧化碳是決定性因子	1.僅 PM2.5 汙染 2.僅 PM10 汙染
前鎮	PM2.5、PM10 有季節差異性，O_3 則無，總汙染次數最多。	PM2.5、PM10 的時序RMSE值中等，O_3 極大。	以空氣汙染因子預測 PM2.5 與以氣象因子預測結果皆良好；以空氣汙染因子預測 PM10 效果較以氣象因子佳。	PM2.5：汙染因子中的平均一氧化碳是決定性因子	1.僅 PM2.5 汙染 2.僅 PM10 汙染 3.PM2.5 和 PM10同時汙染
台東	PM2.5、PM10、O_3無季節差異性，總汙染次數最少。	PM2.5、PM10、O_3 的 RMSE 值都極大。	以空氣汙染物預測 PM2.5 效果較以氣象因子佳；以氣象因子預測 PM10 效果較以汙染因子佳。	PM10：氣象因子中的最大降雨量是決定性因子	1.僅 PM10 汙染 2.PM2.5 和 PM10 同時汙染

陸、結論

一、結論

（一）、PM2.5、 PM10 及 O_3 這三項汙染有一定的關聯性存在。且空氣污染型態 O_3 污染　時多半不會伴隨 PM2.5、PM10 的污染；PM2.5、PM10 幾乎同時發生。

（二）、空氣污染物 CO 、 NO 、 NO_2、 Nx、 SO_2 進行預測分類，不論是季節節律、模型預測中 CO、 Nx 和 NO 與 PM2.5 和 PM10 節律相似關聯大，而 SO_2 則較不明顯。

（三）、單純以天氣因子：風向、風速、相對濕度、溫度、雨量進行分類預測，風向、風速及溫度是明顯重要的因子，其他因子與 PM2.5 和 PM10 污染沒什麼關聯。

（四）、O_3 汙染多為偶發性所以節律不明顯。因為光化反應，南部日照較長，因此 O_3 汙染程度：台南＞前鎮＞沙鹿，而新竹 、松山及台東則很少。

（五）、新竹、台南、前鎮、沙鹿、前鎮的 PM2.5、PM10 皆以汙染因子預測較佳，其中僅以平均一氧化碳 Dav_CO 即有良好的預測結果。台東的 PM2.5 以汙染因子預測較佳，PM10 則以氣象因子預測較佳。台東汙染較不具季節差異性，多為偶發且較輕微。

二、未來展望

人工智慧有別於一般統計及回歸直線分析趨勢，它能更精確且多方面得出數據的時序變化、與其他因子的相互關聯性，我們希望在將來

（一）將時間因子引進kmeans分析，詳細探討時間序列變化與汙染源的關係，現階段只有進行汙染源的聚類分析，無法得知汙染源聚類對應的時間

（二）利用人工智慧得出預測氣象類因子跟污染類因子的 p、z 值，並找出其決定性的因子。

（三）再利用深度學習（Deep Learning）計算（二）找出的決定性因子，

（四）調查各汙染源排放的汙染物種類，藉此更詳細了解各測站周邊地區的汙染物情形，進行污染的管制以及預防。

柒、參考文獻資料

【1】葉惠中主持（2014）。細懸浮微粒（PM2.5）時空分佈特性之研究（102　年度「環保署/國科會空污防制科研合作計畫」期末報告，NSC102－EPA－F－005－001）。中國文化大學土地資源學系。

【2】蔡俊鴻主持（2008）。超細氣懸微粒水溶性無機離子成份特性變異與前驅物及氣象因子關聯性研究（行政院國家科學委員會補助專題研究計畫成果報告，NSC96－2211－E－006－024）。國立成功大學環境工程學系（所）。

【3】李佾、徐如怡、沈姿妤（2020）。走出陰霾－台灣大氣懸浮微粒含量變化與天氣狀況、汙染物種類特徵探討（中華民國第 60 屆中小學科學展覽會作品說明書）。桃園市立內壢高級中等學校。

【4】呂柏諺、何昌倫、李德威（2018）。「清」出於「蘭」－宜蘭地區之氣粒狀汙染物於不同時間尺度下之研究（中華民國第 58 屆中小學科學展覽會作品說明書）。國立宜蘭高級中學。

【5】林華恩、楊于晨、葉家榮（2016）。漫漫黃砂九芎城－宜蘭地區汙染物濃度的探討（中華民國第 56 屆中小學科學展覽會作品說明書）。國立宜蘭高級中學。

【6】張艮輝、陳建鴻。台灣地區臭氧污染控制物種之分析。國立雲林科技大學環境與安全系。

【7】曾于庭、吳治達、龍世俊（2018）。應用土地利用迴歸模式推估北部空品區細懸浮微粒之時空分布。航測與遙測學刊，23，191～204。

【8】 Shixiong Wang, Chongshou Li, and Andrew Lim （2021）. Why Are the ARIMA and SARIMA not Sufficient, arXiv：1904.07632v3 [stat.AP]

【9】Savitzky－Golay filter,Retrieved from：https://en.wikipedia.org/wiki/Savitzky%E2%80%93Golay_filter

【評語】051907

1. 計畫概述：本研究作者對於空汙的問題有興趣，想利用人工智慧的各種演算法的計算，來分析造成空汙的原因的幾種可能主要因子，並以台灣六個不同區域（北、中、南、東等）的環保局測站資料，進行計算分析及探討空汙因子。

2. 綜評：

   (1) 報告撰寫非常的完整，利用人工智慧的演算法來進行科學分析（本研究為空汙因子的分析），是個現今科學家很流行的想法及作法，很不錯。在空汙因子的前置文獻閱讀消化上，整理的很詳盡，很好。報告的內容圖文並茂，顏色的選擇看起來很舒服。各類演算法的分析很專業。分析討論的敘述也很詳盡，分成空汙分析及演算法技術層面來討論，非常好。是個很不錯的科學專題研究。

   (2) 投影片的製作上，很有條理，圖文並茂。

   (3) 研究日誌的紀錄，也是很有次序，加上貼圖及手繪，很不錯。

3. 建議：

   (1) 簡報檔做的很好，只是字體有些有點小，讀者閱讀上可能會有一些難度，可以考慮放大一些？也許文字也可以再精簡一點？

   (2) 在高中階段，就已經可以嫻熟的操作幾種不同的數值分析技術，來分析自然界的現象，程度非常好，持續努力，很有前景。如果除了操作現成的軟體外，也可以自己寫程式，將會更可以把不同想法或模式串連起來，也更可以發揮創意。

作品簡報

人工智慧進行汙染因子的模式及時序分析

組別:高級中等學校組

科別:地球與行星科學科

前言

摘要

本研究蒐集松山、新竹、沙鹿、台南、前鎮、台東六個北中南東空品測站的空氣汙染和氣象因子資料,並用人工智慧:決策樹、邏輯回歸、支持向量機、K–平均演算法、ARIMA 分析 PM2.5、PM10、O3 的時空分布模式及汙染預測模型。

研究動機

在現今科技中，仍無法精確對空氣汙染進行預測,而汙染來源的擴散因很難有效的標定，難以用動力學模型來預測懸浮微粒的時空變化。若能夠根據各測站氣象因子與化學污染物觀測值建立懸浮微粒隨時序變化的模式並透過機器學習的方法建立各種氣象因子與化學污染物與 PM2.5、PM10 及 O3 的關聯性,勢必能對健康以及汙染防制會有正面的影響。

研究目的

1. 各監測站之每小時資料進行懸浮微粒污染程度分析，了解 PM2.5、PM10、O3 含量的時空變化。
2. 以人工智慧方法 ARIMA、DT、LR、K–means、SVM 進行因子對 PM2.5、PM10及O3 的分析。
3. 針對汙染的影響因子進行再預測，找出 PM2.5、PM10 空氣汙染特徵的化學汙染物及天氣因子。
4. K–means 聚類方式分析各觀測站 PM2.5、PM10 及 O3 的空氣污染型態、並比較測站間關聯性。

文獻回顧

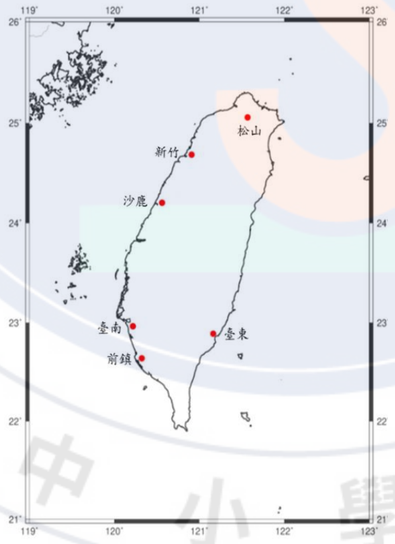
汙染因子

PM2.5　　PM10　　O3

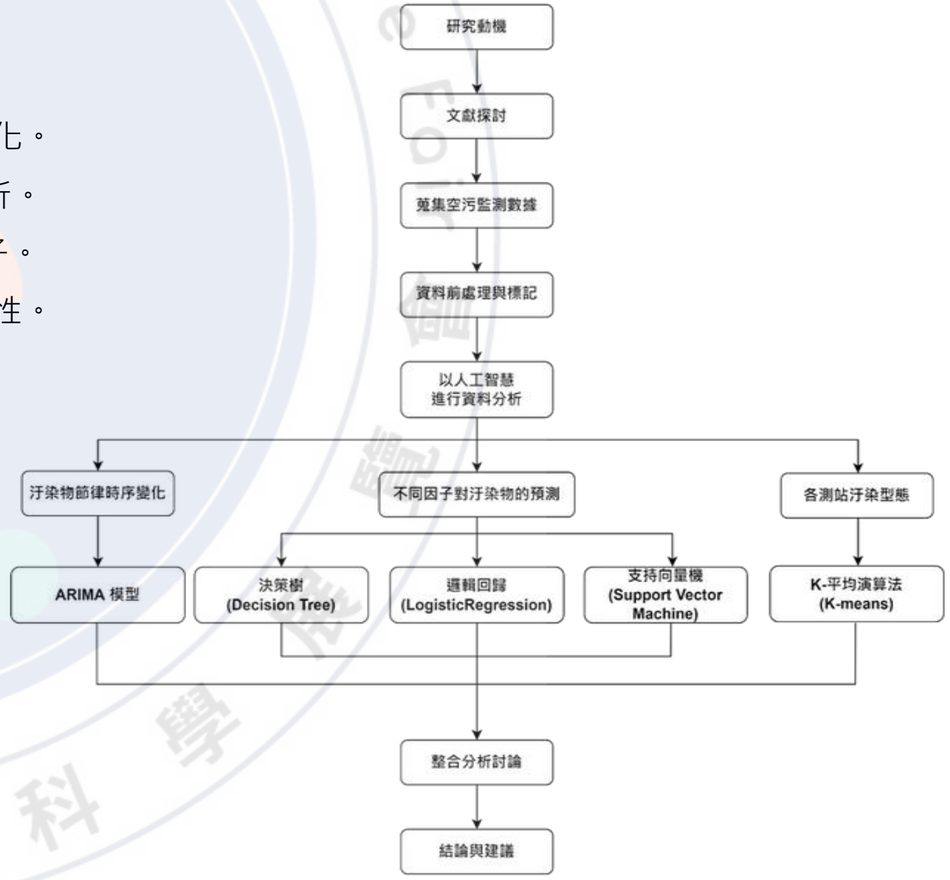
選用測站

松山（15）　新竹（24）　沙鹿（29）

臺南（46）　前鎮（57）　臺東（62）



圖一 各測站在台灣分布位置



圖二 研究流程圖

資料的前處理與標記

將各測站原始資料 CO、NO、 NO2、Nx、SO2、PM2.5、PM10、O3 及風向、風速、相對濕度、溫度、雨量每一小時的觀測值進行 6 小時為時段處理，每日分成 3~9、9~15、15~21、21~3 四個時段，並對每一時段的最大、最小及平均值整理。將高於平均兩個標準差標記為污染，範圍介於保護署的空氣品質汙染指標對敏感族群不健康及對全部族群不健康之間。

空氣品質指標 (AQI)				
AQI 指標	O_3 (ppm) 8 小時平均值	O_3 (ppm) 小時平均值 (1)	$PM_{2.5}$ (μg/m^3) 24 小時平均值	PM_{10} (μg/m^3) 24 小時平均值
良好 0～50	0.000 - 0.054	-	0.0 - 15.4	0-50
普通 51～100	0.055 - 0.070	-	15.5 - 35.4	51-100
對敏感族群不健康 101～150	0.071 - 0.085	0.125 - 0.164	35.5 - 54.4	101-254
對所有族群不健康 151～200	0.086 - 0.105	0.165 - 0.204	54.5 - 150.4	255-354
非常不健康 201～300	0.106 - 0.200	0.205 - 0.404	150.5 - 250.4	355-424
危害 301～400	(2)	0.405 - 0.504	250.5 - 350.4	425 - 504
危害 401～500	(2)	0.505 - 0.604	350.5 - 500.4	505-604

表一 空氣品質指標(AQI)

測站	松山	新竹	沙鹿	臺南	前鎮	臺東
汙染源	PM2.5					
平均	20.419	20.685	22.782	29.636	30.173	11.015
標準差	11.26	12.661	15.161	18.433	19.587	6.563
最大值	134	104	137	140	445	147
兩個標準差	42.939	46.007	53.104	66.502	69.347	24.141
三個標準差	54.199	58.668	68.265	84.935	88.934	30.704
汙染源	PM10					
平均	38.249	41.242	44.877	58.173	60.969	29.285
標準差	21.37	22.835	26.264	35.652	34.948	24.068
最大值	249	460	253	396	489	1000
兩個標準差	80.989	86.912	97.405	129.48	130.87	77.421
三個標準差	102.36	109.75	123.669	165.13	165.81	101.49
汙染源	O_3					
平均	36.85	39.161	40.52	41.441	37.635	30.529
標準差	20.118	17.668	18.495	21.399	22.22	12.017
最大值	168	142	131	143	135	92
兩個標準差	77.086	74.497	77.51	84.239	82.075	54.563
三個標準差	97.204	92.165	96.005	105.64	104.3	66.58

表二 PM2.5、PM10 及 O3 空氣品質平均值及標準差
(2016 年 1 月 1 日至 2021年 6 月 30 日)

● 人工智慧方法

以五種人工智慧:決策樹、邏輯回歸、支持向量機、K–平均演算法、ARIMA 分成三種用途分析 PM2.5、PM10、O3 的時空分布模式及汙染預測模型。

用途一

利用三種人工智慧方法進行前驅汙染物因子及氣象因子對PM2.5、PM10、O3的預測

- 前驅汙染物因子 : 二氧化硫、二氧化氮、一氧化氮、氮化物
- 氣象因子 : 濕度、雨量、溫度、風向、風速

1.決策樹（Decision Tree）: 將標記有空氣污染與空氣品質良好進行樹形結構分類，以最大資訊增益方式演算。

$$H(D,A)=\frac{|S_a|}{|D|}\cdot H(S_a)+\frac{|\widetilde{S}_a|}{|D|}\cdot H(\widetilde{S}_a)$$

增益值公式為 $g(D,a)=H(D)-H(D,A)$

考慮特徵 A 選定值 a 時的熵為 $H_A(a)=-\frac{|S_a|}{|D|}\cdot log_2\left(\frac{|S_a|}{|D|}\right)-\frac{|\widetilde{S}_a|}{|D|}\cdot log_2\left(\frac{|\widetilde{S}_a|}{|D|}\right)$

特徵A 值 a 時的資訊增益比為 $G(D,A,a)=\frac{g(D,a)}{H_A(a)}$

計算對於每個特徵由起使值開始進行梯度向量的疊代計算最小資訊增益比時的 a 值作為選擇 分枝的判準以生成決策樹。

2.邏輯回歸（Logistic Regression）: 多變量的二元分類模型。利用各相關因子找出標記的事件發生與否的二元分類。

對於變量$X=（x_1,x_2,x_3,\cdots,x_n）$，找到權數$w_0,w_1,w_2,\cdots,w_n$，$g(X)=w_0+w_1x_1+w_2x_2+\cdots+w_nx_n$，對於標記的事件，例如汙染的發生以應變量$y=1$表示。是一種多變量的二元分類模型。 利用各相關因子找出標記的事件發生與否的二元分類。

3.支持向量機（SVM）: 採用一般的係數乘積，訓練樣本對每個樣本的資料向量進行分類。

採用一般的係數乘積。訓練樣本為(x_i,y_i)，$1\le i\le N$，x_i 是每個樣本的資料向量，y_i是汙染與否為 0 或 1$w^T\cdot x$表示 w 與 x_i 內積，其中 w 是待決定的權重，ξi 是每一樣本的容許誤差成本。

$$\text{minimize}\ \frac{1}{2}\|w\|^2+C\sum_i\xi_i \quad \text{subject to}\ y_i\left(w^Tx_i-b\right)-1+\xi_i\ge 0 \quad \forall i \quad \xi_i\ge 0 \quad \forall i$$

accuracy: 83.45%

	true false	true true	class precision
pred. false	258	19	93.14%
pred. true	51	95	65.07%
class recall	83.50%	83.33%	

表三 人工智慧預測之混淆矩陣

精確率 : 預測為污染的樣本的正確率

招回率 : 預測事實為污染的樣本的正確率

F1-score : 一種平均值，平衡反映演算法的精確度。

用途二

利用ARIMA得出汙染物時段數季節性的時序節律變化

4.差分整合移動平均自回歸模型ARIMA: 時間序列預測分析:以樣本散點圖、自我相關函數判斷其平穩性

ARIMA（p，d，q）模型可以表示為 $\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d X_t=\left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t$

μ 是 Xt 平均數，ϕi 是自回歸參數，θi 是移動平均參數，εt 是誤差量。

用途三

利用K-means得出各測站汙染物分布類型

5. K-平均演算法K - means: 以聚類的方式進行分類:選定分類效益最大時的 K 值，依 K 類中心分組。

一種聚類（Cluster），原始每一筆資料的經緯度分別為 x,y 坐標，記為 $\boldsymbol{xj}$，一開始選定 k 個初始值資料$\mu 1, \mu 2, \ldots, \mu k$，計算 $J=||\boldsymbol{xj}-\mu i||$， i=1,2,...,k，

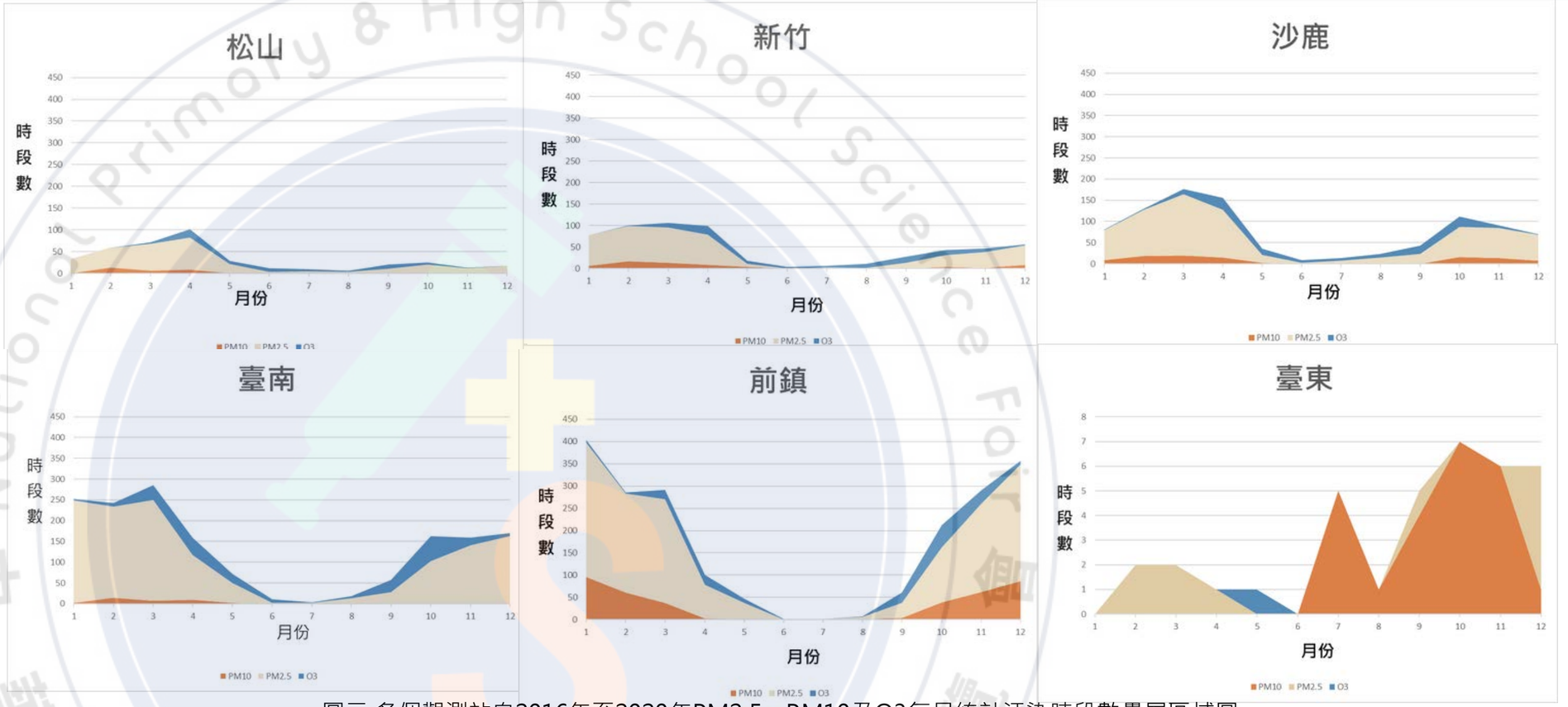
若 J 值最小，則 $\boldsymbol{xj} \in St$，如此原始每一筆資料可分成 $S1, S2, \ldots, Sk$ 類，

再計算 $J=\sum_{i=1}^{k}\sum_{x_j\in S_i}\|x_j-\mu_i\|^2$

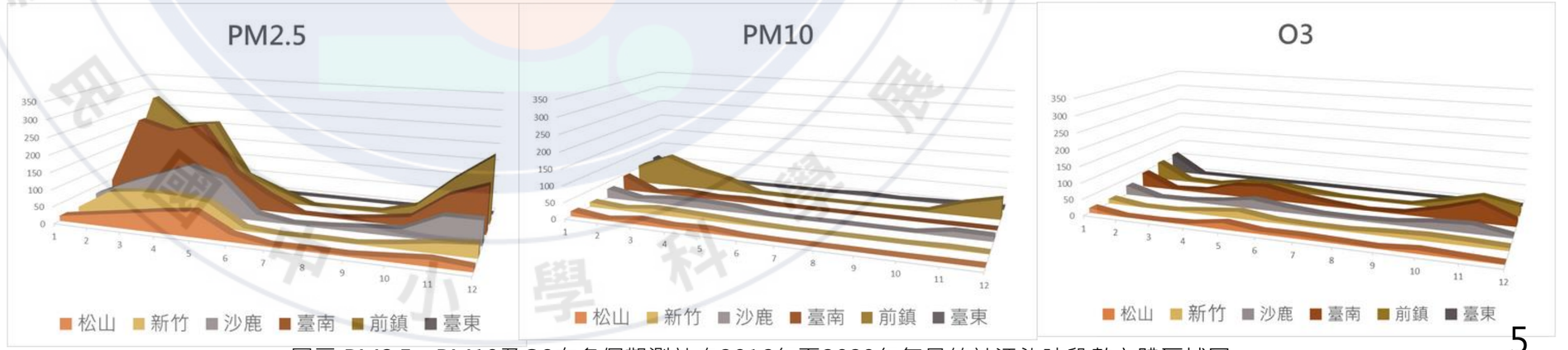
平均中心距離= $\sqrt{\frac{1}{n}\sum_{i=1}^{k}\sum_{x_j\in S_i}\|\boldsymbol{x_j}-\mu_i{}^*\|^2}$

結果一：基本統計

- PM2.5汙染時段數最多
  - PM2.5質量較小，容易長時間在空氣中懸浮
- PM2.5 汙染多發生在 1～4 月；PM10 汙染多發生在 1～5 月；O3 多發生在 3～5 月 及 9～10 月
  - 汙染時段數量多的月份多為東北季風的好發季節
- O3汙染時段數西南部多於東北部
  - 西南部較東北部高的溫度及日照量有助於臭氧的形成
- 2020 年汙染時段數明顯變少
  - 應該是受 COVID - 19 疫情影響，工業及經濟活動減少所致



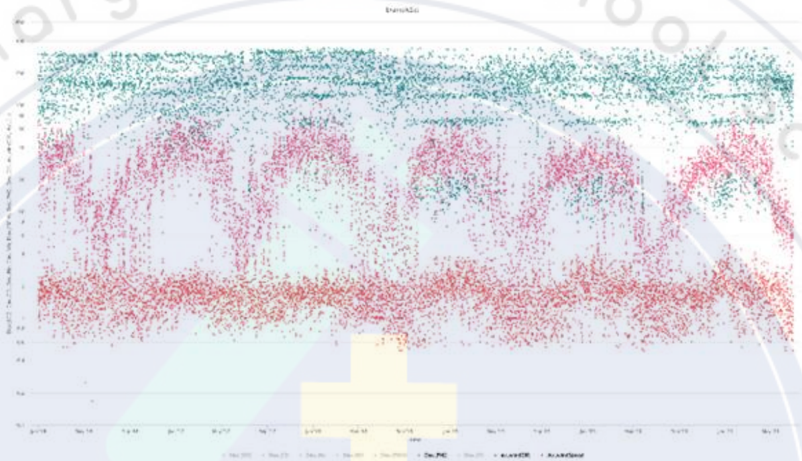
圖二 各個觀測站自2016年至2020年PM2.5、PM10及O3每月統計汙染時段數疊層區域圖



圖三 PM2.5、PM10及O3在各個觀測站自2016年至2020年每月統計汙染時段數立體區域圖

結果二 : 以時序分析 ARIMA 分析各觀測站 PM2.5、PM10 及 O3 的空氣污染型態:

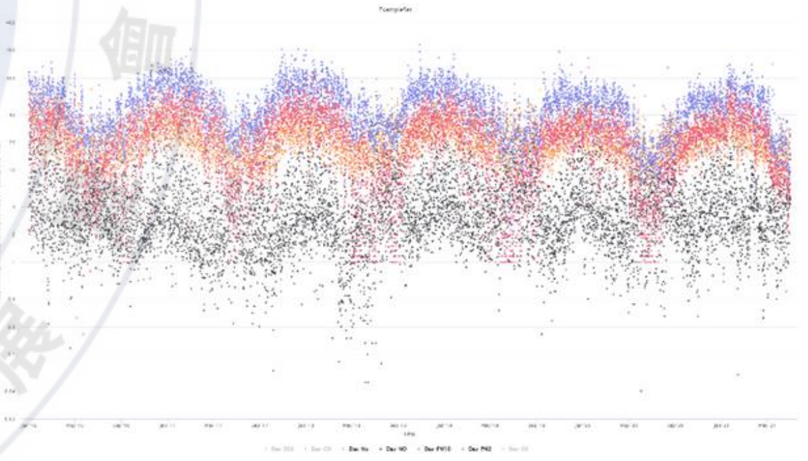
- PM2.5 和 PM10 的時序節律極度相似。O3 的時序節律並不明顯。
- 氣象因子中，風速節律明顯，風向節律不明顯；污染因子中，CO、Nx 和 NO 與 PM2.5 和 PM10 具有相似的節律； SO2 節律較不明顯。
- 松山、新竹、沙鹿 在 PM2.5 及 PM10 的 RMSE 值都優於 台南、前鎮。
- 台東在 PM2.5、PM10、O3 的 RMSE 值皆為極大。
- PM2.5 的 RMSE 值都優於 PM10。



PM2.5、PM10與風速、風向的觀測量的時序變化

PM2.5、PM10與CO、SO2污染物的觀測量的時序變化

PM2.5、PM10、O3污染物的觀測量的時序變化



PM2.5、PM10與NO、Nx污染物的觀測量的時序變化

測站/汙染物	松山	新竹	沙鹿	臺南	前鎮	臺東
PM2	7.074	7.591	8.869	10.502	10.064	極大
PM10	12.203	13.457	15.648	18.906	17.382	極大
O3	極大	極大	極大	極大	極大	極大

表三 各測站ARIMA分析之RMSE值

結果三：透過決策樹、邏輯迴歸、SVM人工智慧方法以汙染物、天氣因子預測

- 前鎮在 PM2.5 上空氣污染物預測結果與氣象因子預測結果 F1–score 值相近。但 PM10 的預測結果有落差。
- 松山以空氣汙染物預測 PM2.5、PM10 的結果為良好，而氣象因子的預測結果很糟且兩 者有明顯落差。
- 新竹、沙鹿以空氣汙染因子在 PM2.5 及 PM10 二種污染物的預測結果優於氣象因子，而 在O3 預測方面兩者有較明顯落差。
- 臺南以空氣汙染物預測 PM2.5 的結果優於氣象因子；PM10的預測結果則為空氣汙染物 和氣象因子相當。
- 臺東以氣象因子預測 PM10 的結果優於空氣汙染物；以空氣汙染物預測 PM2.5 的結 果則優於氣象因子。

PM2.5、PM10的預測-汙染因子

以汙染因子對每個測站的 PM2.5、PM10 進行預測，都有一定效果。

PM2.5、PM10的預測--氣象因子

以氣象因子對每測站的 PM2.5、PM10 進行預測，多沒有良好效果。

O3的預測

以汙染因子對每個測站進行 O3 預測，結果非常好。以氣象因子進行 O3 預測則有時很糟 F1–score 值 0.3898。

但以氣象因子進行 O3 預測的效果較進行 PM2.5、PM10 來得好。

performance							
PM2.5	測站	松山	新竹	沙鹿	臺南	前鎮	臺東
Decision Tree	F1 SCORE	0.759878	0.782353	0.738462	0.786982	0.675768	0.652632
Logistec	F1 SCORE	0.836478	0.843478	0.709898	0.739437	0.708134	0.722772
svm	F1 SCORE	0.802469	0.701493	0.742857	0.779221	0.694444	0.707071
PM10							
Decision Tree	F1 SCORE	0.565445	0.707124	0.782609	0.668675	0.636704	0.586207
Logistec	F1 SCORE	0.650407	0.633588	0.696971	0.600897	0.716279	0.631579
svm	F1 SCORE	0.836478	0.843478	0.709898	0.611872	0.714976	0.672131
O3							
Decision Tree	F1 SCORE	1	1	1	1	0.994819	1
Logistec	F1 SCORE	0.988048	0.963303	0.982143	0.985646	0.984615	1
svm	F1 SCORE	0.943827	0.968615	0.969957	0.949309	0.955224	0.973684
performance(2)							
PM2.5	測站	松山	新竹	沙鹿	臺南	前鎮	臺東
Decision Tree	F1 SCORE	0.643777	0.664975	0.516605	0.637532	0.671011	0.206897
Logistec	F1 SCORE	0.384314	0.546125	0.458498	0.639098	0.58	0.564972
svm	F1 SCORE	0.541806	0.583658	0.527881	0.530435	0.605128	0.543211
PM10							
Decision Tree	F1 SCORE	0.040268	0.683938	0.631579	0.690323	0.594059	0.629635
Logistec	F1 SCORE	0.364583	0.512821	0.412281	0.564593	0.553846	0.739496
svm	F1 SCORE	0.384314	0.546125	0.458498	0.59633	0.553191	0.704762
O3							
Decision Tree	F1 SCORE	0.761246	0.700422	0.733871	0.703704	0.711864	0.263736
Logistec	F1 SCORE	0.778626	0.567143	0.719626	0.703704	0.606742	0.584615
svm	F1 SCORE	0.804272	0.614634	0.690583	0.715596	0.619289	0.544

表四 決策樹、邏輯迴歸、SVM 預測所得 F1–score 值

結果四 : 決定因子再預測

在決策樹、邏輯迴歸、SVM 人工智慧模型預測中，挑選預測結果良好的模型 (AUC 值在 0.88 以上) 分析 PM2.5、PM10 空氣汙染特徵的化學汙染物及天氣因子為何，並挑選出 p 值小於 0.025 者作為決定因子進行再預測。

預測結果中在以汙染因子預測松山、新竹、臺南、前鎮的模型中，平均一氧化碳 Dav_CO 皆有最小的 p 值(0.000)。因此我們從新只以 Dav_CO 預測,得到的混淆矩陣值與原先的比較如下表。僅只以平均一氧化碳 Dav_CO 即有良好的預測結果。

	av_CO						第(1)類資料					
	Decision Tree		Logistec Regration		svm		Decision Tree		Logistec Regration		svm	
測站	新竹		新竹		新竹		新竹		新竹		新竹	
	F	T	F	T	F	T	F	T	F	T	F	T
N	193	8	253	50	257	56	140	12	276	41	247	51
P	83	137	23	95	19	89	62	133	31	194	29	94
精確率	0.6227		0.8051		0.8241		0.6821		0.8622		0.7642	
召回率	0.9448		0.6552		0.6138		0.9172		0.8255		0.6483	
準確率	0.7838		0.8266		0.8219		0.7867		0.8672		0.81	
F1 SCORE	0.7507		0.7224		0.7036		0.7824		0.8435		0.7015	

表五 以空氣汙染因子與以決定因子avCO預測新竹測站之PM2.5、PM10、O3預測結果對照

結果五 : 以 K–means 聚類分析各觀測站 PM2.5 、 PM10 及 O3 的汙染型態

單一測站來看，3 項不同的空品指標分成良好與不良應會有 8 種可能，但分類結果最多只有四大類組，可見 PM2.5、PM10 及 O3 這三項汙染有一定的關聯性存在。且空氣污染型態 O3 污染時多半不會伴隨 PM2.5、PM10 的污染，PM2.5、PM10 幾乎同時發生。

- 松山空氣污染可能的情形是 PM2.5 、PM10 同時或 PM2.5 、O3 同時或僅 O3 污染。
- 新竹與沙鹿空氣污染可能的情形是 PM2.5、PM10 同時或僅 O3 污染。
- 台南空氣污染可能的情形是僅 PM10 或僅 PM2.5 污染。
- 台東空氣污染可能的情形是 PM2.5、PM10 同時或僅 PM10 污染。
- 前鎮空氣污染可能的情形是 PM2.5、PM10 同時或僅 PM10 污染或僅PM2.5 污染。
- 這些觀測站的空氣污染型態 O3 污染時多半不會伴隨 PM2.5、PM10 的污染。

松山												
分類	cluster0	cluster1	cluster2	cluster3	cluster4	cluster5	cluster6	cluster7	cluster8	cluster9	cluster10	cluster11
PM10	20.11	92.39	45.77	35.29	63.57	160.33	30.64	15.64	75.52	62.87	36.65	46.62
O3	14.38	36.59	78.25	54.46	52.23	48.19	35.32	36.70	92.54	18.67	15.24	37.10
PM2.5	11.44	46.50	27.32	19.14	32.90	54.02	15.54	10.15	44.23	35.43	19.15	23.64
新竹												
PM10	22.22	39.65	70.30	18.02	116.94	260.40	36.27	86.66	55.80	46.72	31.53	60.38
O3	16.88	18.54	24.61	37.69	41.08	68.64	57.07	57.12	77.95	38.11	38.32	49.77
PM2.5	10.47	20.76	40.05	9.45	59.67	55.00	18.66	43.52	29.79	22.33	14.58	29.62
沙鹿												
PM10	20.12	102.81	42.25	146.47	36.10	17.83	56.74	69.52	92.16	73.12	33.37	51.86
O3	18.12	37.33	19.54	49.23	57.76	37.45	74.24	24.58	74.14	53.32	36.90	44.24
PM2.5	9.74	54.01	21.96	76.63	17.54	8.67	29.34	38.48	50.29	36.00	16.32	25.66
臺南												
PM10	60.35	20.15	185.34	75.26	55.94	37.00	39.10	92.64	119.06	311.94	129.30	86.94
O3	27.30	23.73	56.46	49.48	71.62	23.25	47.89	28.82	52.73	58.28	49.05	77.76
PM2.5	32.47	10.34	41.13	36.04	30.76	17.97	21.12	56.71	40.22	45.50	86.38	48.96
前鎮												
PM10	93.74	71.41	109.78	122.50	38.16	160.00	281.64	65.04	45.51	61.07	22.25	91.05
O3	52.52	73.53	77.68	32.31	27.19	47.83	55.94	43.70	58.72	18.03	17.57	21.19
PM2.5	46.39	37.40	60.05	71.83	17.12	43.50	71.50	34.90	24.10	31.31	8.47	47.54
臺東												
PM10	53.53	34.36	244.96	143.71	26.17	81.49	42.72	29.58	19.81	666.67	12.72	17.54
O3	47.52	44.02	40.59	36.02	30.46	37.95	28.42	17.08	43.45	24.33	29.91	15.24
PM2.5	21.67	15.28	23.23	22.17	10.79	17.56	14.95	10.29	8.66	29.00	6.77	6.54

表六 k-means聚類PM2.5、PM10、O3結果

討論

汙染因子關聯性

- O3污染多半不會伴隨 PM2.5、PM10汙染的發生；PM2.5、PM10汙染幾乎同時發生。
- 汙染物：CO 、 NO 、 NO2、 Nx、 SO2 進行分類預測，CO、 Nx 和 NO 與 PM2.5 和 PM10 節律相似，而 SO2 則較不明顯。
- 天氣因子 : 風向、風速、相對濕度、溫度進行分類預測，風向、風速是關聯性大的重要因子，其他因子與 PM2.5 和 PM10 污染則沒什麼關聯。
- O3 汙染多為偶發性所以節律不明顯。因為光化反應，南部日照較長，汙染較嚴重，新竹、松山及台東等測站則幾乎沒有汙染。

各測站汙染模式

- 松山、新竹、沙鹿、台南、前鎮等測站皆有來自本地排放的汙染，台東測站則受外來汙染影響大。
- 地形因素：平原地形無阻擋，汙染物容易受自外界擴散而來，如台南、前鎮；盆地地形會使汙染物聚積、不易擴散，受本地的影響增加，如松山、沙鹿。
- 工業發展：重工業的進駐，會加重本地污染的情況，使當地汙染受本地排放的影響增加，如沙鹿、前鎮。

結論

（一）、PM2.5、PM10 及 O3 這三項汙染有一定的關聯性存在。且空氣污染型態 O3 污染時序多半不會伴隨 PM2.5、PM10 的污染；PM2.5、PM10 幾乎同時發生。

（二）、空氣污染物 CO 、 NO 、 NO2、 Nx、 SO2 進行預測分類,不論是季節節律、模型預測中 CO、Nx 和 NO 與 PM2.5 和 PM10 節律相似關聯大，而 SO2 則較不明顯。

（三）、單純以天氣因子:風向、風速、相對濕度、溫度、雨量進行分類預測,風向、風速及溫度是明顯重要的因子，其他因子與 PM2.5 和 PM10 污染沒什麼關聯。

（四）、O3 汙染多為偶發性所以節律不明顯。因為光化反應,南部日照較長，因此 O3 汙染程度 : 台南 > 前鎮 > 沙鹿，而新竹、松山及台東則很少。

（五）、新竹、台南、前鎮、沙鹿、前鎮的 PM2.5、PM10 皆以汙染因子預測較佳，其中僅以平均一氧化碳 Dav_CO 即有良好的預測結果。台東的 PM2.5 以汙染因子預測較佳，PM10則以氣象因子預測較佳。台東汙染較不具季節差異性，多為偶發且較輕微。

未來展望

人工智慧有別於一般統計及回歸直線分析趨勢，它能更精確且多方面得出數據的 時序變化、與其他因子的相互關聯性，我們希望在將來

（一）將時間因子引進kmeans分析，詳細探討時間序列變化與汙染源的關係，現階段 只有進行汙染源的聚類分析，無法得知汙染源聚類對應的時間

（二）利用人工智慧得出預測氣象類因子跟污染類因子的 p、z 值，並找出其決定性 的因子。

（三）再利用深度學習（Deep Learning）計算（二）找出的決定性因子。

（四）調查各汙染源排放的汙染物種類，藉此更詳細了解各測站周邊地區的汙染物情 形，進行污染的管制以及預防。

參考文獻

【1】葉惠中主持(2014)。細懸浮微粒(PM2.5)時空分佈特性之研究(102 年度「環保署/國科會空污防制科研合作計畫」期末報告,NSC102–EPA–F–005–001)。中國文化大學土地資源學系。

【2】蔡俊鴻主持(2008)。超細氣懸微粒水溶性無機離子成份特性變異與前驅物及氣因子關聯性研究(行政院國家科學委員會補助專題研究計畫成果報告,NSC96–2211–E–006–024)。國立成功大學環境工程學系(所)。

【3】李佾、徐如怡、沈姿妤（2020）。走出陰霾 - 台灣大氣懸浮微粒含量變化與天氣狀況、汙染物種類特徵探討（中華民國第 60 屆中小學科學展覽會作品說明書）。桃園市立內壢高級中等學校。