

# 以少數類別過抽樣技術提升太陽能供電異常偵測—以微型空品監測系統為例

汪諭<sup>1</sup> 林良憲<sup>2\*</sup> 陳祉好<sup>1</sup> 劉丞偉<sup>3</sup>

## 摘要

二氧化碳(CO<sub>2</sub>)的大量排放是全球暖化和生態破壞的重要原因。火力發電為臺灣主要的發電方式，然而火力發電所排放之空氣汙染物及CO<sub>2</sub>對於人體健康及環境造成危害，因此臺灣已實施能源轉型，減少燃煤、增加燃氣及發展綠能。隨著物聯網(Internet of Things)逐漸普及，臺灣企業的用電需求也隨之增加。一些研究建議以太陽能作為物聯網系統的主要供電能源。然而由於不確定的氣候變化，以太陽能供電為基礎的物聯網頻繁出現低電壓與電壓中斷等異常事件。在此情境下，在本研究，我們發展一個以太陽能供電為基礎的微型空品感測器系統，使用訊息佇列遙測傳輸協定(MQTT)傳遞感測資料至在樹莓派(Raspberry Pi)上的SQLite資料庫。此外，本研究以少數類別過抽樣技術(SMOTE)改善支持向量機(SVM)模型對不平衡分布之正常與異常事件分類能力。最後，我們也使用Node-RED工具開發網頁介面呈現即時感測器資料。

**關鍵詞：**太陽能供電，溫室效應，異常事件，即時感測資料

## 1. 背景

### 1.1 燃煤發電污染

因應能源轉型政策，我國近幾年以減少燃煤、增加燃氣及發展綠能，進而使全球溫室氣體減量為目標。台灣電力公司指出110年電力總發電為2,488.1億度，其中發購電能源包含核能、燃煤、燃氣、燃油、汽電共生、再生能源與抽蓄水力七種電力來源，而火力發電占比高達79.6%，為現階段最主要的電力來源；燃氣發電占比為52.8%；燃煤為44.6% (台灣電力公司，2022a)。另外，根據我國環保署2019年統計資料顯示，排放之溫室氣體中二氧化碳(CO<sub>2</sub>)

約占95.28%，其中90.13%來自於能源燃料燃燒 (行政院環境保護署，2022a)。Agrawal (2005)研究指出，火力發電以化石作為燃料，該發電方式會產生二氧化碳(CO<sub>2</sub>)、懸浮微粒(PM)、氮氧化物(NO<sub>x</sub>)以及硫氧化物(SO<sub>x</sub>)等污染物，其中二氧化碳(CO<sub>2</sub>)為造成空氣汙染與溫室效應之主要因素。而大氣中的溫室氣體主要包括二氧化碳(CO<sub>2</sub>)、甲烷(CH<sub>4</sub>)、氧化亞氮(N<sub>2</sub>O)、氟氯碳化物(CFCs)與臭氧(O<sub>3</sub>)。當濃度累積過高時，造成地球必須重新調整熱平衡，進而出現顯著的氣候變遷，導致極端氣候形成，使得地球溫室效應逐漸加劇。

由於二氧化碳(CO<sub>2</sub>)和空氣汙染物沒有特殊的顏色及味道，因此民眾容易將其忽略。懸

<sup>1</sup>國立臺北護理健康大學資訊管理系 大專生

<sup>2</sup>國立臺北護理健康大學資訊管理系 助理教授

<sup>3</sup>工業技術研究院綠能與環境研究所 資深研究員

\*通訊作者，電話：02-28227101#1234，電郵：[lianghsien@ntunhs.edu.tw](mailto:lianghsien@ntunhs.edu.tw)

收到日期: 2022年08月31日

修正日期: 2022年11月11日

接受日期: 2022年11月23日

浮微粒(PM)等空汙氣體體積小，可以透過呼吸作用自肺部滲透到血液中，並引起全身性發炎反應或其他心血管疾病的發生(Donaldson *et al.*, 2005)。然而二氧化碳(CO<sub>2</sub>)不同於空氣汙染物對人體造成直接性損害，而是影響到人類生存的環境，二氧化碳(CO<sub>2</sub>)的大量排放已成為全球暖化和生態破壞的重要原因(Bai *et al.*, 2020)，造成全球氣候變遷，隨著地球大氣溫度的升高，導致極地冰原融化速度加快、海平面上升與降水量增加等直接或間接地影響生物多樣性而造成不可挽回的損害(Bastianoni *et al.*, 2004；Patil *et al.*, 2021)。

## 1.2 淨零排放與再生能源

2015年聯合國所簽訂的《巴黎協定》，致力於推動減碳政策，將全球氣溫升幅控制在攝氏2度內，並提出「淨零碳排」一詞。關於溫室效應所造成的環境問題已經成為國際關注之議題。我國環保署為了有效管制國內碳排放量，對國內企業設定每年排放額度上限，若超過核配額度，則須繳納額外費用。臺灣企業配合政府政策進行能源轉型，如台積電(TSMC)、宏碁集團(Acer)與華碩(ASUS)等接續參加RE100，為氣候組織(The Climate Group)與碳揭露計畫(Carbon Disclosure Project, CDP)主導的全球再生能源倡議，以電力需求端的角度，共同努力提升使用綠能的友善環境，參與之企業皆公開承諾在2020年至2050年間達成100%使用綠能，並逐年提報用電數據。政府亦鼓勵民眾透過改變生活模式來減少碳排放，如使用大眾運輸、選購節能家電與親近再生能源等。

我國電力來源以化石燃料為主要，此發電方式容易產生嚴重的碳排放等空氣汙染。對此，學者Martins *et al.* (2019)研究指出歐洲國家過於依賴化石燃料發電而造成負面環境影響，且於未來化石燃料不足情況下，勢必以發展再生能源為首要任務。臺灣的再生能源包括水力、太陽能、風力、生質能、廢棄物及地熱發電六個種類，根據台灣電力公司報告(台灣電

力公司, 2022b)指出統計110年再生能源發電量為174.2億度，其中太陽能發電量為79.6億度，於所有再生能源種類中占比最高，由此可知，太陽能發電已經成為臺灣再生能源的主流，與其他再生能源相比，太陽能之成本效益較高，且能長期取之不盡(Ryabushev, 2022)，於可用性 & 效率方面具有優勢(Kannan and Vakeesan, 2016)。

## 1.3 太陽能為供電基礎之物聯網

近年，物聯網(Internet of Things, IoT)技術在工業與機械領域皆受到廣泛的應用，系統長期以24小時運作需要消耗大量的電力。如我國環保署與臺灣16個地方政府合辦布建覆蓋282個區級行政區與111個工業區及科學園區，負責感知工廠空汙概況的10,000個空汙感測器，進而導致耗電費大幅增加(行政院環境保護署, 2022b)。為了節省電力成本，許多研究提出可以使用再生能源作為感測器之主要供電來源(Eltamaly *et al.*, 2021；Ichikawa *et al.*, 2019；Jack *et al.*, 2020)，如Ichikawa *et al.* (2019)建議以太陽能做為物聯網供電來源。但太陽能屬於間歇性能源，易受氣候及光照程度等環境因素影響，具有不穩定及不可控之特性。根據Sinsel *et al.* (2020)研究指出太陽能之不可控特性勢必衝擊到網絡上之設備運轉的穩定性，而產生供電的疑慮。Glavic and Van Cutsem (2011)研究指出長期的供電不穩定將提高設備損害率。在此情境下，以太陽能為供電基礎的物聯網系統容易遭遇供電不穩情況，導致感測器失效而遺失重要資訊。供電不穩之系統經常發生如低電壓事件或是電壓中斷事件等，這些都會造成感測器的壽命減短或是偵測性能降低(Sinsel *et al.*, 2020)。

Hoiem *et al.* (2020)收集系統上感測器用電特徵，透過Support Vector Machine (SVM)模型來偵測以太陽能供電異常事件。然而，當正常事件與異常事件的資料數量比例不平衡時，使用SVM模型無法對異常事件取得較佳的分類結

果。為了解決類別分布不均學習問題，我們採用現今被廣泛使用的演算法Synthetic Minority Oversampling Technique (SMOTE)過採樣方法將新樣本添加到少數別類中以平衡訓練資料集，用於改善SVM對太陽能供電異常事件之分類。

## 1.4 研究動機

本研究模擬一個以太陽能作為主要供電能源之微型空品感測器系統。系統透過太陽能穩壓充電版提供電源給Arduino Uno板以及五個低成本感測器，如溫濕度感測器(DHT11)、二氧化碳濃度感測器(SGP30)、空氣品質(PM<sub>2.5</sub>)感測器(GP2Y1014AU)、光源感測器(TW7706)以及電壓電流感測器(INA219)。將感測資料透過Message Queuing Telemetry Transport (MQTT)通訊協定技術傳至在Raspberry Pi 4上的資料庫SQLite。系統更進一步將分析結果與偵測資料透過Node-RED視覺開發工具呈現於網頁上。在模擬的系統上，透過SVM分類演算法偵測系統用電是否出現低電壓或電壓中斷等異常事件，並採用SMOTE來改善SVM模型對異常事件分類性能。此外，計算提出的方法在三個分類項目上的結果以驗證提出的方法之有效性。實驗結果顯示，SMOTE方法比另外兩種方法可以達到更好的異常事件分類正確性。

本研究其餘部分如：第二章詳細說明所提出的系統、MQTT傳輸協定、SMOTE演算法以及SVM機器學習模型；第三章提供了平台網頁、異常事件性能評估之實驗結果；第四章為結論與未來研究。

## 2. 太陽能供電之微型空品監測系統

於此章節，我們詳細地說明本研究建構一個以太陽能充電板為基礎的IoT系統。透過MQTT通訊協定儲存感測資料至SQLite資料庫，並將感測資料以網頁呈現於用戶端。

### 2.1 系統物聯網架構

本研究的IoT架構呈現如圖1，以太陽能作為主要供應能源，鋰電池作為輔助供電設備。透過Lipo Rider Pro太陽能充電板提供5V電源給Arduino Uno板及五個感測器進行數據收集，包含DHT11溫濕度感測器、SGP30之二氧化碳濃度感測器、GP2Y1014AU之PM2.5感測器、Grove Sunlight光源感測器及INA219電壓電流感測器。系統的感測資料透過MQTT傳送至Raspberry Pi 4之SQLite資料庫。

依此架構，本研究使用的感測器與材料，

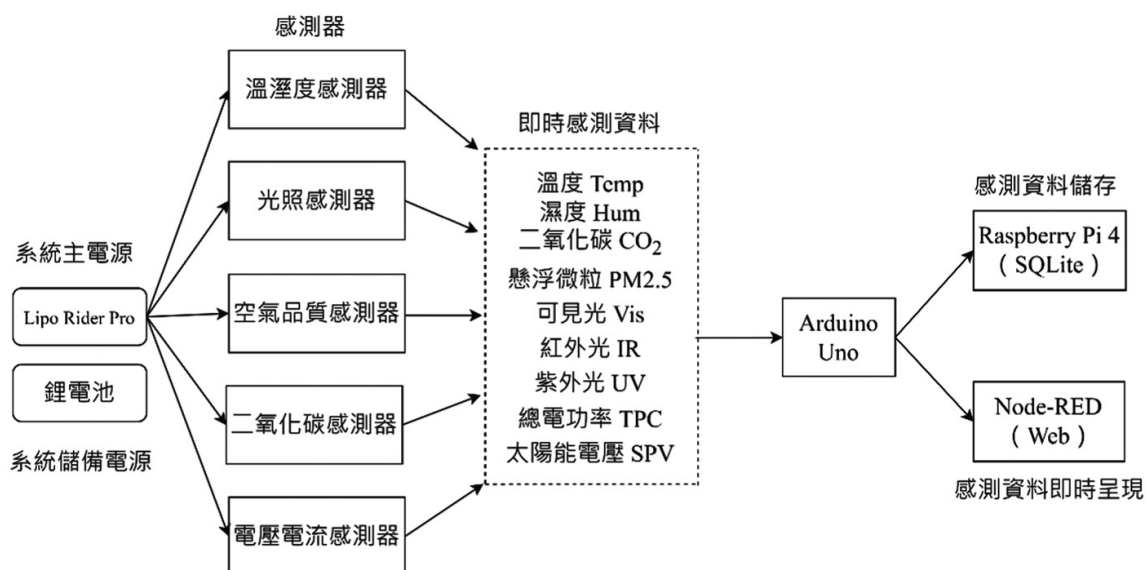


圖1 本研究IoT架構



說明如下：

■ Arduino Uno板：微型控制器。

■ Lipo Rider Pro太陽能充電板：

Lipo Rider Pro是太陽能鋰電池充電板(輸出穩壓專業板)，透過USB銜接到設備，具有穩定的輸出5V/1A供電以及提供四個LED指示燈，用以提示目前鋰電池之電量與充電狀態。Lipo Rider Pro的穩定輸出電壓為5V/1A，當監測到輸出電壓低於5V時，鋰電池的電力會被直接切換為主供應電源。

■ 溫濕度感測器(型號：DHT-11)：

DHT-11溫濕度感測器是一款經過校準過且直接以數位訊號輸出之溫濕度感測器。溫度及濕度的檢測範圍為0°C至50°C，工作電壓為3.3V~5V直流電。

■ 二氧化碳濃度感測器(型號：SGP30)：

SGP30為一款配備I2C接口的室內空氣質量傳感器。輸出的二氧化碳濃度以ppm為單位。具有高穩定性，具有I2C接口和完全校準的輸出信號，測量值內的典型精度為15%。

■ Sharp空氣品質感測器(型號：GP2Y1014AU)：

是一種光學空氣品質感測器，其內部安裝紅外線發光二極體與光電晶體，使其能夠探測空氣中塵埃粒子之反射光，經常運用在空氣淨化系統中。

■ Sunlight Sensor光源感測器(型號：TW7706)：

為多通道數字光線傳感器，具有檢測紫外光，可見光和紅外光的能力。

■ 電壓電流監控感測器(型號：INA219)：

INA219為電壓電流監控感測器可同時監控分流電阻器上的電壓降以及供應電源之電功率並直接以毫安培(mA)為單位讀取電流，以毫瓦特(mW)為讀取電源之單位。

■ Raspberry Pi 4：

英國樹莓派基金會(Raspberry Pi Foundation)所開發的SoC開發版Raspberry Pi 4。CPU為Cortex-A72，採用四核處理器的晶片，記憶體為8GB的LPDDR4 SDRAM，提供40根I/O腳位，體積僅為85.6\*56.5 mm ("Raspberry

Pi")。佈署於Raspberry Pi 4上的作業系統為Ubuntu 20.04，並以該開發版作為智能閘道(smart gateway)取得感測數據，最後將數據傳輸至雲端伺服器紀錄。

## 2.2 訊息佇列遙測傳輸協定(MQTT)

1999年Andy stanford-Clark及Eurotech的Arlen Nipper博士提出訊息佇列遙測傳輸(Message Queuing Telemetry Transport, MQTT) (*The MQTT protocol*, 2012)，為一種輕量級的發布/訂閱消息傳遞協議。其採用星型拓撲(Star Topology)的簡單Publisher/ Subscriber結構，透過代理者(Broker)做為溝通橋樑，以最大限度地減少網路頻寬及設備資源的需求，且能於低功耗與少量儲存空間之設備上輕鬆實現(Sanjuan *et al.*, 2020)。

MQTT中具有三個角色(Role)，發布者(Publisher)負責將訊息分為不同的主題(Topic)，且決定是否要保留訊息(Retain)；代理者(Broker)負責角色之間的消息交換，轉發來自發布者(Publisher)的主題(Topic)以及負責網絡中所有角色的身分驗證；訂閱者(Subscriber)負責接收被訂閱的主題(Topic)。其中，用戶端(Client)的發布者(Publisher)會向代理者(Broker)發布特定的主題(Topic)，並且以代理者(Broker)為伺服器端(Server)作為此通訊協定之中繼站。用戶端(Client)間只需透過主題(Topic)來取得所需資訊，無須取得對方的IP位址。

本研究以Arduino Uno微型控制器擔任發布者(Publisher)傳送DHT-11、UV-TW7706 Sensor、INA219和SGP30等感測器資料至MQTT服務代理者(Broker)，並以Raspberry Pi 4與Node-RED作為訂閱者(Subscriber)將感測資料分別存放於資料庫SQLite與即時呈現於網頁上，本研究之MQTT架構如圖2所示。

## 2.3 系統供電異常偵測

面對系統供電不平衡類別之感測資料，本研究應用SMOTE技術處理訓練資料集之類別不

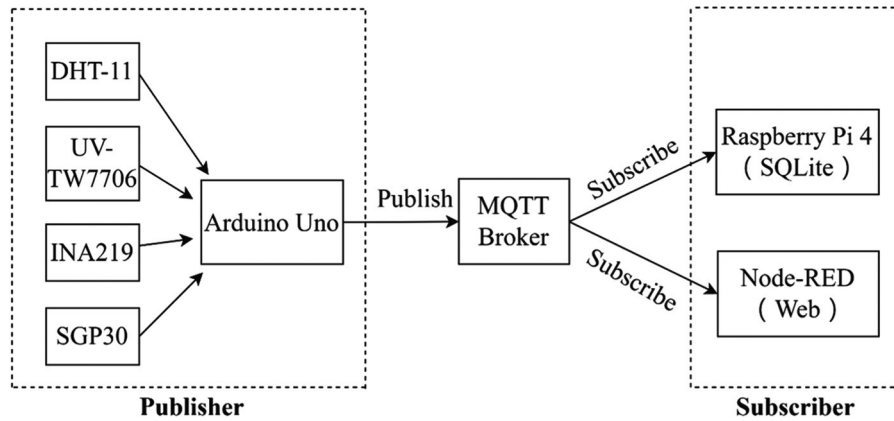


圖2 本研究MQTT架構

平衡問題，以及使用SVM分類模型偵測以太陽能為供電基礎的IoT系統之供電異常事件。

### 2.3.1 隨機過抽樣技術(ROS)

Menardi *et al.* (2014) 提出一種隨機過抽樣技術(Random Over-sampling Examples, ROS)，此方法為最容易使用的過採樣技術。該方法對於少數類樣本進行隨機抽取，將其複製後加入數據集當中。然而，抽取原始樣本作為新樣本的方式可能導致學習模型產生過度擬合(Over-fitting)之問題。

### 2.3.2 少數類別過抽樣技術(SMOTE)

為了避免同樣的樣本被過度學習而產生過度擬合問題，Chawla *et al.* (2002)提出一種抽樣方法SMOTE演算法，對少數類樣本進行人工合成產生新樣本新增至原始資料集。SMOTE方法是透過直線內插方式找出相似於原始樣本的人工樣本。SMOTE採用直線內插生成新的少數類別樣本，其中 $\delta$ 為0到1的隨機數值，如公式(1)。

$$x_{\text{synthetic}} = x_{\text{original}} + (x_{\text{nearest}} - x_{\text{original}}) \times \delta \quad (1)$$

生成過程會隨機選擇一少數樣本點( $x_i$ )，並計算該點與其他少數類樣本之距離，接著根據設定之採樣比例N以及近鄰值k，從近鄰樣本中選取k個樣本點。透過增加少數類樣本的數量，達到資料集的平衡並幫助分類器提高泛化能力

(Chawla *et al.*, 2002)。為了詳細說明SMOTE方法，我們以四個鄰近點為例，選擇樣本點( $x_i$ )延伸連接至4個樣本點( $x_{i1} - x_{i4}$ )，再從4個少數類樣本隨機選擇並帶入公式(1)，最後生成4個合成樣本點( $n_1 - n_4$ )，如圖3。

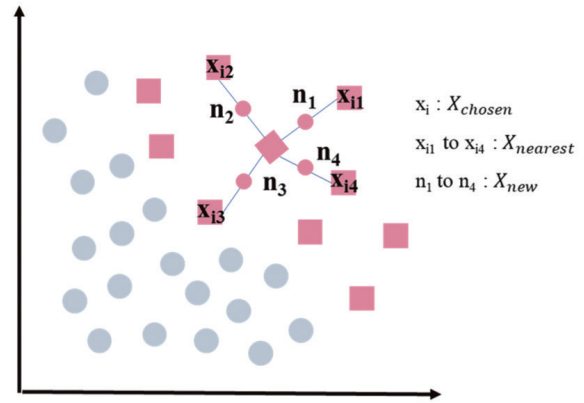


圖3 SMOTE生成樣本

依據SMOTE方法的公式(1)，我們可於原始樣本鄰近點間生成人工樣本，而這些人工樣本限制生成於資料範圍內而不會變成離群值，因此可有效的改善分類器的學習正確性。更進一步地，透過SMOTE技術，少數類別的數量被增加到與多數類別相同，因而平衡資料類別分佈。應用SMOTE改善這種類型資料的研究如：針對天氣不穩定下旋風預測之不平衡資料集，Novitasari *et al.* (2020)應用SMOTE技術改善SVM分類器的預測正確性以及Al Kharusi *et al.* (2022)應用該SMOTE技術改善機器模型對輸電線路故障分類的正確性。

### 2.3.3 支持向量機(SVM)

1995年貝爾實驗室Vapnik等人以統計學習理論基礎提出SVM(Cortes and Vapnik, 1995)，為一種監督式分類與迴歸分析之學習模型演算法(Mahesh, 2020)。基於最小化誤差來訓練模型(Zhang, 2012)，在建構超平面(Hyperplane)之複雜度和訓練誤差之間取捨(Trade-off)，並降低過度擬合(Over-fitting)的機率。透過微調分類器獲得訓練數據集中的最大準確度(Pisner and Schnyer, 2020)。應用決策函式Kernel與超平面(Hyperplane)將各類別資料進行分類，透過找尋最大邊際(Margin)，也就是資料集之數據所相距的最大距離，使超平面(Hyperplane)達到最佳化。而落在該邊際落之數據則稱為支援向量(Support vector)，為提供資訊的主要來源。假設樣本有兩類，分別以 $y_i = 1$ 及 $y_i = -1$ 表示。參數 $x_i$ 為樣本點( $i = 1, 2, \dots, M$ )， $M$ 為樣本的數量， $w$ 代表向量， $b$ 代表常數，用於定義分隔超平面(Hyperplane)的位置。若資料屬於線性關係，則以 $f(x) = 0$ 表示，如公式(2)。

$$f(x) = w^T x + b = \sum_{i=1}^M w_i x_i + b = 0 \quad (2)$$

不同的分隔超平面(Hyperplane)應該滿足約束，如式子(3)與式子(4)

$$f(x_i) = 1 \text{ if } y_i = 1 \quad (3)$$

$$f(x_i) = -1 \text{ if } y_i = -1 \quad (4)$$

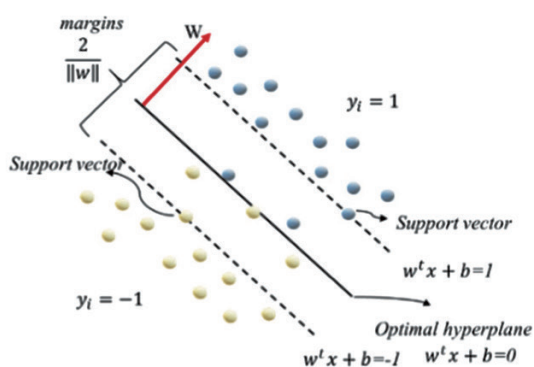


圖4 核函數轉換原始資料至高維度空間

當資料為低維度空間之無法線性分類數據時，則需透過核函數(Kernel function) (Widodo and Yang, 2007)進行非線性投影( $\phi$ )，將資料轉換至更高維度的空間(Noble, 2006)。資料由核函數(Kernel function) ( $\Phi(x_i), \Phi(x_j)$ )轉換至高維度的特徵空間中，使資料之分散程度提升，如圖4。

### 2.4 Node-RED視覺化開發

本研究將即時數據與分析後之結果，透過Node-RED視覺化開發工具轉換成圖表(Dashboard)並呈現於平台上。本研究使用的Node-RED由IBM Emerging Technology於2013年所研發，以Node.js開放原始碼為基底架構的視覺化開發工具。使用流程基礎程式設計(Flow-based Programming, FBP)的資料流(Dataflow)來建立視覺化流程的程式。將硬體設備與應用程式接口(Application Programming Interface, API)相連，將感測相關數據快速整合並產出圖表(Dashboard)，且與網路通訊協定進行資料交換，進而組成一個物聯網(Heath, 2014; Lewis, 2016)。本研究設計的資料流程如圖5。

## 3. 實驗結果與討論

此章節我們將說明實驗資料集、輸入特徵之選取、正規化及學習模型參數設定，以及介紹本研究開發之即時感測數據介面，最後為分



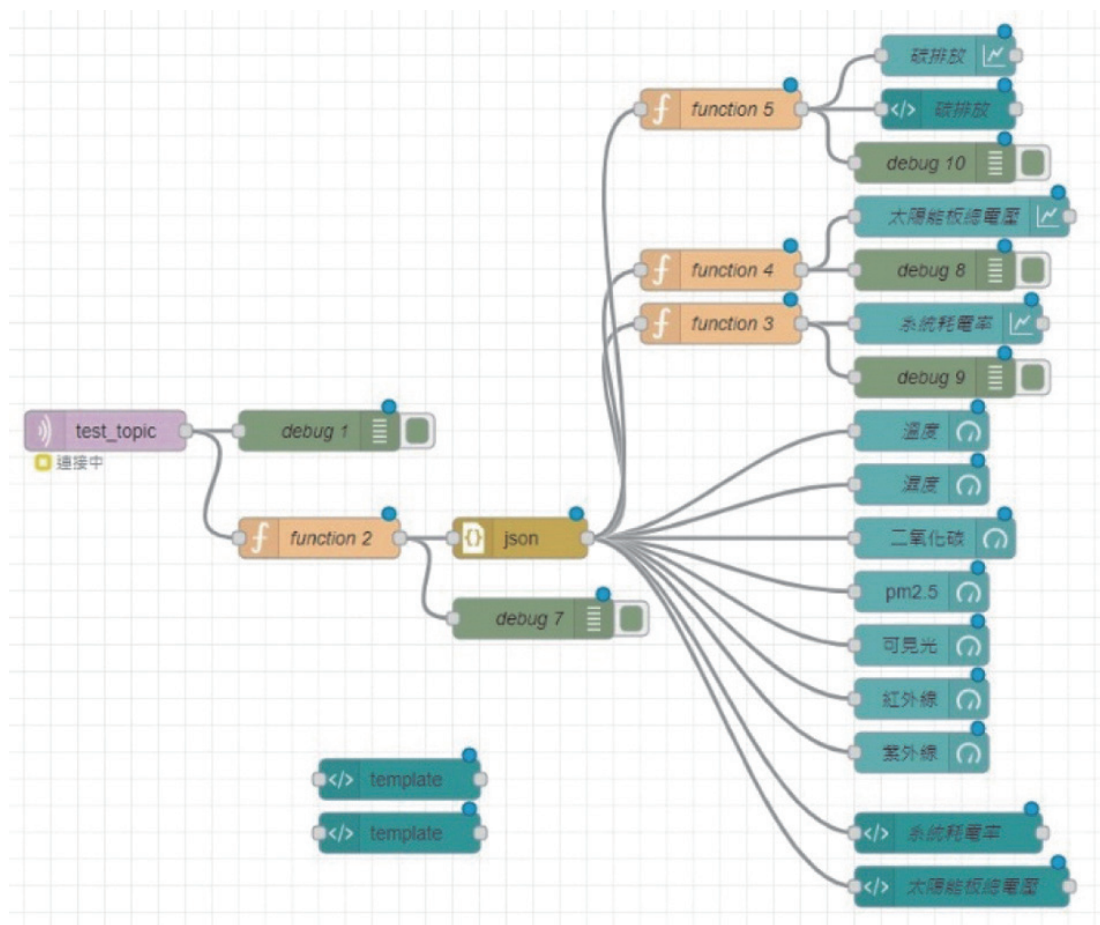


圖5 Node-RED流程

類性能評估指標。

### 3.1 感測資料特徵選取

本研究以感測器數據作為建構支持向量機(SVM)之輸入特徵，用於偵測太陽能供電之異常事件。提取的溫度(Temp)、濕度(Hum)、二氧化碳濃度(CO<sub>2</sub>)、懸浮微粒濃度(PM<sub>2.5</sub>)、可見光(VIS)、紅外光(IR)、紫外光(UV)、全部感測器之電功率(Total Power Consumption, TPC)和太陽能板之電壓(Solar Power Voltage, SPV)為SVM模型之輸入特徵。於每五分鐘計算一次感測器數據的平均值，如表1。並將太陽能供應電源之狀態定義為兩類別：以class 0表示太陽能供電穩定為正常事件，並以class 1表示低電壓或電壓中斷事件為異常事件。實驗資料收集時間為2022年8月2日11:56:20至14:20:40、8月3日10:02:00至15:08:40、8月8日10:15:40

至10:26:20、8月9日11:35:40至13:19:00、8月12日10:38:20至11:32:00、8月22日10:25:20至14:15:40、8月23日12:24:00至15:37:00及8月25日09:43:00至11:34:40。本研究提出的系統設定資料取樣頻率為20秒，並以五分鐘感測資料之平均值作為一筆資料。在資料收集期間，因陽光照射角度問題使太陽能板無法穩定地供電，而導致系統處於低電壓狀況。這個狀況將造成系統發生電力不足而無法啟動感測器，因此收集的資料於某些時段缺少感測資料。最後，本研究共蒐集212筆資料整理如表1。另外，Arduino Uno微型控制器的工作電壓為5V，在這五分鐘內若太陽能SPV輸出電壓出現低於5V時，代表系統處於低電壓或電壓中斷狀況，則設定為異常事件(class=1)；反之，SPV數值大於等於5V，則設定為正常事件(class=0)。以此設定為基礎，我們可取得20個樣本為異常事件類

表1 感測資料

id	date	輸入特徵									輸出
		Temp	Hum	CO <sub>2</sub>	PM <sub>2.5</sub>	VIS	IR	UV	TPC	SPV	class
異常事件樣本											
1	2022/8/2 11:55	23.34	57.00	360.78	3334.11	269.78	295.11	0.00	199.42	5.91	1
2	2022/8/2 14:00	27.14	71.22	414.78	3253.89	267.00	285.00	0.00	184.96	0.86	1
3	2022/8/2 14:15	27.63	64.50	300.00	3341.50	264.25	276.50	0.00	178.50	1.05	1
4	2022/8/3 10:00	26.12	67.50	613.33	3397.33	1225.67	8803.83	4.67	199.27	8.56	1
5	2022/8/3 10:10	32.84	53.38	350.00	3273.75	746.88	4372.50	2.13	164.68	8.72	1
...	...	...	...	...	...	...	...	...	...	...	...
20	2022/8/25 10:15	22.70	62.00	200.00	3568.00	268.00	286.50	0.00	218.40	0.00	1
正常事件樣本											
1	2022/8/2 12:00	25.03	68.00	509.92	3047.92	265.50	291.17	0.00	172.55	6.15	0
2	2022/8/2 12:05	25.98	67.85	533.08	3050.00	266.69	288.85	0.00	172.63	6.06	0
3	2022/8/2 12:10	26.37	68.08	521.25	3032.08	267.17	289.83	0.00	173.95	5.97	0
4	2022/8/2 12:15	26.51	68.67	491.25	3033.75	267.00	288.58	0.00	173.37	5.89	0
5	2022/8/2 12:20	26.78	68.00	459.42	3038.58	266.83	286.75	0.00	172.43	5.74	0
...	...	...	...	...	...	...	...	...	...	...	...
192	2022/8/25 11:30	22.60	55.83	400.00	609.75	264.83	272.83	0.00	194.13	3.20	0

別，以及192個樣本為正常事件類別。

### 3.2 即時感測數據介面

本研究之網頁圖表(Dashboard)數據從邊緣運算裝置(Raspberry Pi)裡取得，透過Node-RED、CSS與JavaScript工具開發網頁介面用於查看即時數據，如圖6。本研究開發之即時感測資料Dashboard介面可提供給與本地管理員相同網域之使用者查閱。本研究開發之平台上以顏色區分感測器之數值範圍。在提出的系統上，我們分別以綠色、黃色、紅色等三種顏色標示各感測器數值區間如表2。在表2，我們列出溫

度、濕度、二氧化碳濃度、PM<sub>2.5</sub>、紅外線、可見光、紫外線、太陽能板即時總電壓以及系統即時耗電功率等各感測器數值範圍。例如，二氧化碳濃度為508 ppm、PM<sub>2.5</sub>為80 µg/m<sup>3</sup>、紅外線為1,261 nm、可見光為462 nm、系統耗電功率203 mW。

### 3.3 異常事件偵測模型

本研究使用Python 3.10.1執行機器學習模型之建模實驗，使用套件如矩陣計算Numpy 1.20.0、資料前處理Pandas 1.4.3及建構模型Scikit-learn 1.1.2等模組。更進一步地介紹用於



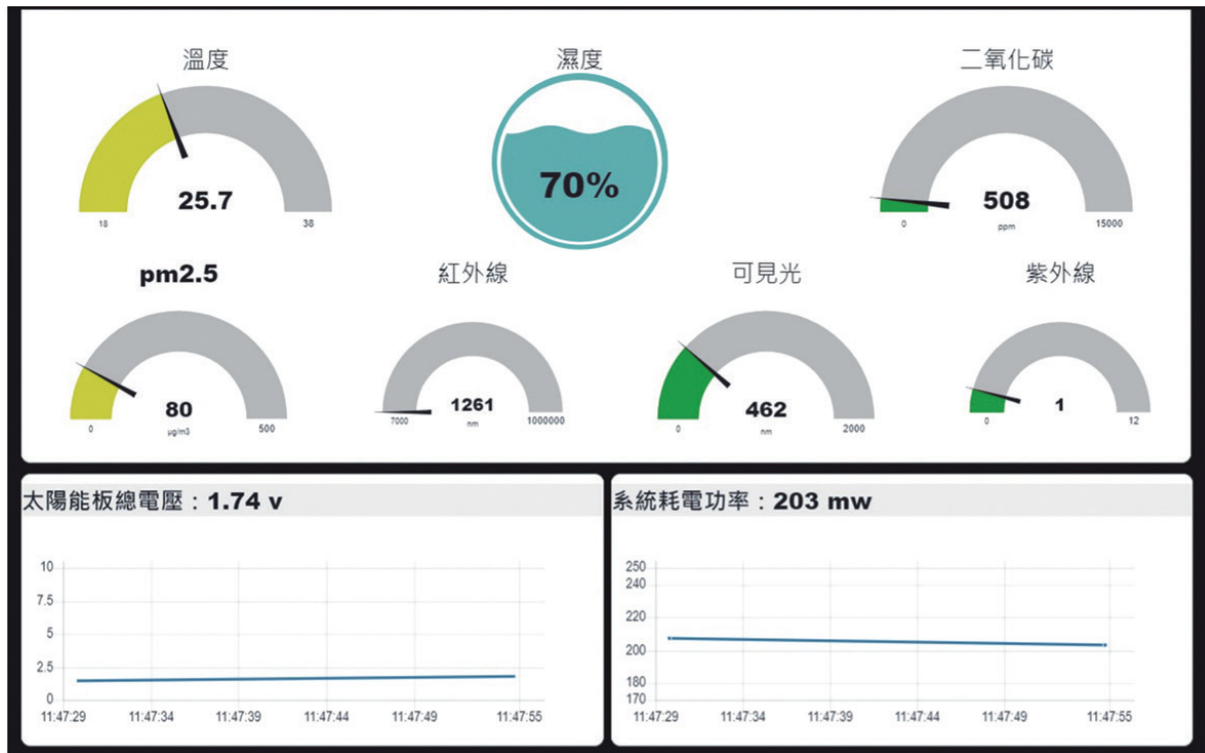


圖6 即時感測資料

表2 區間顏色對照表

種類	綠色	黃色	紅色
溫度	0~18 (度)	18~26 (度)	26~38 (度)
二氧化碳	0~850 (ppm)	850~2,500 (ppm)	2,500~15,000 (ppm)
PM <sub>2.5</sub>	0~35 (µg/m <sup>3</sup> )	35~150 (µg/m <sup>3</sup> )	150~500 (µg/m <sup>3</sup> )
紫外線	0~5 (nm)	5~8 (nm)	8~11 (nm)
紅外線	IR-A	IR-B	IR-C
	700 (nm)~1,400 (nm)	1,400 (nm)~3,000,000 (nm)	3,000,000 (nm)~1,000,000 (nm)

\*可見光(波長)360 nm~400 nm / 760 nm~830 nm。

評估不平衡類別異常事件偵測性能之三個指標如G-mean、F1與IBA。

### 3.3.1 訓練資料正規化

最大值最小值正規化(Min-Max Normalization)是一種線性轉換(Linear Transformation)，為了避免數據不同的變量尺度影響模型訓練之準確度，因此在保持原始示例之間的相對距離下將原始數據轉換成[0,1]之間的數值，以達到數據標準化，如公式(5)。Al-Shalabi (2006)研究顯示，使用最大值最小值正規化後的模型，依

然保持相當高的模型準確度。

$$x_{i,j}^* = \frac{x_{i,j} - \min_j}{\max_j - \min_j} \in [0,1] \quad (5)$$

$x_{i,j}$ 代表原始數據， $\max_j$ 為 $x_j$ 中的最大值， $\min_j$ 為 $x_j$ 中的最小值， $x_{i,j}^*$ 則代表轉換過後的值。

### 3.3.2 學習模型參數設定

本研究SVM模型分別採用線性核函數(Linear Kernel)、多項式核函數(Polynomial Kernel)以及徑向基核函數(Radial basis function (RBF) Kernel)三種類型之核函數(Kernel

function)，其中Linear Kernel的部分，目標函數之懲罰係數C設定為1，核函數gamma之係數設定為auto，Polynomial-3為degree=3、Polynomial-4為degree=4；Radial basis function Kernel的部分，其餘參數設定皆採用預設值。SMOTE參數指定近鄰之個數N\_neighbors為5。

### 3.3.3 分類性能評估指標

一般研究會以準確度(Accuracy)作為評估模型好壞的指標，但當資料集為不平衡資料集時，準確度會無法反映出少數類別於分類模型中真實的預測性能(Han *et al.*, 2005)。不平衡資料指資料集的某一類別在數量上遠超過於另一個類別。在此情況下，針對二元類別資料，一般學者將數量較大的負類別設定為多數類別(Majority Class)與數量較少的正類別設定為少數類別(Minority Class)。若我們以不平衡資料集訓練傳統的分類器，由於少數類別的樣本數量過於稀少而分類器對該類別學習不足，使得分類器預測結果偏向多數類別。本研究中選擇三個經常被使用於衡量不平衡資料的指標，G-mean(Geometric Mean)、F1與IBA(Index of Balanced Accuracy)做為評估指標。為了說明三個指標，我們列出混淆矩陣(Confusion Matrix)如表3，其中行代表實際類別(Actual Class)、列代表模型預測類別(Predicted Class)。其中TP(True Positive)代表真陽性，為正確識別為陽性數目；FN(False Negative)代表錯誤識別為陰性數目；FP(False Positive)代表假陽性，為錯誤識別為陽性數目；TN(True Negative)代表真陰性，為正確識別為陰性數目。對不平衡資料的分類問題，學者們於評估正類別與負類別的分類正確率項目上，提出Recall評估所有正類別

個體中被正確地判斷為正類別比例與Specificity評估所有負類別個體中被正確地判斷為負類別比例。然而，正確地辨識正類別資料對於研究者來說更為重要。以Recall與Specificity為基礎，過往研究提出以G-mean與F1評估指標，綜合地評估不平衡資料集的分類性能。針對各類別分類正確率，G-mean將兩指標相乘取得幾何平均數，另一方面，F1為調和平均數用以評估不平衡分類資料集之分類正確性。更進一步地，García *et al.* (2009) 提出一個平衡正確性指標(Index of Balanced Accuracy, IBA)以評估學習模型之不平衡資料分類性能。

- Accuracy：為所有被分類為正確個體之比率，如公式(6)：

$$ACC = \frac{TP + TN}{TP + FN + TN + FP} \quad (6)$$

- Precision：為所有被判斷為陽性的個體中，也確實為陽性之比率，如公式(7)：

$$Prec = \frac{TP}{TP + FP} \quad (7)$$

- Recall：為所有陽性個體中，被正確判斷為陽性之比率，如公式(8)：

$$Rec = \frac{TP}{TP + FN} \quad (8)$$

- Specificity：為所有陰性個體中，被正確判斷為陰性之比率，如公式(9)：

$$Spec = \frac{TN}{TN + FP} \quad (9)$$

- G-mean：為Precision與Recall的加權調和平均數，用以評估模型之有效性。如公式(10)：

$$G\text{-mean} = \sqrt{(Rec \times Spec)} \quad (10)$$

- F1：來以評估模型偏差的程度，如公式(11)：

表3 混淆矩陣(Confusion Matrix)

		Predicted class	
		Positive	Negative
Actual class	Positive	TP (True Positive)	FN (False Negative)
	Negative	FP (False Positive)	TN (True Negative)

$$F1 = \frac{2 \times \text{Rec} \times \text{Prec}}{\text{Rec} + \text{Prec}} \quad (11)$$

- 平衡精度指標(Index of Balanced Accuracy, IBA)：García *et al.* (2009)所提出將正類別與負類別的分類正確性皆納入考慮，用於評估分類系統中具有資料不平衡情形之性能，如公式(12)。

$$\text{IBA} = [1 + (\text{Rec} - \text{Spec})] \times \text{Rec} \times \text{Spec} \quad (12)$$

### 3.4 異常事件分類結果

於本研究的實驗中，按照不平衡率(Imbalanced Ratio, IR)：192/20 = 9.6，對原始數據集隨機抽取80%的樣本作為訓練資料，其餘的為測試資料。並透過實驗來驗證Linear SVM、Polynomial SVM與RBF SVM等三種核函數(Kernel function) SVM模型之分類有效性。

此外，在四個不平衡資料分類評估指標上，本研究以原本的不平衡資料集IMB (Imbalance Dataset)作為實驗控制組，並比較使用隨機過採樣方法的ROS以及SMOTE方法。於圖7，長條圖中分別顯示Linear SVM、Polynomial-3 SVM、Polynomial-4 SVM和RBF SVM於評估指標之結果，發現透過Accuracy進行評估後之性能差異不大。其原因為本實驗資料類別分佈不平衡率高達9.6，屬於不平衡分布資料集。如先前所提，使用不平衡資料集訓練學習模型，該模型預測結果大部分為多數類別。因此，IMB、ROS與SMOTE在Accuracy項目上的差異並不顯著。另一方面，由SMOTE資料前處理技術後取得的訓練資料類別比例已平衡為1：1，因此使用SMOTE方法於SVM模型上對G-mean、F1與IBA項目能取得較佳預測性能，並且於使用Polynomial SVM進行分類之分類性

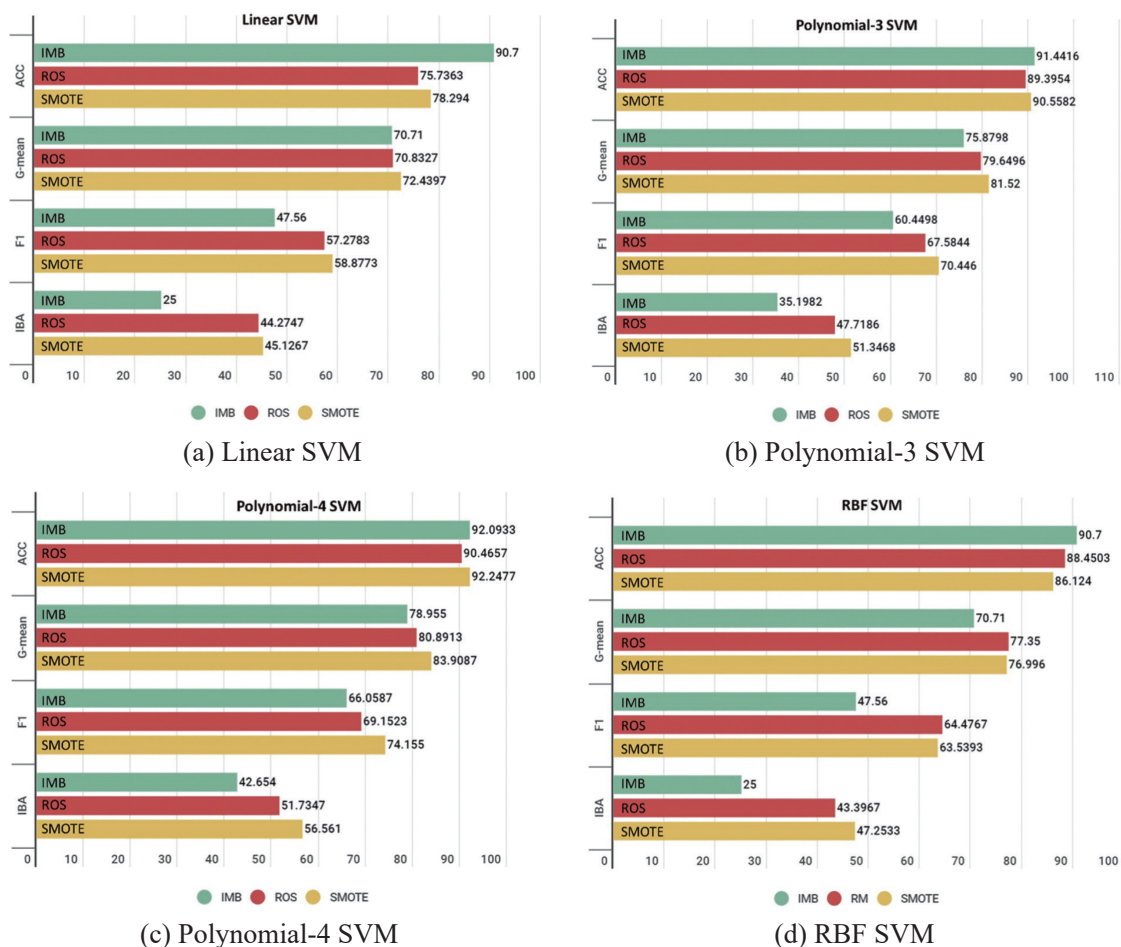


圖7 SVM分類結果

能較好。

## 4. 結論與未來展望

本研究模擬一個感測器供電情境，透過太陽能穩壓充電版提供電源架設低成本空品感測器之物聯網系統。系統透過MQTT通訊協定將感測器所蒐集之數據傳至邊緣運算設備(Raspberry Pi)，存放於SQLite資料庫。此外，透過Node-RED網頁平台呈現感測器即時資料幫助開發者有效管理感測資料。在異常事件偵測上，我們收集感測資料作為機器學習模型的輸入特徵，但由於感測資料集內多數類別屬於供電穩定狀態，少數類別為發生低電壓或電壓中斷等是異常事件，因此本研究採用現今於過抽樣方法中流行的SMOTE解決此不平衡數據集的問題，透過合成少數類樣本來平衡原始數據的分佈，並以SVM分類演算法模型偵測本研究之物聯網系統中太陽能供電的情形是否有異常事件。當使用提出的SMOTE演算法搭配SVM模型辨識到太陽能供電系統出現異常事件，並且發生的時間點為太陽照射的時間範圍內，代表太陽能供電系統可能發生退化或者陽光照射不到等問題，因而需更換太陽能板或位置。在更換期間為確保資料傳送不中斷，建議先以市電作為主供應電源，待更換完成後，切換回太陽能作為主供應電源。最後，實驗結果顯示透過SMOTE進行資料前處理，並使用Polynomial SVM進行分類之分類性能較好。在未來的研究中，可考慮三個研究方向：將氣候因素作為偵測異常事件的輸入特徵與擴展更多用於偵測太陽能低成本低功率感測器；以太能陽替代利用市電所運作之空品感測器，進而達成聯合國於2050年淨零碳排的目標；在未來，我們預計將開發的網頁建置於Heroku託管伺服器(2017)，便於其他的使用者透過網路監測系統供電情況。

## 致 謝

感謝國立臺北護理健康大學與國家科學及技術委員會(National Science and Technology Council, NSTC)以及工業技術研究院(Industrial Technology Research Institute, ITRI)提供本研究相關支援。本研究運算設備與實驗材料由國家科學及技術委員會經費編號MOST 110-2222-E-227-001-MY2贊助。

## 參考文獻

- 台灣電力公司，2022a。火力電廠環境保護。<https://www.taipower.com.tw/tc/page.aspx?mid=216#b01>。
- 台灣電力公司。2022b。再生能源發展概況。<https://www.taipower.com.tw/tc/page.aspx?mid=204>。
- 行政院環境保護署，2022a。行政院環境保護署溫室氣體排放統計。<https://www.epa.gov.tw/Page/81825C40725F211C/6a1ad12a-4903-4b78-b246-8709e7f00c2b>。
- 行政院環境保護署，2022b。認識空污感測物聯網。[https://airtw.epa.gov.tw/cht/Encyclopedia/AirSensor/AirSensor\\_4.aspx](https://airtw.epa.gov.tw/cht/Encyclopedia/AirSensor/AirSensor_4.aspx)。
- Agrawal, M., 2005. Effects of air pollution on agriculture: An issue of national concern. *National Academy Science Letters*, 28(3-4), 93-106.
- Al-Shalabi, L., 2006. Coding and normalization: The effect of accuracy, simplicity, and training time. In: RCED.
- Al Kharusi, K., A. El Haffar and M. Mesbah, 2022. Fault Detection and Classification in Transmission Lines Connected to Inverter-Based Generators Using Machine Learning. *Energies*, 15(15), 5475.
- Bai, C., L. Zhou, M. Xia and C. Feng, 2020. Analysis of the spatial association network



- structure of China's transportation carbon emissions and its driving factors. *Journal of environmental management*, 253, 109765.
- Bastianoni, S., F. M. Pulselli and E. Tiezzi, 2004. The problem of assigning responsibility for greenhouse gas emissions. *Ecological economics*, 49(3), 253-257.
- Chawla, N. V., K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.
- Cortes, C. and V. Vapnik, 1995. Support-vector networks. *Machine learning*, 20(3), 273-297.
- Donaldson, K., N. Mills, W. MacNee, S. Robinson and D. Newby, 2005. Role of inflammation in cardiopulmonary health effects of PM. *Toxicology and applied pharmacology*, 207 (2), 483-488.
- Eltamaly, A. M., M. A. Alotaibi, A. I. Alolah and M. A. Ahmed, 2021. Iot-based hybrid renewable energy system for smart campus. *Sustainability*, 13(15), 8555.
- García, V., R. A. Mollineda J. S. Sánchez, 2009. Index of balanced accuracy: A performance measure for skewed class distributions. Iberian conference on pattern recognition and image analysis.
- Glavic, M. and T. Van Cutsem, 2011. A short survey of methods for voltage instability detection. 2011 IEEE Power and Energy Society General Meeting.
- Høiem, K. W., V. Santi, B. N. Torsæter, H. Langseth, C. A. Andresen and G. H. Rosenlund, 2020. Comparative study of event prediction in power grids using supervised machine learning methods. 2020 International Conference on Smart Energy Systems and Technologies (SEST).
- Han, H., W.-Y. Wang and B.-H. Mao, 2005. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. International conference on intelligent computing.
- Heath, N., 2014. How IBM's Node-RED is hacking together the internet of things. *TechRepublic*. *TechRepublic*, March, 14.
- Heroku, 2017. <https://www.heroku.com/home>.
- Ichikawa, H., S. Yokogawa, Y. Kawakita, K. Sawada, T. Sogabe, A. Minegishi and H. Uehara, 2019. An approach to renewable-energy dominant grids via distributed electrical energy platform for IoT systems. 2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm).
- Jack, K. E., A. Usoro, E. S. Udofa and L. A. Johnson, 2020. Real Time Energy Data Monitoring Model for Integrated Renewable Energy System with other Collaborative Energy Supply.
- Kannan, N., and D. Vakeesan, 2016. Solar energy for future world:-A review. *Renewable and Sustainable Energy Reviews*, 62, 1092-1105.
- Lewis, K., 2016. Node-RED visual programming for the Internet of Things (IoT) is now a JS Foundation Project. In.
- Mahesh, B., 2020. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*. 9, 381-386.
- Martins, F., C. Felgueiras, M. Smitkova and N. Caetano, 2019. Analysis of fossil fuel energy consumption and environmental impacts in European countries. *Energies*, 12(6), 964.
- Menardi, G. and N. Torelli, 2014. Training and assessing classification rules with imbalanced data. *Data mining and knowledge discovery*, 28(1), 92-122.

- The MQTT protocol, 2012. <http://www.mqtt.org>.
- Noble, W. S., 2006. What is a support vector machine? *Nature biotechnology*, 24(12), 1565-1567.
- Novitasari, D., A. Foeady, R. Nariswari, A. Asyhar, N. Ulinnuha, Y. Farida, D. Santi and F. Setiawan, 2020. Whirlwind Classification with Imbalanced Upper Air Data Handling using SMOTE Algorithm and SVM Classifier. *Journal of Physics: Conference Series*.
- Patil, J., P. Sharma and K. Mhatre, 2021. Global warming Induced Stress and its Impact on Biodiversity. *Science and Technology*, 6(3).
- Pisner, D. A. and D. M. Schnyer, 2020. Support vector machine. In *Machine learning* (pp. 101-121). Elsevier.
- Ryabushev, Y. A., 2022. Solar energy as alternative energy source: advantages and disadvantages.
- Sanjuan, E. B., I. A. Cardiel, J. A. Cerrada C. Cerrada, 2020. Message queuing telemetry transport (MQTT) security: a cryptographic smart card approach. *IEEE Access*, 8, 115051-115062.
- Sinsel, S. R., R. L. Riemke and V. H. Hoffmann, 2020. Challenges and solution technologies for the integration of variable renewable energy sources—a review. *Renewable energy*, 145, 2271-2285.
- Widodo, A. and B.-S. Yang, 2007. Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical systems and signal processing*, 21(6), 2560-2574.
- Zhang, Y., 2012. Support vector machine classification algorithm and its application. In *international conference on information computing and applications* (pp. 179-186).

# Using SMOTE Technology to Improve Abnormal Detection of Solar Power Supply – A Case of Micro Air Quality Monitoring System

Yu Wang<sup>1</sup>   Liang-Sian Lin<sup>2\*</sup>   Zhi-Yu Chen<sup>1</sup>   Chen-Wei Liu<sup>3</sup>

## ABSTRACT

The massive emission of carbon dioxide (CO<sub>2</sub>) is a leading cause of global warming and ecological destruction. The main source of electricity in Taiwan is thermal power generation. However, the air pollutants and CO<sub>2</sub> emitted from thermal power generation that causes harm to human health and the environment. Consequently, Taiwan has implemented energy transformation including reduction of coal-burning power plants, embracing natural gas, and developing green energy. Due to the popularization of the Internet of Things (IoT), the industrial demand for electricity is also increased. Some studies have suggested using solar energy as a primary source of energy for the IoT system. However, due to uncertain climate change, the IoT based solar energy system frequently occurs abnormal events such as low voltage and voltage interruption. Under this circumstance, in this study, we developed a micro air quality monitoring system based on solar energy supply. In the suggested system, the message queuing telemetry transport (MQTT) protocol is used to transform sensing data into SQLite database on Raspberry Pi. In addition, we use SMOTE to improve classification ability of the support vector machine (SVM) model for imbalanced distribution of normal abnormal events. Finally, we also develop a user interface (UI) using Node-RED tool to display real-time sensor data.

**Keywords:** solar energy supply; greenhouse effect; abnormal event; real-time sensing data.

---

<sup>1</sup> College, Department of Information Management, National Taipei University of Nursing and Health Sciences.

<sup>2</sup> Associate Professor, Department of Information Management, National Taipei University of Nursing and Health Sciences.

<sup>3</sup> Senior Researcher, Green Energy and Environment Research Laboratories, Industrial Technology Research Institute.

\*Corresponding Author, Phone: +886-2-28227101#1234, E-mail: [lianghsien@ntunhs.edu.tw](mailto:lianghsien@ntunhs.edu.tw)

Received Date: August 31, 2022

Revised Date: November 11, 2022

Accepted Date: November 23, 2022