

# 以機器學習偵測異常狀態：龍井太陽能發電場案例

許志義<sup>1\*</sup> 古典家<sup>2</sup> 張志豪<sup>2</sup> 葉法明<sup>3</sup> 陳宗薊<sup>3</sup>

## 摘 要

太陽能發電方式相較於火力發電、核能發電，發電量對於氣候、溫度、日照長度等環境因素較為敏感，發電量具有不穩定性的特徵。再者，太陽能發電系統機器設備的健康狀態，也會影響發電量，尤其是來自於太陽能發電系統之中變流器(inverter)的健康狀態所導致的不確定性。本研究利用機器學習，對變流器作異常偵測，以便即時掌握變流器的健康狀態，提供台電公司參考，作為降低發電量不確定性的對策之一。本研究利用台電公司再生能源處提供的臺中龍井一期發電的歷史數據，應用預測性維修的故障診斷方法，針對變流器進行異常狀態的偵測。本研究先採用因素分析及主成份分析等方法，為實證資料萃取特徵因素，然後使用邏輯斯迴歸、支援向量機、隨機森林、K最鄰近法等方法，進行分類模型的訓練。此間，評估分類模型之標準，包括：最多正確分類異常資料的資料筆數與精準度、F1分數。本研究實證結果顯示：

1. 各變流器的健康狀態，均適合使用邏輯斯迴歸為初步診斷工具；再進一步與其他工具做評估比較，以利尋找合宜的訓練工具。本實驗中，沒有任何樣本適用K最鄰近法。
2. 本研究發現，4號變流器與7號變流器之KMO取樣適切性量，介於0.8至0.9之間，屬於良好，適合作因素分析；累積解釋變異量偏低的4號變流器與偏高的7號變流器，均合適隨機森林方法。後來的精準度與F1分數都相對顯著，應證了Kaiser (1974)觀點。
3. 在實證結果中，2號變流器、3號變流器、5號變流器與6號變流器之KMO取樣適切性量，介於0.7至0.8之間，屬於中等。使用支援向量機方法的2號變流器與6號變流器，在精準度與F1分數方面呈現差異，發現6號變流器輸入特徵缺乏溫度與電壓相關特徵。

本研究建議：

1. 台電公司宜提前對6號變流器進行預防性維修，其他變流器可繼續按照檢修週期保養。
2. 比起定期維修策略，若採用預測性維修策略，預測每一台變流器可能發生故障的時間，再依照預測的結果制定維修計畫，將能夠大幅降低維修成本。

關鍵詞：太陽能發電，預測性維修，故障診斷，異常偵測，因素分析，主成份分析，邏輯斯迴歸，支援向量機，隨機森林，K最鄰近法

<sup>1</sup> 國立臺北商業大學榮譽講座教授、國立中興大學應用經濟學系暨資訊管理學系 合聘教授

<sup>2</sup> 國立中興大學資訊管理學系所 碩士

<sup>3</sup> 國立中興大學應用經濟學系所 博士生

\*通訊作者電話: 04-2285-0505, E-mail: [hsu@nchu.edu.tw](mailto:hsu@nchu.edu.tw)

收到日期: 2019年09月17日

修正日期: 2020年11月26日

接受日期: 2021年01月05日

# 1. 緒 論

## 1.1 研究動機

近年來，臺灣政府積極發展再生能源，提高再生能源佔總發電量之比例。截至2018年底，再生能源佔臺灣總發電量的4.9%。若以太陽能發電容量來看，2017年其累積裝置容量為138.69萬瓩，2018年增加為234.22萬瓩，成長約68.8%。若以發電量來看，2017年太陽光電累積發電量為16.21億度，2018年增加為26.55億度，成長約63.8%(台灣電力公司，2019)，此顯示太陽能發電在臺灣未來的整體發電系統中日益重要，這是因為臺灣地處亞熱帶，非常適合發展太陽光電系統。

一直以來，太陽能發電預測被認為是非常重要的議題。因為相較於火力發電與核能發電，太陽能發電具有相對不穩定性與不確定性。這是由於外在氣候條件變化的影響，加上太陽能發電系統機器設備不一定處於健康狀態所造成的。因此，當太陽能發電的占比增加時，這種不確定性會造成台電公司電力調度上的困難。

若能藉由機器學習異常偵測，事前預測出太陽能設備不穩定的特徵及其出現的時間點，就能夠即時做出因應的配套措施，來降低太陽能發電量的不確定性對電力系統可能造成的負面影響。

本文的目的為異常診斷。異常診斷是變流器的異常狀態判斷，但判斷的目的其實為預測性維修。也就是說我們透過此異常診斷，知道變流器何時損壞，當我們累積經驗之後，就可以在變流器損壞之前，趕快進行維修，避免變流器臨時故障，臨時維修。

因此，預測性維修可以說是此研究異常診斷之最終目的。與其異常發生之後，才安排人員進行臨時維修，不如當發現到數據些微異常，已經預測到機器設備將要故障，馬上安排人員進行維修，避免臨時故障與臨時維修的狀況發生。

太陽能發電的預測，國內外學者皆提出許多相關研究。研究主分為兩大類。第一類為太陽能發電量與外在環境影響的關聯性。王群翔(2018)的研究實證臺灣四季氣候與乾濕季氣候對太陽能發電量有影響。此外，許志義等人(2018)的研究實證太陽能發電量會受到溫度、濕度、氣壓、日照時間等因素所影響。第二類研究主要為太陽能發電量與太陽能發電機器設備品質的關係。陳建融(2016)的研究對太陽能發電系統作異常狀態偵測。Braun *et al.* (2012)的研究對太陽能面板作異常狀態偵測。此外，Alam *et al.* (2014)的研究同時也做了太陽能面板異常狀態的偵測。換言之，太陽能發電量受到外在環境與機器設備健康狀態的影響。而目前國內研究中，尚無針對太陽能發電系統中變流器作異常偵測，因此本研究利用臺中龍井太陽能電場的變流器歷史發電數據作異常狀態的偵測，找出判斷變流器是否故障的特徵為何，如果能夠精準地監控故障狀態的話，也能夠提升變流器預測異常的精準度。

太陽光變流器(PV inverter)可以將太陽能板產生的可變直流電壓轉換為市電頻率交流電，可以回饋回商用輸電系統，或是供離網的電網使用。太陽光變流器是光伏陣列系統中重要的系統平衡之一，可以配合一般交流供電的設備使用。太陽能變流器有配合光伏陣列的特殊功能，例如最大功率點追蹤及孤島效應保護的機能。

因此當太陽光變流器發生異常時，會造成太陽能板所產生的直流電無法正常輸出。而變流器容易故障的原因有：市電異常、絕緣阻抗過低、漏電流過高、環境溫度或散熱器溫度過高、監控無數據、絕緣阻抗低、漏電流故障、變流器開機無回應、電網故障、直流過壓保護等。

外在環境因素影響無法由人為控制，但是能透過機器學習技術預測未來不同季節的太陽能發電量。相較於外在環境因素，機器

設備的狀態是人為可以控制的。首先可以選擇適用的機器設備與設定機器運行的最佳參數，使機器能夠穩定地運行。再者可以預防機器發生故障，實踐方法除了定期維修(Periodic Maintenance)之外，透過預測性維修(Predictive Maintenance)的導入，可以預測未來機器設備的故障時間、故障型式與使用壽命。此時台電公司可以依照預測出的故障時間點與故障的型式，預先制定維修計畫，減少機器發生臨時性故障的可能性，並提升太陽能發電量。

## 1.2 研究目的

本研究目的以機器學習偵測太陽能發電之異常狀態。以龍井太陽能發電廠為實證案例。

## 1.3 研究方法

為了達成以上目的，採用邏輯斯迴歸、支援向量機、隨機森林、K最鄰近法進行資料分析訓練並以混淆矩陣、精準度、F1分數作為檢驗標準，最佳結果的輸入特徵即為判斷每一台變流器是否故障的關鍵特徵。

# 2. 文獻探討

## 2.1 預測性維修

預測性維修(Predictive Maintenance)指在機器運行時，對機器的零組件進行連續性的狀態監測及故障診斷，依零組件之狀態及可能的故障時間、故障模式，制定預先之維修計畫。預測性維修包含四個範疇，狀態監測、故障診斷、狀態預測、維修決策。狀態監測指布建感知器於機器內部，對想要監控之零組件記錄其在某一時間點之溫度、濕度、振動頻率、壓力等指標數據。故障診斷指檢測機器之故障與否。狀態診斷指依據現行零組件之狀態，預測未來機器可能發生故障之時間、模式，像是性能衰退預測。維修決策指從人員、資源、時間、費用等角度出發，根據狀態監測、故障診

斷、狀態預測之結果進行維修可行性分析，制定預先維修計畫(Mobley, 2002)。

## 2.2 因素分析與主成份分析

因素分析為統計學中用來處理維度縮減之方法，在進行因素分析前，必須通過兩個檢定，第一個檢定為KMO (Kaiser-Meyer-Olkin)，KMO是使用相關矩陣來計算，根據Kaiser (1974)提出KMO分數之解釋，如表1。

表1 KMO分數及對應解釋(Kaiser, 1974)

KMO分數	英文釋義	中文釋義
$\geq 0.9$	Marvelous	卓越的
$\geq 0.8$	Meritorious	良好的
$\geq 0.7$	Middling	中等的
$\geq 0.6$	Mediocre	平凡的
$\geq 0.5$	Miserable	悲慘的
$< 0.5$	Unacceptable	不可接受的

第二個檢定為Bartlett球形檢定，是使用相關係數來計算，Bartlett球形檢定的顯著性必須小於0.05，才算通過檢定。接著計算出因素分析的結果後，累積解釋變異量必須大於0.5，才足以代表母體資料。

通過上述兩個檢定後，使用因素分析中的主成份分析進行因素的萃取，主成份分析(Principal Component Analysis, PCA)是由Pearson (1901)所發明，之後經由Hotelling (1930)發展並命名的一種統計方法。主成份分析是根據特徵之間的相關性，使用正交轉軸法，將n個特徵加以轉換，得到P個成份( $P \leq n$ )與其成份的特徵值，成份中在正交軸的變異數為最大，且成份間彼此無關(林清山，1991；傅粹馨，2002)。利用主成份分析可以降低特徵的維度，找出對分類有重要影響的特徵。

## 2.3 機器學習

### 2.3.1 邏輯斯迴歸

邏輯斯迴歸(Logistic Regression)為二元分

類的線性模型，根據選取的特徵及權重，對樣本進行機率分類。邏輯斯迴歸以 $\log$ 函數計算樣本屬於某一類的機率，如公式1：

$$\log p(1-p) \quad (1)$$

$p$ 值代表A類事件， $1-p$ 代表B類事件，當 $W_0X_0 + W_1X_1 + \dots + W_nX_n$ 愈大時，判斷為A類事件的機率愈大；反之，當 $W_0X_0 + W_1X_1 + \dots + W_nX_n$ 愈小時，判斷為A類事件的機率愈小，亦即判斷為B類事件的機率愈大(Raschk, 2016)。

### 2.3.2 支援向量機

支援向量機(Support Vector Machine)屬於監

督式學習之演算法，可用來處理線性、非線性分類與迴歸的問題，支援向量機之目的在於找到一個決策邊界(Decision Boundary)，使兩類資料之間的邊界(Margins)最大化(Raschk, 2016)，支援向量機示意圖如圖1。

### 2.3.3 隨機森林

隨機森林(Random Forest)屬於監督式學習之演算法，用來處理分類與迴歸的問題，為一建立在決策樹上的整體學習法。決策樹之示意圖如圖2。

根據訓練集的特徵，決策樹模型會出一系列問題，並依此推斷樣本的類別標籤，隨機森

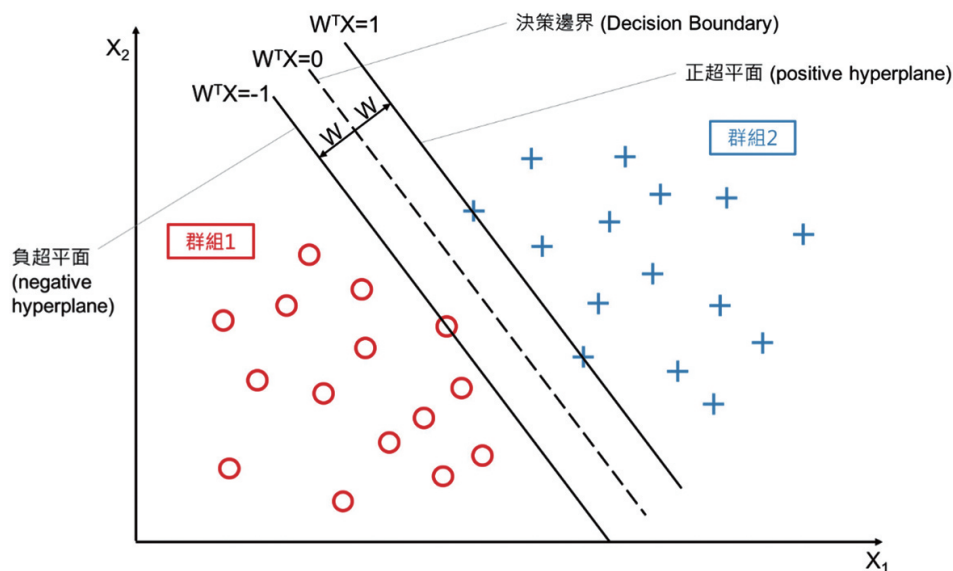


圖1 支援向量機示意(Raschk, 2016)

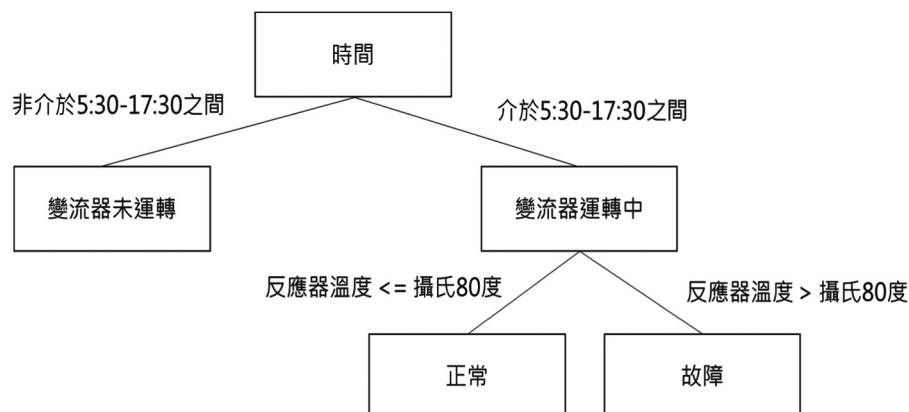


圖2 決策樹示意(本研究繪製)

註：圖中設定之數字為假想情況，與真實情況不同



林是根據整體學習(ensemble learning)的方法，由多棵決策樹的組合而成(Raschk, 2016)。在隨機森林的參數調整，主要有`n_estimators`參數與`n_jobs`參數，前者表示樹木的多寡，後者表示要使用多少個電腦的核心進行運算，本研究將`n_estimators`設定為1000，`n_jobs`設定為-1。

### 2.3.4 K最鄰近法

K最鄰近法(K-nearest neighbor)屬於監督式學習之演算法，用來處理分類與迴歸的問題，在使用K最近鄰時，需選定一個K值與距離度量(distance metric)，K值代表想要分類之最鄰近樣本的數量，最後以多數決的方式指定此樣本的類別標籤，亦即在樣本鄰近K個值之多數的類別標籤為何，該樣本的類別標籤即為何(Raschk, 2016)。K最近鄰居法主要的參數為`n_neighbors`參數，表示要分類之最鄰近樣本的數量，本研究將`n_neighbors`設定為5。

## 2.4 應用機器學習於再生能源領域國內外相關文獻

許志義等人(2018)的研究以太陽能發電資料與氣象局2015年全年資料(共14項特徵)，包含兩個實證分析的部分，實證(一)3個方法實作。方法一，使用時間序列將資料切割為「春、夏、秋、冬、不分季」，以支援向量機建立預測模型。方法二，將預處理後的資料以Wrapper(包裝法)作特徵選取，之後以支援向量機建立預測模型。方法三同方法二，另在預處理階段時，增加聚類處理，分為日照時數少、中、多三類。接著以RMSE值(Root Mean Squared Error, RMSE)比較各模型預測結果，得出在不同季節、日照時數下最佳發電量預測模型。實證(二)，將原本14項特徵，新增「月份」、「前一日最高溫」與「前一日最低溫」，並依照文獻及試誤法剔除不適合之特徵，篩選出七組模型，採三種核函數(Linear、

Radius Basis Function (RBF)、Polynomial)的SVM (Support Vector Machine, SVM)建立模型，預測結果顯示新增特徵後可提升預測準確性，最後使用倒傳遞類神經網路建立模型，藉由調整隱藏層神經元個數，提升預測準確性。研究結果發現，倒傳遞類神經網路得到最佳之太陽能發電量預測準確率。

張家源(2018)以風力發電機數據與氣象觀測資料作為輸入變數，使用隨機森林、支援向量機迴歸、人工神經網路、最近鄰居法與線性迴歸，依序不分季、季風、海陸風等天氣因素，建立風力發電量預測。研究結果發現編號5號之風機為最佳結果，編號26號風機為最差結果，另外，將不分季、東北季風與海陸風時段的風力發電量建立預測模型，隨機森林可以得到最低的誤差值。最後，該研究使用SFS演算法找出個模型最相關特徵，得到更低之預測誤差值。

王群翔(2018)以太陽能發電量資料與氣象觀測資料作為輸入變數，並以最小絕對壓縮挑選機制做特徵選取，使用支援向量機、倒傳遞類神經網路及隨機森林，分析臺灣四季與乾濕季的變化對太陽能發電量的影響程度，研究結果發現夏季的預測結果為所有時間序列中最好之結果，支援向量機的RMSE值為30.79kwh，倒傳遞類神經網路的RMSE值為28.11kwh，隨機森林的RMSE值為26.50kwh，隨機森林為三個演算法中表現最佳之預測模型。最後驗證模型的訓練資料與測試資料比例，會影響模型的預測準確度。

王柏勝(2017)以太陽能發電數據與觀測資料查詢系CODIS<sup>1</sup>中的歷史氣候資料作為輸入變數，並以Wrapper做特徵選取，最後將數據聚類，使用支援向量機建立太陽能發電量預測模型，研究結果發現在四季、不分季與聚類之模型結果有好有壞。

蔡松諺(2017)對嘉義、高雄、金門等3個地

<sup>1</sup>CODIS是一種分散式Redis的解決方案，CODIS是Wandoujia Infrastructure Team開發的一個分散式Redis服務，使用者可以看成是一個無限記憶體的Redis服務。

區做太陽能發電量預測，以日照值、面板溫度與氣象局網站資料作為輸入變數，使用類神經網路與支援向量機進行模型之訓練，研究結果發現使用類神經網路預測太陽能發電量較佳。此外，因為臺灣四季變化、梅雨季與颱風侵襲的影響，作者加入了月份、前天最高溫、前天最低溫等變數，實證結果證實月份與前天最高溫、前天最低溫能有效地提高預測準確率。而最佳預測結果所使用的輸入變數為月份、溫度、露點溫度、測站氣壓、海平面氣壓、相對溼度、日照時間、日照值、前天最高溫度、前天最低溫度。

陳建融(2016)以集群分析(K-means)、最大期望演算法(Expectation-maximization, EM)、兩種基於密度的聚類演算法(Density-based spatial clustering of applications with noise, DBSCAN)、Ordering points to identify the clustering structure, OPTICS) 做太陽能發電系統的異常偵測，其中最大期望演算法兩種基於密度的聚類演算法(DBSCAN、OPTICS)能夠成功預測異常狀態。

Alam *et al.* (2014)利用Spread Spectrum Time Domain Reflectometry (SSTDR)方法對太陽能面板做異常偵測，研究結果證實此方法能夠預測太陽能面板的異常狀態。

Braun *et al.* (2012)利用最小共變異數決定Robust minimum Covariance Determinant (MCD)方法，以arc-fault與ground-fault數據作異常偵測。研究結果發現Robust minimum Covariance Determinant (MCD)方法可以適用於多種太陽能面板的異常型態。

從上述文獻得知，國內的學者利用過支援向量機、隨機森林、K最鄰近法、線性迴歸、類神經網路、倒傳遞類神經網路，以氣象資料搭配歷史發電數據實證太陽能發電量受到外在環境因素影響。另外，國內外學者利用最大期望演算法以及兩種基於密度的聚類演算法(DBSCAN、OPTICS)、Spread Spectrum Time Domain Reflectometry (SSTDR)、Robust minimum Covariance Determinant (MCD)，成功

對太陽能發電系統的機器設備作異常偵測。實證使用歷史發電資料可以偵測異常狀態，且太陽能發電系統中機器設備的故障與否，會影響太陽能發電量。

### 3. 研究方法

#### 3.1 實證資料

本研究實證資料來自於台灣電力公司再生能源處提供之臺中龍井地區太陽能發電場一期變流器(inverter)之歷史發電數據。資料擷取時間範圍自2015年1月1日0時0分起至2018年10月1日0時0分止。資料時間間隔為每十分鐘一筆變流器發電數據。變流器發電數據變數名稱為以下15種：時間、有效功率、總效能、AC電流、頻率、今日效能、變流器狀態、環境溫度、外殼溫度、正常模式、冷卻溫度、PV電流、PV電壓、無效電能、反應器溫度。

在臺中龍井地區太陽能發電場一期中共有7台變流器。每一台變流器之資料擷取時間範圍及資料時間間隔皆為10分鐘。臺中龍井地區太陽能發電場一期每一台變流器含有197,137筆資料。

#### 3.2 資料預處理

資料預處理的第一步驟為判斷變流器資料是否具有可解釋性，編號1號的變流器，於資料擷取時間內，每一筆資料數值皆相同，顯然該變流器資料傳輸或接收發生問題，形成無效資料，因此本研究不予採用。

資料預處理的第二步驟為刪除原始資料中數值為Bad(缺失值)的資料，亦即資料品質不佳，無法作為機器學習之用，應予排除。表2列示2號至7號變流器中，每一台變流器資料含有Bad值的資料筆數。

資料預處理第三步驟為觀察每一台變流器每一項特徵的數值變化。表3列示2號至7號變流器中2015年、2016年、2017年、2018年的總效

表2 龍井一期2號至7號變流器資料欄位含有Bad值之資料筆數(本研究整理)

變流器編號	資料筆數
2號	4344筆
3號	4344筆
4號	4344筆
5號	4248筆
6號	4344筆
7號	4344筆

表3 歷年度2號變流器總效能值(本研究整理)

單位：度

變流器編號	2015年	2016年	2017年	2018年
2號	359,394.84	333,639.63	347,314.09	44,973.66
3號	335,449.38	325,394.91	284,346.56	284,346.56
4號	350,336.72	332,607.16	281,264.72	0
5號	288,778.72	328,916.94	340,188.13	44,440.64
6號	224,865.56	224,865.56	224,865.56	224,865.56
7號	286,582.50	286,582.50	286,582.50	286,582.50

能。總效能為一個年度的效能累積值。在原始資料中，台電公司設置總效能的歸零重新計算時間為每一年的1月1日23時50分。因此，2015年度總效能計算的時間為2015年1月1日23時50分至2016年1月1日23時40分。2016年度總效能計算的時間為2016年1月1日23時50分至2017年1月1日23時40分。2017年度總效能計算的時間為2017年1月1日23時50分至2018年1月1日23時40分。而2018年度總效能值，因本研究的資料擷取時間只到2018年10月1日0時0分，因此2018年度總效能計算的時間為2018年1月1日23時50分至2018年10月1日0時0分。

2015年年度總效能最佳至最差依序為：2號變流器(359,394.84 kWh)、4號變流器(350,336.72 kWh)、3號變流器(335,449.38 kWh)、5號變流器(288,778.72kWh)、7號變流器(286,582.50 kWh)、6號變流器(224,865.56 kWh)。此外，6號變流器在2015年9月9日7時50分開始，總效能的數值維持在224,865.56 kWh，無繼續增加，表示自2015年9月9日7時50分開始，6號變流器可能停止運作或是產生故

障。7號變流器2015年11月1日6時50分開始，總效能的數值維持在286,582.50 kWh，無繼續增加，表示自2015年11月1日6時50分，7號變流器可能停止運作或是產生故障。

2016年年度總效能最佳至最差依序為：2號變流器(333,639.63 kWh)、4號變流器332,607.16 kWh)、5號變流器(328,916.94 kWh)、3號變流器(325,394.91 kWh)。

6號、7號變流器因總效能的數值不再變化，可能是資料的傳輸、接收產生問題，或是該變流器已經停止運作或故障，所以不計入6台變流器的2016年年度總效能排名。

2017年年度總效能最佳至最差依序為：2號變流器(347,314.09 kWh)、5號變流器(340,188.13 kWh)、3號變流器(284,346.56 kWh)、4號變流器(281,264.72 kWh)。

此外，3號變流器在2017年10月21日8時0分開始，總效能的數值維持在284,346.56 kWh，無繼續增加，表示自2017年10月21日8時0分開始3號變流器可能停止運作或是產生故障。6號、7號變流器因總效能的數值不再變



化，可能是資料的傳輸、接收產生問題，或是該變流器已經停止運作或故障，所以不計入6台變流器的2017年年度總效能排名。

2018年年度總效能最佳至最差依序為：2號變流器(44,973.66 kWh)、5號變流器(44,440.64 kWh)。此外，2號變流器在2018年3月7日12時40分開始，總效能的數值維持在44,440.64 kWh，無繼續增加，表示自2018年3月7日12時40分開始2號變流器可能停止運作或是產生故障。3號、6號、7號變流器因總效能的數值不再變化，可能是資料的傳輸、接收產生問題，或是該變流器已經停止運作或故障，所以不計入6台變流器的2018年年度總效能排名。4號變流器2018年度總效能為0 kWh，表示該變流器於2018年度無運作或是故障。

根據表3與上述探討得知，2號變流器在2015年至2018年年度總效能皆為表現最佳的變流器，也是異常資料最少的變流器。此外，將2號變流器各年度總效能值比較，2015年年度總效能值(359,394.84 kWh)最高。因此接下來將以2015年2號變流器資料為例，呈現冬季、夏季，有效功率、AC電流、頻率、今日效能、環境溫度、外殼溫度、正常模式、冷卻溫度、PV電流、PV電壓、無效電能、反應器溫度等12個特徵的數值變化與比較。冬季的資料時間為：2015年1月1日0時0分至2015年1月8日0時0分。夏季的資料時間為：2015年7月1日0時0分至2015年7月8日0時0分。

圖3為2015年2號變流器冬季資料，圖4為2015年2號變流器夏季資料。可以得知在冬季每一天有效功率的最大值普遍高於夏季每一天有效功率的最大值，原因為冬季的日照時間短、溫度低，因此冬季的發電效率高。此外，冬季有效功率顯著變化的時間約自5時30分至17時30分止；而夏季有效功率顯著變化的時間約從5時30分至19時0分止。表示夏季的日照時間長，溫度高，因此夏季的發電量高。

再者於圖3中，1月1日與1月3日的有效功率曲線表現穩定，而1月6日、1月7日以及圖4的

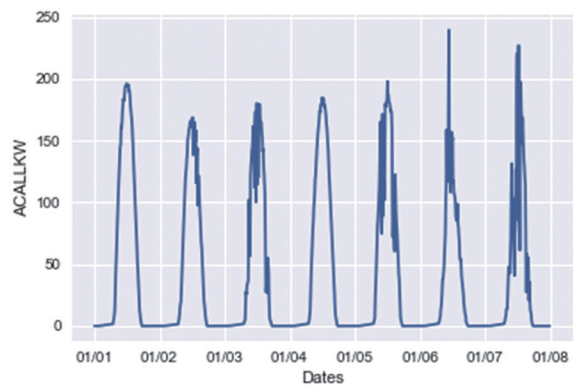


圖3 典型冬季一周七天之太陽能有效功率曲線(以2015年2號變流器為例)(本研究繪製)

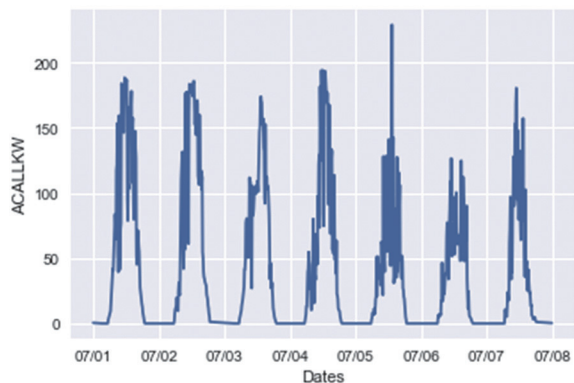


圖4 典型夏季一周七天之太陽能有效功率曲線(以2015年2號變流器為例)(本研究繪製)

7月1日至7月7日有效功率曲線表現較差，可能的原因為受到外在環境因素(包含天氣、溫度、降雨量、沙塵、外來物等)，影響太陽能面板接收太陽輻射量，最後影響到變流器的發電效率。

由表4與表5可以得知每一天變流器的效能。在冬季資料中，表現最好的是1月1日(1211.31 kWh)，表現最差的是1月7日(761.98 kWh)。該周今日效能的平均數為990.10 kWh。在夏季資料中，表現最好的是7月1日(1210.98 kWh)，表現最差的是7月6日(644.30 kWh)，該周今日效能的平均數為938.73 kWh。表示冬季的發電效率比夏季高。此外，7月5日至7月7日的效能表現差，可能是受到外在環境的影響，導致這三天的發電效率差。

圖5為2015年冬季一周2號變流器馬達頻率資料，圖6為2015年夏季一周2號變流器馬達頻



表4 2號變流器冬季資料今日效能一覽(本研究整理)

單位：度

今日效能	1月1日	1月2日	1月3日	1月4日	1月5日	1月6日	1月7日
最大值	1211.31	995.96	949.81	1131.92	1057.79	761.98	821.92

表5 2號變流器夏季資料今日效能一覽(本研究整理)

單位：度

今日效能	7月1日	7月2日	7月3日	7月4日	7月5日	7月6日	7月7日
最大值	1210.98	1210.47	1028.42	1052.98	693.45	644.30	730.54



圖5 典型冬季一周七天之太陽能頻率折線圖 (以2015年2號變流器為例)(本研究繪製)

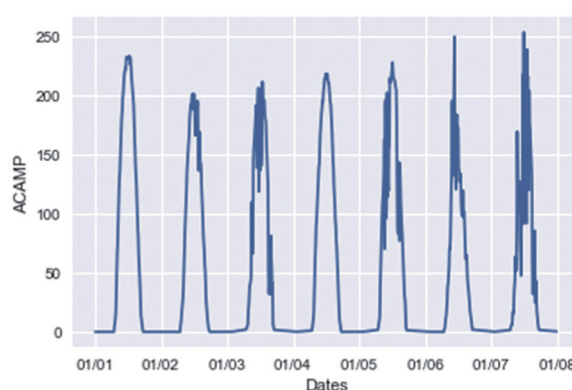


圖7 典型冬季一周七天之太陽能AC電流曲線 (以2015年2號變流器為例)(本研究繪製)

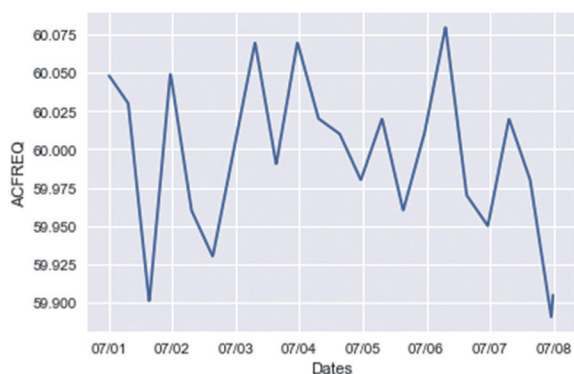


圖6 典型夏季一周七天之太陽能頻率折線圖 (以2015年2號變流器為例)(本研究繪製)

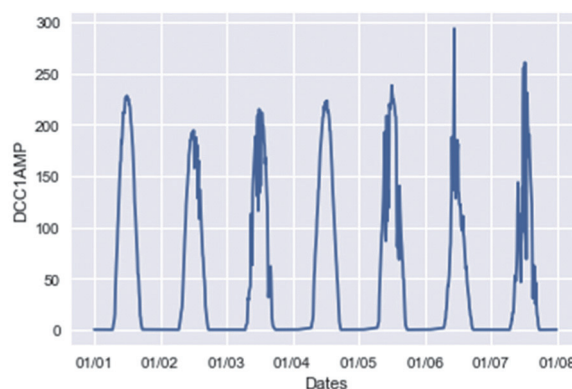


圖8 典型冬季一周七天之太陽能PV電流曲線 (以2015年2號變流器為例)(本研究繪製)

率資料。而頻率為變流器馬達每秒的轉速，單位為赫茲(Hz)/秒(s)。由圖5、圖6得知變流器的頻率維持在59.9 Hz/s至60.075 Hz/s之間跳動，不會因為季節而有顯著性的變化。因此頻率的穩定有助於變流器正常運作與否。

圖7為2015年冬季一周2號變流器AC電流曲線，圖8為2015年冬季一周2號變流器PV電流曲線。圖9為2015年夏季一周2號變流器AC電流曲

線，圖10為2015年夏季一周2號變流器PV電流曲線。而AC電流指變流器中交流電之電流，單位為安培(A)。PV電流指太陽能板輸入至變流器的電流，單位為安培(A)。由圖7至圖10可以得知AC電流與PV電流呈高度正相關的關係，另外，也與有效功率呈高度正相關的關係。因此可見得溫度高低會影響到發電功率也會影響

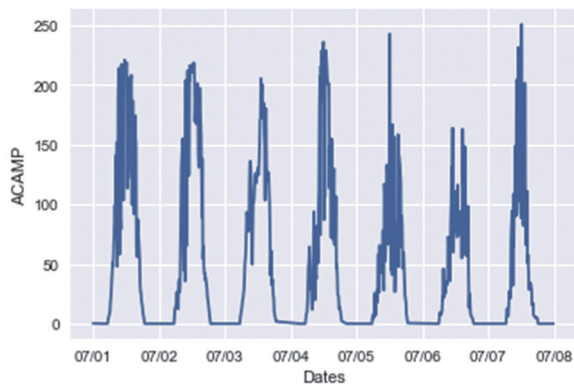


圖9 典型夏季一周七天之太陽能AC電流曲線  
(以2015年2號變流器為例) (本研究繪製)

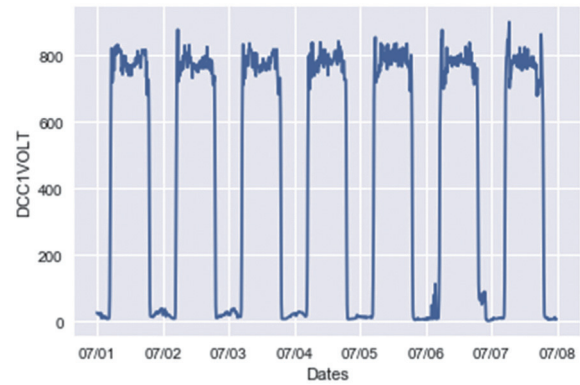


圖12 典型夏季一周七天之太陽能PV電壓曲線  
(以2015年2號變流器為例) (本研究繪製)

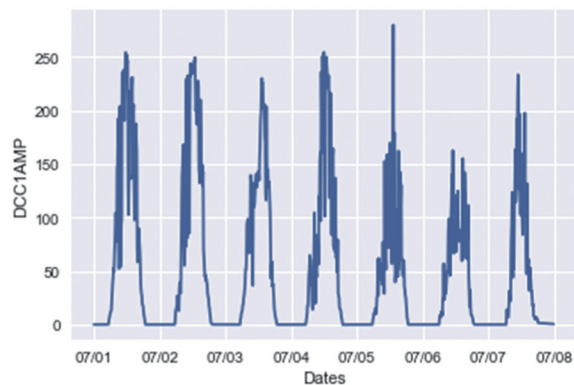


圖10 典型夏季一周七天之太陽能PV電流曲線  
(以2015年2號變流器為例) (本研究繪製)

到AC電流與PV電流。

PV電壓指太陽能板輸入至變流器的電壓，單位為伏特(V)。由圖11與圖12可以得知冬季的電壓較高。

環境溫度為存放變流器的場域之室內環境的溫度，單位為攝氏( $^{\circ}\text{C}$ )。外殼溫度為變流

器外殼的溫度，單位為攝氏( $^{\circ}\text{C}$ )。反應器溫度為變流器中反應器的溫度，單位為攝氏( $^{\circ}\text{C}$ )。冷卻溫度為變流器中冷卻風扇的溫度，單位為攝氏( $^{\circ}\text{C}$ )。此四個溫度特徵互相呈高度正相關的關係。此外，溫度會影響到發電效率，溫度愈高，發電效率愈差；亦會影響變流器零件的壽命，溫度愈高，變流器零件的壽命愈縮減。圖13至圖20可以得知夏季的環境溫度、外殼溫度、反應器溫度、冷卻溫度較冬季來得高，因此夏季的發電效率低。

無效電能產生的原因來自於交流電電壓與電流的交互變化，單位為乏(VAR)。在一般情況下，無效電能的值會在0的附近，如圖21。圖22的12月4日與12月5日，無效電能的值高達165.4 VAR與195.3 VAR，當無效電能的值忽然升高時，表示變流器可能在短時間內產生故障

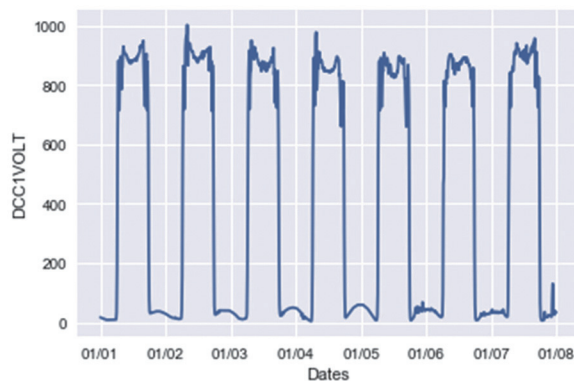


圖11 典型冬季一周七天之太陽能PV電壓曲線  
(以2015年2號變流器為例) (本研究繪製)

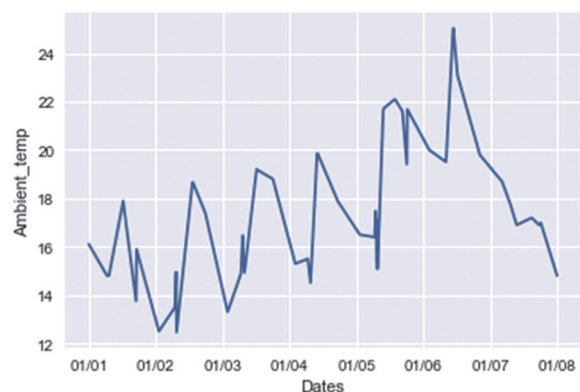


圖13 典型冬季一周七天之太陽能環境溫度曲線  
(以2015年2號變流器為例) (本研究繪製)

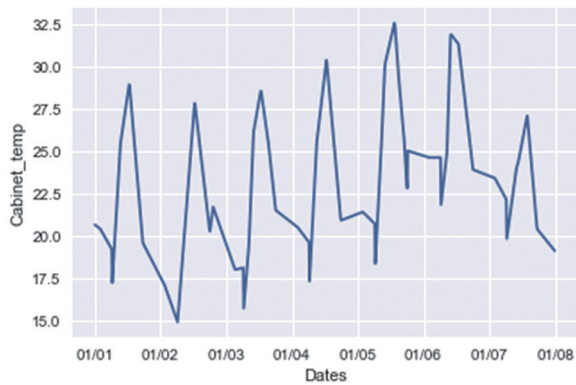


圖14 典型冬季一周七天之太陽能外殼溫度曲線(以2015年2號變流器為例) (本研究繪製)

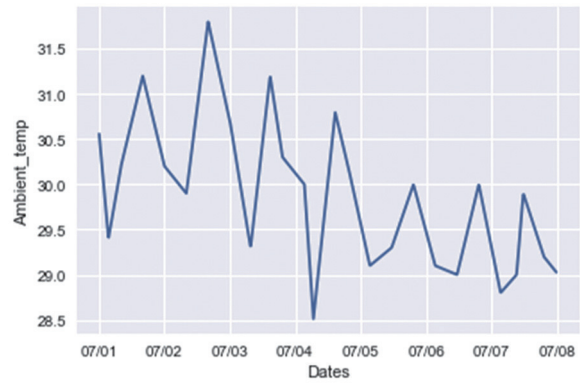


圖17 典型夏季一周七天之太陽能環境溫度曲線(以2015年2號變流器為例) (本研究繪製)

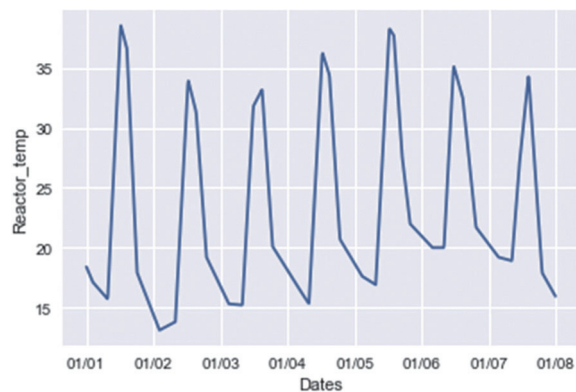


圖15 典型冬季一周七天之太陽能反應器溫度曲線(以2015年2號變流器為例) (本研究繪製)

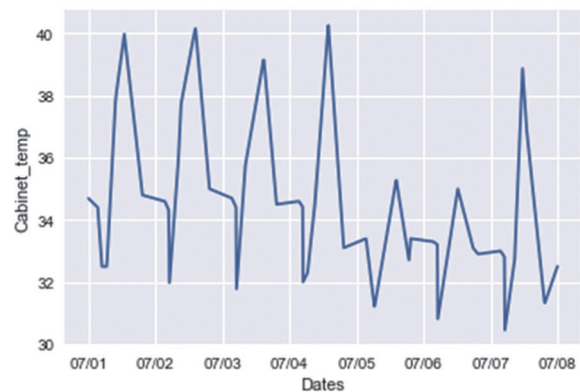


圖18 典型夏季一周七天之太陽能外殼溫度曲線(以2015年2號變流器為例) (本研究繪製)

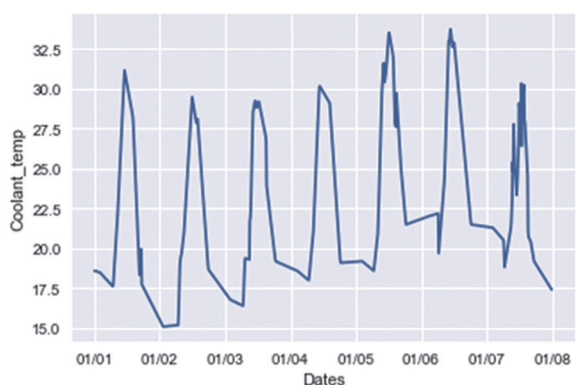


圖16 典型冬季一周七天之太陽能冷卻溫度曲線(以2015年2號變流器為例) (本研究繪製)

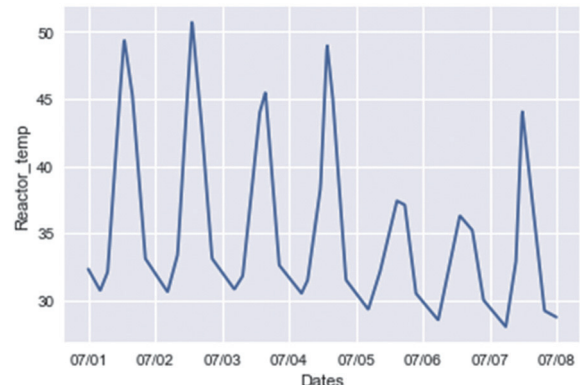


圖19 典型夏季一周七天之太陽能反應器溫度曲線(以2015年2號變流器為例) (本研究繪製)

的情形發生。

正常模式指變流器正負極的平均與接地的電壓差，單位為伏特(V)。圖23與圖24得知在變流器於一天中開始運轉與停止運轉時，正常模

式會大幅地上升。在冬季運轉時的正常模式會較夏季來的穩定。

圖3至圖24比較冬季與夏季有效功率、AC電流、頻率、今日效能、環境溫度、外殼



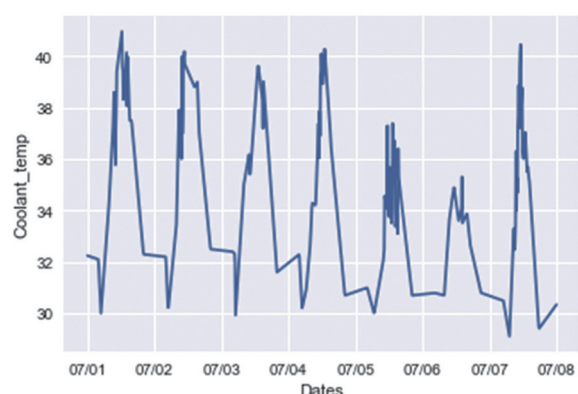


圖20 典型夏季一周七天之太陽能冷卻溫度曲線(以2015年2號變流器為例) (本研究繪製)

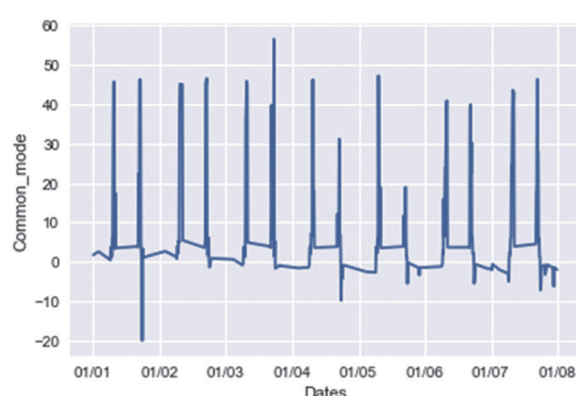


圖23 典型冬季一周七天之太陽能正常模式曲線(以2015年2號變流器為例) (本研究繪製)

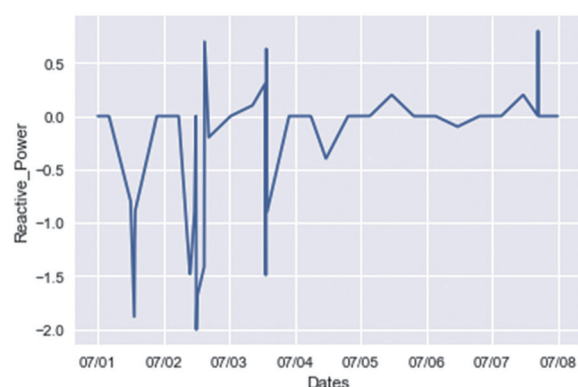


圖21 正常情形下之太陽能無效電能變化(以2015年2號變流器為例) (本研究繪製)



圖24 典型夏季一周七天之太陽能正常模式曲線(以2015年2號變流器為例) (本研究繪製)

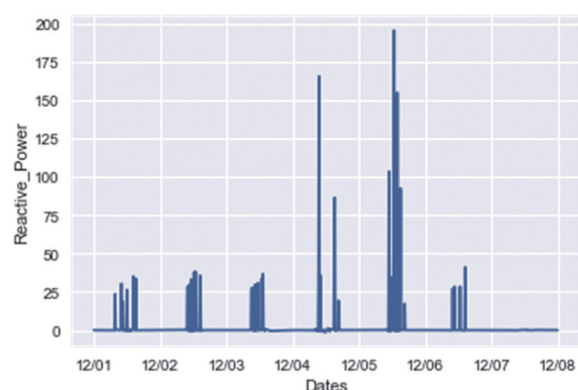


圖22 非正常情形下之太陽能無效電能變化(以2017年2號變流器為例) (本研究繪製)

溫度、正常模式、冷卻溫度、PV電流、PV電壓、反應器溫度等11個特徵的數值變化，以及正常與非正常情況下無效電能的數值變化。可以發現到季節對於變流器發電效率是有影響

的。此外，無效電能的值突然升高代表變流器可能於短時間內產生故障。另外，也得知這些特徵的數值是隨時在改變的，因此在變流器資料中總效能、有效功率、AC電流、頻率、今日效能、環境溫度、外殼溫度、正常模式、冷卻溫度、PV電流、PV電壓、反應器溫度、無效電能等13個特徵於連續時間內數值未改變，表示變流器可能發生故障，本研究在做資料預處理時將變流器狀態由False (假)調整為True (真)。

經過資料預處理的三個步驟後，各變流器正常資料與異常資料的資料筆數如表6。經過資料預處理後，將此資料使用3.3的兩個研究模型、3.4檢驗標準，實證結果於第四章實證。

表6 龍井一期各變流器資料預處理後正常/異常資料筆數(本研究整理)

變流器編號	正常資料筆數	異常資料筆數
2號	161,960筆	30,833筆
3號	141,835筆	50,958筆
4號	140,510筆	52,284筆
5號	150,779筆	42,110筆
6號	32,740筆	160,053筆
7號	43,651筆	149,142筆

### 3.3 建立研究模型

#### 3.3.1 模型1：因素分析與主成份分析

本研究使用SPSS 20軟體中的因素分析降低特徵的維度，輸入變數為有效功率、總效能、AC電流、頻率、今日效能、環境溫度、外殼溫度、正常模式、冷卻溫度、PV電流、PV電壓、無效電能、反應器溫度等13個特徵。在因素分析的方法中，SPSS軟體參數設定如下，首先使用KMO檢定與Bartlett球形檢定(IBM, 2019)，判斷該資料是否適合使用因素分析。通過KMO檢定與Bartlett球形檢定後，接著選用主成份分析作為萃取的方法。轉軸法的部分，本研究使用最大變異法，將彼此相關性高的特徵歸為一類。在選項的部分，本研究設定在同一成份中的所有特徵，每個特徵的因素負荷量絕對值必須大於0.6。最後會篩選出特徵值大於1的所有成份，並檢查這些成分的累積變異解釋量必須大於60%，表示篩選出的成份能夠代表母體的變異量。若通過累積變異解釋量大於60%的檢定，將篩選出成份的特徵組合，作為輸入變數，至模型2的四個機器學習演算法進行分類及檢驗。

#### 3.3.2 模型2：機器學習訓練

使用模型1萃取之因素，套入邏輯斯迴歸、支援向量機、隨機森林、K最鄰近法進行分類訓練，並使用3.4的兩項檢驗標準。第一項為混淆矩陣中最多正確分類異常資料的資料

筆數。第二項為精準度與F1分數，比較使用何種因素作為輸入變數與搭配哪種機器學習演算法進行訓練，會得出精準度與F1分數最佳之結果。

### 3.4 檢驗標準

#### 3.4.1 混淆矩陣

混淆矩陣(Confusion Matrix)是一種描述學習演算法效能的矩陣，包含「真陽性」(true positive, TP)表示預測結果為真，實際結果為真；「真陰性」(true negative, TN)表示預測結果為假，實際結果為假；「假陽性」(false positive, FP)表示預測結果為假，實際結果為真；「假陰性」(false negative, FN)表示預測結果為真，實際結果為假。換言之，真陽性、真陰性表示預測結果與實際結果一致，亦即正確分類；假陽性、假陰性表示預測結果與實際結果不同，亦即錯誤分類。利用混淆矩陣可以得知，經機器學習演算法之訓練後，正確分類及錯誤分類的資料筆數，示意圖如表7。

表7 混淆矩陣示意圖(本研究整理)

	預測結果	
	真陽性(TP)	假陰性(FN)
實際結果	假陽性(FP)	真陰性(TN)

#### 3.4.2 精準度、F1分數

精準度(precision)與F1分數(F1-score)為利用混淆矩陣計算出之數值，用來評估分類模

型的結果，精準度與F1分數的值介於0至1之間，愈接近1表示分類之準確率愈高(Raschk, 2016)。

精準度為樣本正確被分類為陽性的比率，精準度的公式如公式2：

$$PRE = \frac{TP}{TP + FP} \quad (2)$$

召回率，為真陽率，召回率的公式如公式3：

$$REC = \frac{TP}{FN + TP} \quad (3)$$

F1分數是使用精準度和召回率計算出的一個組合，F1分數的公式如公式4：

$$F1\text{-score} = 2 \frac{PRE + REC}{PRE \times REC} \quad (4)$$

## 4. 實證結果與討論

### 4.1 臺中龍井地區太陽能發電場一期2號至7號變流器實證結果解析

#### 4.1.1 臺中龍井地區太陽能發電場一期2號至7號變流器因素分析結果

從表8得知，4號變流器與7號變流器之KMO取樣適切性量，根據Kaiser (1974)觀點介於0.8至0.9之間，屬於良好，適合作因素分析。2號變流器、3號變流器、5號變流器與6號變流

器之KMO取樣適切性量，根據Kaiser (1974)觀點介於0.7至0.8之間，屬於中等，適合作因素分析。另外各組變流器之Bartlett球形檢定顯著性皆為0小於0.05，適合作因素分析。

從表9中得知龍井一期2號變流器共萃取4個特徵值大於1的共同因素，因素一為有效功率、PV電流、AC電流、PV電壓所組成，解釋變異量為41.348%；因素二為環境溫度、外殼溫度、冷卻溫度、反應器溫度所組成，解釋變異量為16.977%；因素三為正常模式組成，解釋變異量為9.177%；因素四為無效電能所組成，解釋變異量為7.730%。四個因素加總之累積解釋變異量為75.232%，能代表母體資料。

表8結果顯示2號變流器通過因素分析事前檢定。表9結果得到2號變流器之四個因素且累積解釋變異量足以代表母體資料。因此本研究將2號變流器之四個因素分別代入模型2機器學習演算法之分類及檢驗，比較哪一個因素及機器學習演算法之組合會得到最佳之結果。

從表10中得知龍井一期3號變流器共萃取5個特徵值大於1的共同因素，因素一為有效功率、PV電流、AC電流、PV電壓所組成，解釋變異量為35.833%；因素二為環境溫度、外殼溫度、反應器溫度所組成，解釋變異量為18.440%；因素三為頻率、今日效能、總效能所組成，解釋變異量為11.925%；因素四為正常模式、冷卻溫度所組成，解釋變異量為9.469%；因素五為無效電能組成，解釋變異量為7.832%，五個因素加總之累積解釋變異量為83.499%，能代表原始資料。

表8 個變流器KMO與Bartlett球形檢定結果(本研究整理)

	KMO取樣適切性量數	Bartlett球形檢定顯著性
2號變流器	0.786	0
3號變流器	0.744	0
4號變流器	0.82	0
5號變流器	0.792	0
6號變流器	0.741	0
7號變流器	0.828	0



表9 龍井一期2號變流器因素萃取結果(本研究整理)

	因素一	因素二	因素三	因素四
有效功率	0.954			
PV電流	0.950			
AC電流	0.930			
PV電壓	0.783			
環境溫度		0.900		
外殼溫度		0.863		
冷卻溫度		0.794		
反應器溫度		0.685		
總效能				
今日效能				
正常模式			0.898	
頻率				
無效電能				0.993
特徵值	5.375	2.207	1.193	1.005
解釋變異量(%)	41.348	16.977	9.177	7.730
累積解釋變異量(%)	41.348	58.326	67.502	75.232

表10 龍井一期3號變流器因素萃取結果(本研究整理)

	因素一	因素二	因素三	因素四	因素五
有效功率	0.970				
PV電流	0.966				
AC電流	0.964				
PV電壓	0.767				
環境溫度		0.914			
外殼溫度		0.777			
反應器溫度		0.717			
頻率			0.787		
今日效能			0.726		
總效能			0.699		
正常模式				0.904	
冷卻溫度				0.659	
無效電能					0.976
特徵值	4.658	2.397	1.550	1.231	1.018
解釋變異量(%)	35.833	18.440	11.925	9.469	7.832
累積解釋變異量(%)	35.833	54.273	66.198	75.667	83.499

表8結果顯示3號變流器通過因素分析事前檢定。表10結果得到3號變流器之五個因素且累積解釋變異量足以代表原始資料。因此本研究將3號變流器之五個因素分別代入機器學習演算法之分類及檢驗，比較哪一個因素及機器學習演算法之組合，會得到最佳化之結果。

從表11中得知共萃取3個特徵值大於1的共同因素，因素一為有效功率、AC電流、PV電流、PV電壓所組成，解釋變異量為45.999%；因素二為環境溫度、外殼溫度、冷卻溫度、反應器溫度、總效能所組成，解釋變異量為14.30%；因素三為正常模式組成，解釋變異量

表11 龍井一期4號變流器因素萃取結果(本研究整理)

	因素一	因素二	因素三
有效功率	0.953		
AC電流	0.950		
PV電流	0.948		
PV電壓	0.642		
環境溫度		0.907	
外殼溫度		0.896	
冷卻溫度		0.830	
反應器溫度		0.796	
總效能		0.666	
今日效能			
正常模式			0.639
頻率			
無效電能			
特徵值	5.980	1.859	1.120
解釋變異量(%)	45.999	14.300	8.619
累積解釋變異量(%)	45.999	60.298	68.917

為8.619%，三個因素加總之累積解釋變異量為68.917%，能代表母體資料。

表8結果顯示通過因素分析事前檢定。表11結果得到四個因素且累積解釋變異量足以代表原始資料。因此本研究將三個因素分別代入機器學習演算法之分類及檢驗，比較哪一個因素

及機器學習演算法之組合，會得到最佳化之結果。

從表12中得知共萃取4個特徵值大於1的共同因素，因素一為有效功率、PV電流、AC電流、PV電壓所組成，解釋變異量為39.677%；因素二為環境溫度、外殼溫度、反應器溫度所

表12 龍井一期5號變流器因素萃取結果(本研究整理)

	因素一	因素二	因素三	因素四
有效功率	0.899			
PV電流	0.892			
AC電流	0.885			
PV電壓	0.708			
環境溫度				
外殼溫度		0.919		
冷卻溫度		0.896		
反應器溫度		0.809		
總效能				
今日效能				
正常模式			0.803	
頻率				
無效電能				0.974
特徵值	5.158	1.861	1.117	1.003
解釋變異量(%)	39.677	14.313	8.591	7.713
累積解釋變異量(%)	39.677	53.991	62.582	70.295

組成，解釋變異量為14.313%；因素三為正常模式組成，解釋變異量為8.591%；因素四為無效電能所組成，解釋變異量為7.713，四個因素加總之累積解釋變異量為70.295%，能代表母體資料。

表8結果顯示通過因素分析事前檢定。表12結果得到四個因素且累積解釋變異量足以代表原始資料。因此本研究將四個因素分別代入機器學習演算法之分類及檢驗，比較哪一個因素及機器學習演算法之組合，會得到最佳化之結果。

從表13中得知共萃取4個特徵值大於1的共同因素，因素一為有效功率、AC電流、PV電流所組成，解釋變異量為39.985%；因素二為外殼溫度、環境溫度、冷卻溫度、反應器溫度所組成，解釋變異量為26.327%；因素三為PV電壓、正常模式、今日效能、總效能組成，解釋變異量為11.323%；因素四為無效電能所組成，解釋變異量為8.045，四個因素加總之累積解釋變異量為85.680%，能代表母體資料。

表8結果顯示通過因素分析事前檢定。表

13結果得到四個因素且累積解釋變異量足以代表原始資料。因此本研究將四個因素分別代入模型2機器學習演算法之分類及檢驗，比較哪一個因素及機器學習演算法之組合，會得到最佳化之結果。

從表14中得知共萃取4個特徵值大於1的共同因素，因素一為有效功率、PV電流、AC電流、冷卻溫度、反應器溫度所組成，解釋變異量為47.976%；因素二為環境溫度、外殼溫度、冷卻溫度、反應器溫度所組成，解釋變異量為15.840%；因素三為正常模式、總效能、PV電壓、今日效能所組成，解釋變異量為13.194%；因素四為無效電能所組成，解釋變異量為7.706，四個因素加總之累積解釋變異量為84.716%，能代表母體資料。

表8結果顯示通過因素分析事前檢定。表14結果得到四個因素且累積解釋變異量足以代表母體資料。因此本研究將四個因素分別代入模型2機器學習演算法之分類及檢驗，比較哪一個因素及機器學習演算法之組合，會得到最佳化之結果。

表13 龍井一期6號變流器因素萃取結果(本研究整理)

	因素一	因素二	因素三	因素四
有效功率	0.966			
PV電流	0.962			
AC電流	0.959			
環境溫度		0.935		
外殼溫度		0.923		
冷卻溫度		0.890		
反應器溫度		0.761		
PV電壓			-0.912	
正常模式			0.886	
今日效能			0.691	
總效能			-0.667	
頻率				
無效電能				0.981
特徵值	5.198	3.423	1.472	1.046
解釋變異量(%)	39.985	26.327	11.323	8.045
累積解釋變異量(%)	39.985	66.312	77.635	85.680



表14 龍井一期7號變流器因素萃取結果(本研究整理)

	因素一	因素二	因素三	因素四
有效功率	0.971			
PV電流	0.968			
AC電流	0.966			
環境溫度		0.949		
外殼溫度		0.822		
冷卻溫度	0.609	0.725		
反應器溫度	0.650	0.675		
頻率				
正常模式			0.856	
總效能			0.844	
PV電壓			0.841	
今日效能			0.666	
無效電能				0.999
特徵值	6.237	2.059	1.715	1.002
解釋變異量(%)	47.976	15.840	13.194	7.706
累積解釋變異量(%)	47.976	63.816	77.010	84.716

#### 4.1.2 臺中龍井地區太陽能發電場 一期2號至7號變流器資料分 析結果

表15為龍井一期2號變流器至7號變流器經過特徵選取選出的輸入特徵組合、搭配使用的機器學習演算法訓練資料後，得到最多正確分

類異常資料筆數、錯誤分類異常資料筆數、精準度、F1分數。正確分類異常資料筆數與錯誤分類異常資料筆數來自混淆矩陣的結果。精準度與F1分數的計算方式，包括正確分類異常資料筆數、錯誤分類異常資料筆數、正確分類正常資料筆數、錯誤分類正常資料筆數。

由表15可得知，2號變流器以環境溫度、

表15 各變流器特徵選取及其異常資料分析結果(本研究整理)

變流器編號	輸入特徵	最佳之機器學習演算法	正確分類異常資料之筆數	錯誤分類異常資料之筆數	精準度	F1分數
2號	環境溫度、外殼溫度、冷卻溫度、反應器溫度	邏輯斯迴歸	6,002	218	0.0046	0.0092
3號	頻率、今日效能、總效能	邏輯斯迴歸	10,263	0	0.0002	0.0004
4號	有效功率、AC電流、PV電流、PV電壓	隨機森林	6,030	4,554	1.000	0.9247
5號	外殼溫度、冷卻溫度、反應器溫度	邏輯斯迴歸	8,437	0	0.0007	0.0007
6號	正常模式	邏輯斯迴歸	32,005	0	0	0
	無效電能	邏輯斯迴歸	32,005	0	0	0
7號	正常模式、總效能、PV電壓、今日效能	邏輯斯迴歸	29,882	0	0.0003	0.0003
	無效電能	邏輯斯迴歸	29,882	0	0.0023	0.0023

外殼溫度、冷卻溫度、反應器溫度作為輸入變數，並使用邏輯斯迴歸作為訓練模型，得到6,002筆正確分類的異常資料、218筆錯誤分類的異常資料、149筆正確分類的正常資料、32,190筆錯誤分類的正常資料，精準度與F1分數分別僅有0.46%與0.92%。表示該組合僅適用於分類異常資料，不適合用於分類正常資料。

由表15可得知，3號變流器以頻率、今日效能、總效能作為輸入變數，並使用邏輯斯迴歸作為訓練模型，得到10,263筆正確分類的異常資料、0筆錯誤分類的異常資料、5筆正確分類的正常資料、28,291筆錯誤分類的正常資料，精準度與F1分數僅有0.02%與0.04%。表示該組合僅適用於分類異常資料，不適合用於分類正常資料。

由表15可得知，4號變流器以有效功率、AC電流、PV電流、PV電壓作為輸入變數，並使用隨機森林作為訓練模型，得到6,030筆正確分類的異常資料、4,554筆錯誤分類的異常資料、27,975筆正確分類的正常資料、0筆錯誤分類的正常資料，精準度為100%，F1分數為92.47%。表示該組合非常適合用於分類正常資料，但是分類異常資料的結果表現普通。

由表15得知，5號變流器以外殼溫度、冷卻溫度、反應器溫度作為輸入變數，並使用邏輯斯迴歸作為訓練模型，得到8,437筆正確分類的異常資料、0筆錯誤分類的異常資料、21筆正確分類的正常資料、30,120筆錯誤分類的正常資料，精準度與F1分數分別僅有0.07%與

0.07%。表示該組合僅適用於分類異常資料，不適合用於分類正常資料。

由表15可得知，6號變流器分別以正常模式與無效電能作為輸入變數，並使用邏輯斯迴歸作為訓練模型，得到32,005筆正確分類的異常資料、0筆錯誤分類的異常資料、0筆正確分類的正常資料、6,554筆錯誤分類的正常資料，精準度為0%，F1分數為0%。表示該組合能夠完全正確地分類異常資料，無法正確地分類正常資料。

由表15可得知，7號變流器以正常模式、總效能、PV電壓、今日效能作為輸入變數，並使用邏輯斯迴歸作為訓練模型，得到29,882筆正確分類的異常資料、0筆錯誤分類的異常資料、3筆正確分類的正常資料、8,674筆錯誤分類的正常資料，精準度為0.03%，F1分數0.03%。另外，7號變流器以無效電能作為輸入變數，並使用邏輯斯迴歸作為訓練模型，得到29,882筆正確分類的異常資料、0筆錯誤分類的異常資料、20筆正確分類的正常資料、8,657筆錯誤分類的正常資料，精準度為0.23%，F1分數0.23%。此兩個組合，適合用於分類異常資料，不適合用於分類正常資料。

表16呈現透過公式3.1、公式3.2與公式3.3進行分析2號至7號變流器精準度與F1分數之最佳結果，在比較精準度與F1分數時，訓練結果必須要符合正確分類正常資料的資料筆數，大於錯誤分類正常資料的資料筆數，且正確分類異常資料的資料筆數，也必須大於錯誤分類異

表16 各變流器整合正常與異常資料分析結果之精準度與F1分數(本研究整理)

變流器編號	輸入變數/特徵	最佳之機器學習演算法	精準度	F1分數
2號	有效功率、PV電流、AC電流、PV電壓	邏輯斯迴歸	0.9488	0.9704
3號	無效電能	支援向量機	0.8333	0.9035
4號	有效功率、AC電流、PV電流、PV電壓	隨機森林	1.000	0.9247
5號	無效電能	邏輯斯迴歸	0.8166	0.9036
6號	無效電能	支援向量機	0.3318	0.5025
7號	正常模式、總效能、PV電壓、今日效能	隨機森林	0.9714	0.9821

常資料的資料筆數。

由表16可得知，2號變流器以有效功率、PV電流、AC電流、PV電壓作為輸入變數，並使用邏輯斯迴歸作為訓練模型，得到30684筆正確分類的正常資料、1,655筆錯誤分類的正常資料、6,002筆正確分類的異常資料、218筆錯誤分類的異常資料，精準度為94.88%，F1分數為97.04%。為2號變流器的16個組合中，最佳的精準度與F1分數結果。

由表16可得知，3號變流器以無效電能作為輸入變數，並使用支援向量機作為訓練模型，得到23,654筆正確分類的正常資料、4,732筆錯誤分類的正常資料、9,946筆正確分類的異常資料、317筆錯誤分類的異常資料，精準度為83.33%，F1分數為90.35%。為3號變流器的20個組合中，最佳的精準度與F1分數結果。

由表16可得知，4號變流器以有效功率、AC電流、PV電流、PV電壓作為輸入變數，並使用隨機森林作為訓練模型，得到27,975筆正確分類的正常資料、0筆錯誤分類的正常資料、6,030筆正確分類的異常資料、4,554筆錯誤分類的異常資料，精準度為100%，F1分數為92.47%。為4號變流器的12個組合中，最佳的精準度與F1分數結果。

由表16可得知，5號變流器以無效電能作為輸入變數，並使用邏輯斯迴歸作為訓練模型，得到26,816筆正確分類的正常資料、3,325筆錯誤分類的正常資料、6,035筆正確分類的異常資料、2,402筆錯誤分類的異常資料，精準度為81.66%，F1分數為90.36%。為5號變流器的16個組合中，最佳的精準度與F1分數結果。

由表16可得知，6號變流器以無效電能作為輸入變數，並使用支援向量機作為訓練模型，得到2,212筆正確分類的正常資料、4,342筆錯誤分類的正常資料、31,967筆正確分類的異常資料、38筆錯誤分類的異常資料，精準度為33.18%，F1分數為50.25%。為6號變流器的16個組合中，最佳的精準度與F1分數結果。

由表16可得知，7號變流器以正常模式、

總效能、PV電壓、今日效能作為輸入變數，並使用隨機森林作為訓練模型，得到8,429筆正確分類的正常資料、248筆錯誤分類的正常資料、29,830筆正確分類的異常資料、59筆錯誤分類的異常資料，精準度為97.14%，F1分數為98.21%。為7號變流器的16個組合中，最佳的精準度與F1分數結果。

## 5. 結 論

太陽能發電量受到氣候、溫度、日照長度等環境因素影響，因此相較於火力發電、核能發電，太陽能發電具有不穩定性。除此之外，太陽能發電系統機器設備的健康狀態也會造成太陽能發電量不如預期。為了解決來自於太陽能發電系統中變流器(inverter)健康狀態的不確定性，本研究使用機器學習的方法，對變流器作異常狀態的偵測，找出何種特徵組合能夠判別變流器的正常運作或是發生故障。

本研究的實證資料來自於台灣電力公司再生能源處提供之臺中龍井一期太陽能電場變流器歷史發電數據。資料時間自2015年1月1日0時0分起至2018年10月1日0時0分止。資料特徵包含以下15個：時間、有效功率、總效能、AC電流、頻率、今日效能、變流器狀態、環境溫度、外殼溫度、正常模式、冷卻溫度、PV電流、PV電壓、無效電能、反應器溫度。

研究方法分為三個步驟。第一步為資料預處理，由於1號變流器所有資料的數值皆相同，屬於無效資料，因此本研究予以刪除。接著，刪除2號變流器至7號變流器中的無效資料(數值為Bad值)，出現Bad值表示該時間點資料的傳送與接收可能產生問題，而數值的出現會影響資料無法做機器學習的分析，因此予以刪除。第二步為利用因素分析與主成份分析降低資料的維度，找出彼此相關性高的特徵，用來解釋母體資料的變異量。第三步驟為利用邏輯斯迴歸、支援向量機、隨機森林、K最鄰近法將每一台變流器資料做分類訓練，並以精準度、F1



分數比較各分類模型的結果，最佳結果的輸入特徵即為判斷每一台變流器是否故障的關鍵特徵。

研究結果分為三個部分，第一部分為正確分類異常資料的資料筆數。2號變流器以環境溫度、外殼溫度、冷卻溫度、反應器溫度作為輸入特徵，使用邏輯斯迴歸作為訓練模型，得到最多正確分類異常資料的資料筆數。3號變流器以頻率、今日效能、總效能作為輸入特徵，使用邏輯斯迴歸作為訓練模型，得到最多正確分類異常資料的資料筆數。4號變流器以有效功率、AC電流、PV電流、PV電壓作為輸入特徵，使用隨機森林作為訓練模型，得到最多正確分類異常資料的資料筆數。5號變流器以外殼溫度、冷卻溫度、反應器溫度作為輸入特徵，使用邏輯斯迴歸作為訓練模型，得到最多正確分類異常資料的資料筆數。6號變流器分別以正常模式、無效電能作為輸入特徵，使用邏輯斯迴歸作為訓練模型，皆會得到最多正確分類異常資料的資料筆數。7號變流器以正常模式、總效能、PV電壓、今日效能作為輸入特徵，使用邏輯斯迴歸作為訓練模型，得到最多正確分類異常資料的資料筆數。

第二部分為精準度與F1分數的最佳結果。2號變流器以有效功率、PV電流、AC電流、PV電壓作為輸入特徵，使用邏輯斯迴歸作為訓練模型，得到精準度為94.88%、F1分數為97.04%。3號變流器以無效電能作為輸入特徵，使用支援向量機作為訓練模型，得到精準度為83.33%、F1分數為90.35%。4號變流器以有效功率、AC電流、PV電流、PV電壓作為輸入特徵，得到精準度為100%、F1分數為92.47%。5號變流器以無效電能作為輸入特徵，使用邏輯斯迴歸作為訓練模型，得到精準度為81.66%、F1分數為90.36%。6號變流器以無效電能作為輸入特徵，使用支援向量機作為訓練模型，得到精準度為33.18%、F1分數為50.25%。7號變流器以正常模式、總效能、PV電壓、今日效能作為輸入特徵，使用隨機森林

作為訓練模型，得到精準度為97.14%、F1分數為98.21%。

第三部分為6號變流器之精準度為33.18%、F1分數為50.25%，其表現明顯低於其他各組變流器，經查因為正確分類的正常資料只有2,212筆而錯誤分類的正常資料為4,342筆，造成正確分類的正常資料筆數少而錯誤分類的正常資料導致精準度不足，進而造成F1分數較低，故建請針對6號變流器之相關硬體進行檢測，以釐清是否為硬體異常或者是演算法模型需要進行調整。

相較於Alam (2014)使用SSTDR透過反射法對於傳輸線的電器特性進行故障偵測。陳建融(2016)使用K-Means、EM、DBSCAN、OPTICS針對DC\_I、DC\_KW、RADIATION、DC\_I、RADIATION等5項變數分群後進行分析。本研究則是透過邏輯斯迴歸、支援向量機、隨機森林、K最鄰近法等機器學習演算法，針對龍井太陽能光電廠1-1期2015年至2018年六組變流器之時間、有效功率、總效能、AC電流、頻率、今日效能、變流器狀態、環境溫度、外殼溫度、正常模式、冷卻溫度、PV電流、PV電壓、無效電能、反應器溫度等14項變數進行資料分析與預測。透過混淆矩陣進行分類，最後在透過精準度與F1分數進行評定，其研究結果以4號變流器使用隨機森林做分群訓練，得到精準度100%與F1分數為92.47%為最佳表現。因此相較於前兩者，本研究針對變流器之相關狀況，能有更進一步的掌握，而研究結果也顯示，精準度也能有100%的表現。

在研究結果中得知下列兩點：

第一點，本研究以有效功率、總效能、AC電流、頻率、今日效能、環境溫度、外殼溫度、正常模式、冷卻溫度、PV電流、PV電壓、無效電能、反應器溫度作為特徵選取的輸入變數。做完特徵選取後發現環境溫度、外殼溫度、冷卻溫度、反應器溫度彼此的相關性高。原因為當變流器開始運轉時反應器溫度會上升，接著外殼溫度、冷卻溫度會隨之上升，

在變流器運轉會需要散熱，因此變流器周遭的環境溫度緩慢地上升。另外，有效功率、AC電流、PV電流、PV電壓彼此的相關性也很高。原因是PV電流為輸入變流器的電流，會影響到輸出的電流(AC電流)，而有效功率的計算方式與PV電流、PV電壓的數值有關。

第二點，本研究以相同的方法對每一台變流器做特徵選取、機器學習分類訓練，但是得到的結果卻不同。原因為每一台變流的運行狀態皆不同，分別獲得不一樣的發電數據。

本研究實證結果顯示：

1. 各變流器的健康狀態，均適合使用邏輯斯迴歸為初步診斷工具；再進一步與其他工具做評估比較，以利尋找合宜的訓練工具。本實驗中，沒有任何樣本適用K最鄰近法。
2. 本研究發現，4號變流器與7號變流器之KMO取樣適切性量，介於0.8至0.9之間，屬於良好，適合作因素分析；累積解釋變異量偏低的4號變流器與偏高的7號變流器，均合適隨機森林方法。後來的精準度與F1分數都相對顯著，應證了Kaiser (1974)觀點。
3. 在實證結果中，2號變流器、3號變流器、5號變流器與6號變流器之KMO取樣適切性量，介於0.7至0.8之間，屬於中等。使用支援向量機方法的2號變流器與6號變流器，在精準度與F1分數方面呈現差異，發現6號變流器輸入特徵缺乏溫度與電壓相關特徵。

在此情況下，每一台變流器得到的結果不相同。若採用定期維修決策，固定一段時間對該場所全部變流器進行例行檢測、保養，會浪費不必要的維修成本。因此，採用預測性維修策略，預測每一台變流器可能發生故障的時間，再依照預測的結果制定維修計畫，能夠降低維修成本。

另外，由於太陽能板與相關硬體設備之特性，會隨著時間與環境進行改變，因此一個月內，連續三次判斷異常需要台電維護人員進行維護，但維護人員實際查看相關設備後並沒有發現異常。此時應該進行模型的修正，找出目

前最適合的模型。避免人工的浪費，進一步降低維護成本。

本研究建議：

1. 台電公司宜提前對6號變流器進行預防性維修，其他變流器可繼續按照檢修週期保養。
2. 比起定期維修策略，若採用預測性維修策略，預測每一台變流器可能發生故障的時間，再依照預測的結果制定維修計畫，將能夠大幅降低維修成本。

## 誌 謝

作者感謝科技部「新興科技創新營運模式:智慧製造與資訊系統整合創新營運模式建置與維運計畫」整合型研究計畫(編號109-2425-H-005-001-)及科技部臺灣地區太陽光電及風電未來96小時逐時預測系統之建置與測試(編號108-3116-F-005-001-)之部分經費支持，始得完成此研究。同時，感謝台電再生能源處提供相關數據，以及兩位匿名評審提供的寶貴建議，惟文中若有任何疏誤，應由作者們自負文責。

## 參考文獻

- Raschk, S. (劉立民、吳建華譯)，2016。Python機器學習(初版)。新北市：博碩文化(原著初版年代：2015)。
- 王柏勝，2017。以資料探勘技術預測太陽能發電與需量反映方案之整合應用(未出版之碩士論文)。國立中興大學資訊管理學系，臺中市。
- 王群翔，2018。以機器學習預測太陽能發電量：龍井太陽能電廠案例(未出版之碩士論文)。國立中興大學資訊管理學系，臺中市。
- 台灣電力公司，2019。太陽光電發電，2019年7月16日檢自：<https://www.taipower.com.tw/tc/page.aspx?mid=204&cid=>

- [1582&cchk=5b8ce619-7ff5-40e9-9032-bdfd93d197d9](#)。
- 林清山，1991。多變項分析統計法。臺北市：東華書局。
- 許志義、林冠成、王柏勝、蔡松諺與郭晉源，2018。臺灣太陽能發電資料探勘與預測模型建置與應用。台電工程月刊，第839期，頁8-24，臺北市，台灣電力公司。
- 陳建融，2016。資料科學與應用—以太陽能發電異常偵測為例(未出版之碩士論文)。國立雲林科技大學資訊管理系，雲林縣。
- 張家源，2018。以資料探勘技術預測風力發電量之研究：彰工風力發電站案例(未出版之碩士論文)。國立中興大學資訊管理學系，臺中市。
- 蔡松諺，2017。以資料探勘方法預測臺灣地區太陽能發電量之研究(未出版之碩士論文)。國立中興大學資訊管理學系，臺中市。
- 傅粹馨，2002。主成份分析和共同因素分析相關議題之探究，*教育與社會研究*(3)，107-131。
- Alam, M. K., F. H. Khan, J. Johnson and J. Flicher, 2014. PV arc-fault detection using spread spectrum time domain reflectometry(SSTDR). In *2014 IEEE Energy Conversion Congress and Exposition(ECCE)*, pp.3294-3300. Retrieved from: <https://ieeexplore.ieee.org/document/6953848>.
- Braun, H., S. T. Buddha, V. Krishnan, A. Spanias, C. Tepedelenlioglu, T. Yeider and T. Takehara, 2012. Signal Processing for fault detection in photovoltaic arrays. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1681-1684. Retrieved from: <https://ieeexplore.ieee.org/document/6288220>.
- Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6), 417.
- IBM, 2019. KMO and Bartlett's test. Retrieved July 30,2019 from: [https://www.ibm.com/support/knowledgecenter/en/SSLVMB\\_subs/statistics\\_casestudies\\_project\\_ddita/spss/tutorials/fac\\_telco\\_kmo\\_01.html](https://www.ibm.com/support/knowledgecenter/en/SSLVMB_subs/statistics_casestudies_project_ddita/spss/tutorials/fac_telco_kmo_01.html).
- Kaiser, H. F., 1974. An Index of factorial simplicity. *Psychometrika*, 39, pp.31-36.
- Mobley, R. K., 2002. *An introduction to Predictive Maintenance* (Second Edition). Elsevier Science: Massachusetts.
- Pearson, K., 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559-572.

# Detecting False Alarm by Using Machine Learning: Case of Longjing Solar Power Station

Jyh-Yih Hsu<sup>1\*</sup>    Tien-Chia Ku<sup>2</sup>    Zhi-Hao Zhang<sup>2</sup>  
Fa-Ming Yeh<sup>3</sup>    Chung-Chi Chen<sup>3</sup>

## ABSTRACT

Compared with thermal power generation and nuclear power generation, solar power generation is more sensitive to environmental factors such as climate, temperature, and sunshine length, and the power generation is characterized by instability. In addition, the health status of the machinery and equipment of the solar power generation system will also affect the power generation, especially from the uncertainty caused by the health status of the inverter in the solar power generation system. This research uses machine learning to detect anomalies in the inverter, so as to grasp the health status of the inverter in real time and provide Taipower Company as a reference as one of the countermeasures to reduce the uncertainty of power generation. This research uses the historical data of Taichung Longjing Phase I power generation provided by the Renewable Energy Department of Taipower Company, and applies the fault diagnosis method of predictive maintenance to detect abnormal conditions of the inverter. This research first uses factor analysis, principal component analysis and other methods to extract characteristic factors for empirical data, and then uses Logistic Regression, Support Vector Machine, Random Forest, K-Nearest Neighbor method and other methods to train the classification model. Here, the criteria for evaluating the classification model include: the number and accuracy of the most correctly classified abnormal data, and the F1 score. The empirical results of this study show:

1. The health status of each inverter is suitable for using Logistic Regression as a preliminary diagnostic tool; further evaluation and comparison with other tools are made to facilitate the search for suitable training tools. In this experiment, none of the samples applied the K-Nearest Neighbor method.
2. This study found that the appropriateness of KMO sampling for No. 4 inverter and No. 7 inverter is between 0.8 and 0.9, which is good and suitable for factor analysis; the cumulative explanation of No. 4 variable is low Both the current inverter and the higher inverter No. 7 are suitable for Random Forest method. Later accuracy and F1 scores were relatively significant, which confirmed Kaiser's (1974) viewpoint.
3. In the empirical results, the appropriateness of KMO sampling for No. 2 inverter, No. 3. inverter, No. 5 inverter and No. 6 inverter is between 0.7 and 0.8, which is medium. The No. 2 inverter and the No. 6 inverter using the Support Vector Machine method show differences in accuracy and F1 scores. It is found that the input characteristics of the No. 6 inverter lack temperature and voltage related characteristics.

This research suggests:

1. It is recommended that Taipower Company conduct preventive maintenance on the No. 6 inverter in advance, and other inverters can continue to be maintained in accordance with the maintenance cycle.
2. Compared with the regular maintenance strategy, if a predictive maintenance strategy is adopted to predict the possible failure time of each inverter, and then to formulate a maintenance plan based on the predicted results, the maintenance cost will be greatly reduced.

**Keywords:** Solar Power Generation, Predictive Maintenance, Fault Diagnosis, Fault Detection, Factor Analysis, Principal Component Analysis, Logistic Regression, Support Vector Machine, Random Forest, K-Nearest Neighbor.

<sup>1</sup>Distinguished Chair Professor, National Taipei University of Business; Professor, Department of Management Information Systems and Department of Applied Economics, National Chung-Hsing University.

<sup>2</sup>Master, Department of Management Information Systems, National Chung Hsing University.

<sup>3</sup>PhD Student, Department of Applied Economics, National Chung Hsing University.

\* Corresponding Author, Phone: +886-4-2285-0505, E-mail: [hsu@nchu.edu.tw](mailto:hsu@nchu.edu.tw)

Received Date: September 17, 2019

Revised Date: November 26, 2020

Accepted Date: January 5, 2021