國立雲林科技大學工業工程與管理系

碩士論文

Department of Industrial Engineering and Management
National Yunlin University of Science & Technology

Master Thesis

基於 LSTM 方法的風力渦輪機發電量預測

Power Generation Prediction of Wind Turbines Based on LSTM Method

黃堉豪

Yu-Hao Huang

指導教授：洪鈺欣 博士

Advisor：Yu-Hsin Hung, Ph.D.

中華民國 112 年 6 月

June 2023

摘要

過去因為人類對於石化燃料的過度依賴，導致氣候變遷加劇、生態環境的過度破壞，近年，環保意識抬頭，各國開始簽訂國際協定並積極發展綠色能源以達到永續發展的目的，而在眾多綠能發電中，以被視為是最能達到永續發展目的的風力發電為綠能發電主流，除了發電機組的改良提升，也將人工智慧引進該領域來進行發電功率的預測，過往預測中，學者多以風速、風向、氣壓等環境特徵作為輸入藉此進行發電功率的預測。然而，因為風速是一種隨機且不穩定的數據，需要耗費大量的計算跟時間成本進行前處理，導致模型運算成本增加。有鑑於此，本研究提出利用 SCADA 的風力渦輪機狀態數據作為主要輸入變量且在不考慮外部環境因素的影響下結合相互資訊與 LSTM 進行風力發電功率預測。根據本研究分析之結果，26 個特徵組合為對於發電量預測的績效為最佳，R^2、MAE 分別為 0.96 與 62.98，結果表明 SCADA 的風力渦輪機狀態數據對於發電量的預測是有效的。

關鍵字：風力渦輪機、長短期記憶模型、深度學習、Mutual Information

Abstract

In the past, climate change has intensified, and the ecological environment has been excessively damaged due to people's overdependence on fossil fuels. Recently, environmental awareness has increased, and countries are signing international agreements and actively developing green energy to achieve sustainable development. Meanwhile, among many green power generations, wind power generation, considered the most sustainable development goal, is the mainstream of green energy. In addition to improving generator sets, artificial intelligence is introduced into this field to predict power generation. Moreover, in previous predictions, scholars primarily used environmental characteristics, such as wind speed, wind direction, and air pressure, as inputs to expect power generation. However, wind speed is random and unstable data. It requires a lot of calculations and time for preprocessing, which will increase the calculation costs of a model. Considering this, this study proposes using SCADA data on wind turbine status as the primary input variable and combines mutual information and long short-term memory to predict wind power generation without considering the influence of external environmental factors. According to the analysis of this study, the performance of the 26 selected feature combinations for predicting power generation is optimal, with R^2 and MAE values of 0.96 and 62.98, respectively. These results indicate that the SCADA data on wind turbine conditions are effective in predicting power generation.

Keywords: Wind turbines, Long Short-term Memory (LSTM), Deep learning, Mutual Information

目錄

摘要 i
Abstract ii
目錄 iii
表目錄 v
圖目錄 vi
第一章　緒論 1
1.1 研究背景 1
1.2 研究動機 2
1.3 研究目的 3
1.4 研究流程 4
第二章　文獻探討 5
2.1 風力渦輪機簡介 5
2.2 風力發電功率預測方法簡介 5
2.3 相互資訊(Mutual Information, MI) 6
2.4 長短期記憶神經網路(Lomg Short-Term Memory, LSTM) 7
2.5 相關 SCADA 數據分析預測之文獻 8
2.6 小結 9
第三章　研究方法 10
3.1 資料分析流程 10
3.2 資料集介紹 12
3.3 資料前處理 15
3.3.1 盒鬚圖識別異常值與線性插值法 15
3.3.2 平穩性檢驗 17
3.3.4 數值縮放 19
3.4 特徵選擇 19
3.4.1 相互資訊(Mutual Information, MI) 19

3.5 模型建立 20
3.5.1 長短期記憶(Lomg Short-Term Memory, LSTM) 20
3.5.2 參數調整 22
3.6 驗證與績效評估 23
3.6.1 模型驗證流程 23
3.6.2 驗證機制 24
3.6.3 評估指標 25
第四章　研究結果 27
4.1 資料集說明 27
4.2 資料前處理 30
4.2.1 盒鬚圖識別異常值與線性插值法 30
4.2.2 平穩性檢測與轉換 34
4.2.2 數值縮放 35
4.3 特徵選擇 36
4.4 模型建立 39
4.5 績效評估 40
4.5.1 發電量預測結果與分析(重要性大於 1.0 特徵) 40
4.5.2 發電量預測結果與分析(重要性大於 0.9 特徵) 41
4.5.3 發電量預測結果與分析(重要性大於 0.8 特徵) 42
4.5.4 發電量預測結果與分析(重要性大於 0.7 特徵) 43
4.5.5 發電量預測結果與分析(重要性大於 0.6 特徵) 44
4.5.6 發電量預測結果與分析(重要性大於 0.5 特徵) 45
第五章　討論 47
第六章　結論與未來研究 48
6.1 結論 48
6.2 未來研究 48
參考文獻 50

表目錄

表 1 各特徵與目標值的統計資料 12
表 2 平穩性種類的說明 18
表 3 特徵值欄位說明 27
表 4 目標值欄位說明 29
表 5 重新設定後的特徵值欄位說明 31
表 6 發電量 ADF 檢驗結果 34
表 7 發電量 KPSS 檢驗結果 34
表 8 資料集拆分結果 39
表 9 參數的設定與範圍 39
表 10 重要性大於 1.0 特徵之參數調整結果 40
表 11 重要性大於 1.0 特徵之績效表 40
表 12 重要性大於 0.9 特徵之參數調整結果 41
表 13 重要性大於 0.9 特徵之績效表 41
表 14 重要性大於 0.8 特徵之參數調整結果 42
表 15 重要性大於 0.8 特徵之績效表 42
表 16 重要性大於 0.7 特徵之參數調整結果 43
表 17 重要性大於 0.7 特徵之績效表 43
表 18 重要性大於 0.6 特徵之參數調整結果 44
表 19 重要性大於 0.6 特徵之績效表 44
表 20 重要性大於 0.5 特徵之參數調整結果 45
表 21 重要性大於 0.5 特徵之績效表 45

圖目錄

圖 1 2021 年全球風能建置容量比例 2
圖 2 研究流程架構圖 4
圖 3 資料分析流程圖 11
圖 4 盒鬚圖示意圖 16
圖 5 變數之間的關係圖 20
圖 6 LSTM 內部結構示意圖 21
圖 7 模型驗證表示圖 23
圖 8 k 折交叉驗證示意圖 24
圖 9 盒鬚圖異常值檢測圖 30
圖 10 線性補值後結果示意圖 31
圖 11 正規化前的發電量趨勢圖 35
圖 12 正規化後的發電量趨勢圖 35
圖 13 重要性大於 1.0 的特徵圖 36
圖 14 重要性大於 0.9 的特徵圖 36
圖 15 重要性大於 0.8 的特徵圖 37
圖 16 重要性大於 0.7 的特徵圖 37
圖 17 重要性大於 0.6 的特徵圖 38
圖 18 重要性大於 0.5 的特徵圖 38
圖 19 人機介面模擬圖 49
圖 20 人機介面結果產出模擬圖 49

第一章、緒論

1.1 研究背景

從 18 世紀第一次工業革命開始後，世界進入工業時代，人民的生活方式、品質因此快速上升，在煤炭之後，石化能源的運用更帶給人民舒適、方便的生活，沒有節制的使用致使環境汙染、溫室效應加劇，劇烈氣候變遷不僅威脅了海洋、大地以及空氣中的生命，也對人類生活中的許多領域引發負面影響(Celik, 2020)。

在 20 世紀末數次石油危機加上環保意識抬頭，永續發展的理念成為各國政府及科學家努力的目標，各國政府也簽訂了《京都議定書》和《巴黎協定》來減少碳排，所謂的碳排，主要就是燃燒像是煤炭、石油、天然氣等化石燃料所組成氣體，在工業盛行的國家對於化石燃料的使用更是家常便飯，根據氣候媒體（Carbon Brief, CB）在 2021 年的統計，美國是世界上碳排累積量最多的國家，中國、俄羅斯、巴西緊隨其後，但為了達到碳量減排及永續發展同時又不對於國家經濟、人民生活水平造成影響，以這兩大方向為目標，綠色能源、替代能源的開發也就應運而生了，像是太陽能、風能、水力發電都是耳熟能詳的，也是世界各國努力投入研究的一塊，根據世界風能協會（Global Wind Energy Council, GWEC）的數據統計，2021 年世界的風能發電容量新增了 93,600,000 瓩，與 2020 年相比成長了 12.4%，如圖 1 所示。而據能源知識庫統計，丹麥在 2017 年風力發電發電量在全國用電量占比已經達到 43%，並承諾在 2030 年會完全廢棄燃煤發電的使用，綜合上述，可以看出綠色能源、永續發展已經成為一個趨勢。

而在綠色能源中，風能因為被視為是對於永續發展中最有發展性的發電方式(Lin et al., 2020)，也是近年來成長幅度最快的一種發電方式(Zhao et al., 2016)，而有別於火力發電的用電量規劃是燃煤燒水燒出蒸氣帶動渦輪機，風力發電因為是依靠氣壓變化形成的風進行運轉渦輪機帶動發電，變化較為不固定，所以比起傳統火力發電，更需要做好發電量預測，才能進行預測性維護、保養安排及用電規劃。

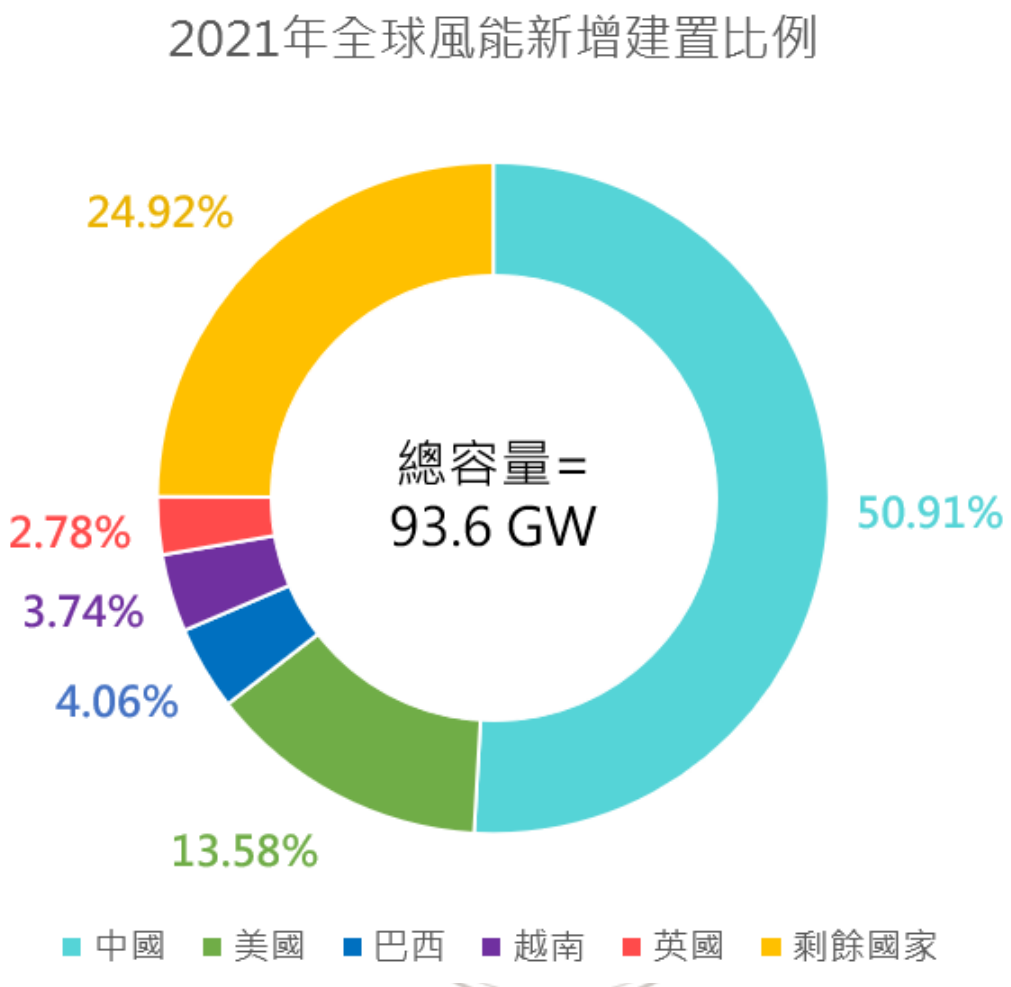

圖 1 2021 年全球風能建置容量比例(資料出處:世界風能協會, GWEC)

1.2 研究動機

隨著綠色能源的重要性和使用率日益增加，加上綠色能源的發電功率並不像傳統火力發電的功率較為穩定，對此的電量規劃及功率預測便越發重要，而適逢巨量資料處理的時代，機器學習和深度學習在各產業也被廣泛應用。近年來已有學者將人工智慧導入到此一領域，以機器學習和深度學習應用於功率預測及渦輪機組的異常檢測上，當中絕大部份研究都是與風力發電的檢測相關(Hong et al., 2019; Leng et al., 2018)，主要是因為風力發電為近代使用率最高以及最被推崇的一種綠色能源(Gazafroudi, 2015)。而在過往風電發電量的研究中，大多都是以天氣數值預報(NWP)中的風速、氣溫、氣壓、風向等環境因素作為輸入特徵，來預測目標值發

電量(Zhang et al.,2019; Aly, 2020; Xiong et al., 2021)。而在風電的異常檢測，都會透過監控系統 Supervisory control and data acquisition(SCADA) 所採集到的風機狀態數據並結合深度學習的方法進行預測(Xiang et al., 2021; Ferguson et al., 2019)。另外，林學者在 2020 便提出利用環境溫度、風速及 SCADA 數據裡面的風力發電機上的葉片槳角作為模型輸入，證實該做法可以降低風電預測的計算成本時間(Lin et al., 2020)。在風電的研究中，利用 SCADA 數據和風速作為特徵已是常態。但是上述研究中，因為風力渦輪機使用的場域特徵是難以做預測的，通常是利用數值天氣資料搭配大量的前處理方法，才有可能得到一個有效的輸入特徵(Wang et al., 2021)。由於風的隨機性高且變化較大，進而導致風力發電廠的安全性與穩定性受到了極大的挑戰，因此如何精準地預測風力發電量就非常的重要，而風力發電模型的預測結果，可以使電廠能更有效地去調度、管理與優化不同電網系統中的運作(Wang et al., 2021)。在上述文獻中也有提到若是能使用更多的監控資料作為輸入特徵便能使預測的準確度提高，因此本研究提出利用 SCADA 監控數據作為主輸入結合深度學習模型，期望能建構出一個運算成本低準確度又高的風力發電模型。

1.3 研究目的

基於上述動機，本研究將透過由 SCADA 監測系統所得到的風力渦輪機機組內部狀態的數據，來進行風力發電量的預測。並提供給發電廠的作業人員做預測參考，以利於發電廠實行排程規劃的安排，例如:停機維護、電力調度、負載跟蹤等作業流程。

本研究藉由 Kaggle 資料科學競賽平台取得的渦輪機狀態和發電量進行風電預測的建模。先將資料進行異常值檢測，避免感測器感應不良或環境問題導致數據資料不平衡，透過平穩性檢測，並使用差分法處理非平穩性的問題，接著選取重要特徵，並以 LSTM 模型預測目標值。

本研究目的如下:

1. 探討風力渦輪機機組狀態對於發電量的影響
2. 藉由不同特徵數的組合建立準確度較高的模型

1.4 研究流程

本研究以 Kaggle 公開資料平台提供的風力渦輪機狀態資料作為本研究數據，將本研究分為 5 個主要章節，其架構如圖 2 所示。第一章為研究背景、研究動機與研究目的，說明相關產業領域的近況與發展。第二章為文獻探討，探討近年來風力發電預測的方法，包含特徵選取、LSTM 模型以及相關 SCADA 數據應用的文獻。第三章為研究方法，首先會介紹數據資料與分析流程，接著進行資料前處理，以利於模型的建立與績效評估。第四章、第五章會呈現本研究的結果並對其進行說明與討論。第六章為結論與未來研究。

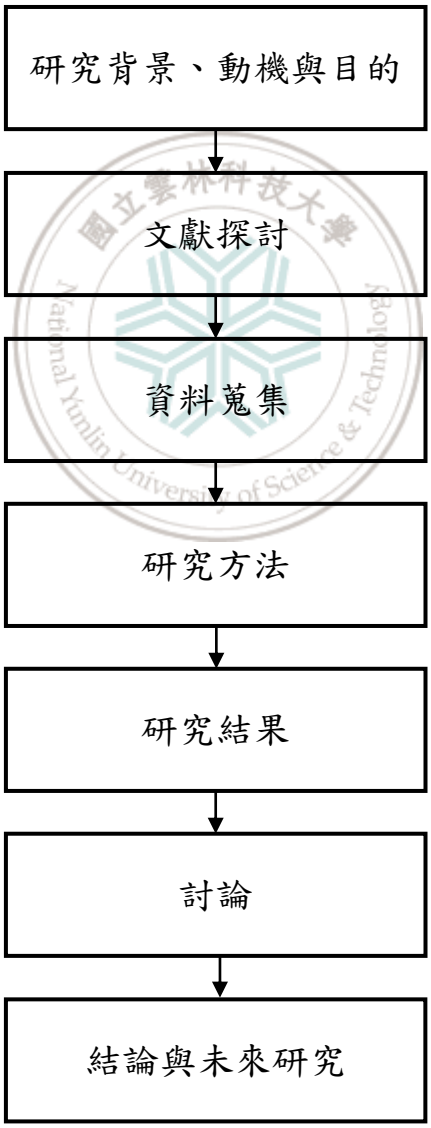

圖 2 研究流程架構圖

第二章、文獻探討

2.1 風力渦輪機簡介

風力渦輪機（Wind Turbine）又稱作風力發動機，是一種能將氣流中的動能轉化為機械能的裝置，是風力發電廠的組成必要元素之一。風力渦輪機會通過風力來帶動風車葉片使其旋轉，再通過增速機使旋轉的速度提升，並促使發電機發電。風力渦輪機主要由機頭、葉片、轉體和尾翼組成，而其中葉片被用來接收風力並通過機頭轉變為電能，尾翼使葉片永遠保持面向風來的方向藉此獲得最大的風能，轉體使機頭能夠靈活地轉動以實行調整尾翼方向的功用。風力渦輪機被廣泛地運用在多風的地理環境，和內燃機（Internal combustion engine）獲得的發電成本相較之下，風力渦輪機獲得的電力成本相對低得多。

(Evans et al., 2009)認為，相比於太陽能發電、地熱能發電、燃煤發電、燃氣發電和水力發電等發電的方式，風力發電被確定為溫室氣體排放量最低的發電方式、同時用水需求也是最低的。

2.2 風力發電功率預測方法簡介

(Jung et al., 2014)指出根據不同電力系統運行的要求，風力發電預測可分為四個不同時間段，分別為極短期 (幾秒至 30 分鐘)、短期 (30 分鐘至 6 小時)、中期 (6 至 24 小時)和長期 (1 至 7 天)，極短期主要用於渦輪機的控制和負載的追蹤，短期用於預負載的共享，中期應用於能源的交易和電力系統的管控，長期則用於風力渦輪機的維護調整。Hanifi, S et al. (2020)研究中提到根據不同的應用方法，風力發電預測模型還可分為物理法、時序模型和人工神經網路 (ANN)，這些方法的差異在於所需要的輸入數據、不同時間範圍下的準確性和運算過程的複雜性。

(Hanifi et al., 2020)提到風力發電的物理法預測的方式主要是藉由收集到的數值天氣預報 (NWP)來進行，而此法需要對特定區域進行描述，例如地形的粗糙度、附近的障礙物以及大氣中的溫度、氣壓等天氣預報的數據，這些收集到的變量會被用來建立複雜的風速模型，接著將預測的風速帶入由渦輪機製造商提供的風力渦輪機發電量曲線進行風力的預測，雖然此法不需要輸入歷史的數據來訓練，但非常

仰賴於物理數據並且計算非常的耗時。時序模型預測方法的開發主要基於數值天氣預報 (NWP)數據，透過收集到的天氣數據像是風速、風向和溫度來探討與發電量之間線性與非線性之間的關係，時序模型包含自回歸模型 (AR)、自回歸移動平均模型 (ARMA)、移動平均模型 (MA)和自回歸綜合移動平均模型 (ARIMA)，這些模型又屬於傳統型統計方法，主要使用歷史數據進行訓練。雖然這些傳統統計方法容易建模且開發成本又低，但其精準度會隨著預測時間的增加而降低(Chang, 2014)。人工神經網路 (ANN)為現今被最廣泛運用的風力發電量的預測方法之一，使用此方法的原因之一是能夠有效地避免機械結構在風力渦輪機中的複雜性，主要可以辨別輸入特徵變量和輸出功率數據之間存在的非線性關係，其模型架構主要由輸入層、單個或多個中間的隱藏層和一個輸出層所組成，並對其輸入歷史數據與特徵進行訓練以及測試，而能夠影響人工神經網路表現性能的因素有非常多，其中包括數據的預處理、數據的結構、網路的學習方法、數據在輸入和輸出中的連接等等。

2.3 相互資訊(Mutual Information, MI)

(Garan et al., 2022)透過資料採集與監視系統(SCADA)所收集到的風力渦輪機感測器數據進行預測性維護的分類與回歸預測任務，預測的目標組件包含齒輪箱、變壓器、發電機、發電機軸承和液壓組，研究中藉由 MI、特徵重要性(FI)、相關性濾波器(CF)、獨立成分分析(ICA)以及主成分分析(PCA)等方法進行比較，並將收集到的 121 個特徵數據集進行特徵選擇並使用決策樹(DT)模型進行預測，結果顯示 MI 對於發電機與液壓組的回歸預測任務中有良好的表現，此法將原數據的 121 個特徵數降為 40 個，能夠有效地篩選出最重要的特徵，以降低模型計算的時間成本與維修成本，並提高模型的準確度。

(Wang et al., 2020)使用 MI 與相關定量分析結合倒傳遞類神經網路(BPNN)建構混合模型，進行柴油發動機的氧氮化物(NO_X)排放預測，研究中藉由 MI 分析每個特徵輸入信號和目標輸出之間的相關性，將 67 個特徵降為 15 個特徵當作模型的輸入變量，結果顯示 MI 可以明顯地減少數據集中的維度和繁冗，進而提高模型的準確度並減少模型的計算時間。

(Liu et al., 2020)提出了一種基於 MI 和堆疊降噪自動編碼器(SDAE)結合 LSTM 模型進行風速預測，研究指出 MI 方法已經被普遍應用在氣象時序建模、語音辨識等領域，MI 能夠計算每個變量之間的統計相關性，並幫助模型篩選出重要的特徵因素藉此來提高模型的性能。

(Chelgani et al., 2018)使用 MI 對煤浮選的變量進行特徵選擇，並結合支持向量回歸(SVR)模型對煤浮選的速率和回收率進行預測，研究指出 MI 可以有效地量化變量之間的相互作用，並篩選出重要的特徵當作模型的輸入變量，結果顯示加入 MI 的 SVR 的判定係數(R^2)從 0.72 提高到 0.93，大幅地提高了模型的可解釋性，並且可降低模型過度擬合的可能性，另外研究也提到特徵選擇是能夠有效地節省在測量關係不大參數時，所產生的額外時間與成本。

2.4 長短期記憶神經網路(Long Short-Term Memory, LSTM)

(Zhang et al.,2019)應用數值天氣預報(NWP)數據中的風速數據進行風力發電量短期的預測，研究中比較了 LSTM、徑向基函數(RBF)、深度信念網路(DBN)、倒傳遞類神經網路(BPNN)、小波以及 Elman 神經網路(ELMAN)等六個模型的準確性與運算時間，而結果顯示 LSTM 雖然比其他模型的運算時間更長，但其預測準確度為佳，而研究中也提到 LSTM 對預測前兩天風力渦輪機的輸出功率時間序列穩定有效。

(Xiong et al., 2021)提出了以 LSTM 與門控循環單元(GRU)兩個不同模型為基礎結合深度神經網路(DNN)進行短期風力發電量的預測，研究使用 DNN 來衡量風速、風向以及溫度三個不同物理數據的重要性，並對預測模型的輸入特徵賦予相異的權重，此組合模型能更有效地提取物理特徵並減少模型的複雜度，結果顯示 DNN-LSTM 比 DNN-GRU 有更準確的預測能力與穩健性。

(Jaseena et al., 2019)在風速預測上比較了 LSTM 和自回歸綜合移動平均模型(ARIMA)，研究中進行了短期、中期和長期不同時間間隔的風速預測，結果顯示在 4 個模型衡量標準(MSE、RMSE、MAE、MAPE)中，LSTM 的預測能力都能夠完全優於 ARIMA，研究中也提到由於風速是屬於非平穩性的數據，所以 LSTM 能夠比 ARIMA 更好地處理非線性的數據。

(Hossain et al., 2020)應用 LSTM 對短期光伏發電的預測，預測的時段可分為冬季(12 至 2 月)、春季(3 至 5 月)、夏季(6 至 8 月)與秋季(9 至 11 月)，且以皮爾森積動差相關係數(PPMCC)將氣溫、風速、降水量、大氣壓力、相對濕度以及太陽照幅度等多個特徵進行相關性分析並篩選出重要的輸入特徵，最後與廣義迴歸神經網絡(GRNN)與非線性自回歸外生模型(NARX)兩種方法進行比較，其分析的結果顯示在不同季節的時段下 LSTM 的預測能力較為準確，並且模型整體的性能表現也優於其他兩個模型。

(Zhang et al., 2019)應用 LSTM 建構下一個小時的風力發電功率模型，為了降低模型的複雜度，使用了自動編碼器(AutoEncoder)減少數據的維度，輸入的特徵包含測量的時間點、大氣壓力、氣溫、風向以及不同高度的風速，與基礎 LSTM、支持向量機(SVM)進行比較，研究發現 Auto-LSTM 與 LSTM 比 SVM 更能接近真實的情況，結果顯示雖然 Auto-LSTM 和 LSTM 的預測能力非常接近，但由於前者有經過降維的處理，所以在訓練計算的時間上能夠比後者來的更加快速。

2.5 相關 SCADA 數據分析預測之文獻

(Xiang et al., 2021)結合 CNN、LSTM 與注意力機制(AM)開發出一個新穎的模型，對風力渦輪機的 SCADA 監控系統採集到的數據進行發電機故障預測，數據包含環境溫度、功率、軸承溫度、電流、發電機轉速等輸入特徵，研究中提到雖然 SCADA 數據包含許多的機組狀態數據，但過多的輸入變量進到預測模型中，會因為數據冗餘導致模型精準度大幅地降低，結果顯示該方法透過 SCADA 數據能夠有效地預測風力渦輪機的故障與異常。

(Su et al., 2019)使用 LSTM 進行超短期風力發電量的預測，研究中先利用小波包分解(WPD)與經驗模態分解(EEMD)將風速特徵進行前處理，並且增加偏航誤差和轉子速度兩種由 SCADA 監控系統採集的風力渦輪機狀態作為輸入特徵，其結果顯示 LSTM 對於超短期風力發電量這類時間序列的問題能夠獲得良好的表現，且新增了兩個風力渦輪機狀態的特徵能夠使預測結果更符合真實的風力發電廠情況，對於模型的預測準確度也能夠大幅地提高，另外研究也指出模型的準確度會隨著風力渦輪機狀態特徵的增加而有所提高，因為這會使實驗考慮更多實際的條件。

(Eseye et al., 2018)結合小波轉換(WT)、粒子群優化(PSO)以及 SVM 進行光伏太陽能發電預測，該研究使用 SCADA 發電量數據與數值天氣數據(NWP)作為輸入變量，結果顯示此分析方法對於短期光伏發電的預測是有效的。

2.6 小結

綜合上述文獻的回顧，可以發現 MI 能夠有效地選出與目標值高度相關的特徵，且對於操作者也更方便理解，而 LSTM 可以改善資料因長時間的參數累積與訊息流失，所產生的梯度爆炸或梯度消失，並且 LSTM 是能夠應用於時間序列數據的分析，而風力渦輪機的機組狀態也是由時間順序組成的資料。雖然有許多學者提出各式方法進行風力發電量的預測，但大部分仍是使用風速作為主要輸入變量，故本研究將使用 MI 對 SCADA 數據進行特徵選取，選出對於目標輸出發電量相關性最高的特徵當作輸入變量，接著使用 LSTM 進行發電量的預測，並探討不同特徵數對於模型績效的影響。

第三章、研究方法

3.1 資料分析流程

透過圖 3 可以了解本研究分析流程主要分為四個階段，分別為資料前處理、特徵擷取、模型建立與驗證與績效評估，而資料前處理包含資料觀察與清洗、平穩性檢測與轉換與數值縮放，接著為特徵選擇、模型建立與參數調整，最後則是模型驗證與績效評估。

首先在本研究取得數據資料後，會先使用盒鬚圖(Box Plot)觀察資料的分佈，在觀察資料的同時也能很清楚地識別異常值並將其剔除，接著會使用線性插值法將缺值進行填值，以使資料完整便於後續的分析。隨後會透過 Augmented Dickey-Fuller(ADF)與 Kwiatkowski Phillips-Schmidt-Shin(KPSS)這兩種方法進行平穩性的檢驗，來判斷資料有無單位根問題，當資料存在單位根的問題時，代表數據目前屬於非平穩性的數據，換言之為常見的趨勢性問題，到時可藉由簡單差分法或是季節差分法來調整數據中的數值。接著利用最小值與最大值正規化將資料等比例縮放轉換到特定數值區間，並使用相互資訊(Mutual Information, MI)方法選出與目標值相關性最高之特徵。

第二步將建立能夠預測時間序列的 LSTM 模型，並搭配貝葉斯優化(Bayesian Optimization)進行參數調整，使模型參數能達到最佳化狀態。

最後會將訓練好的模型進行驗證，並透過 R^2(R-squared)、均方根誤差(RMSE)、平均絕對誤差(MAE)與均方誤差(MSE)評估模型的績效，以此判斷模型是否合適。

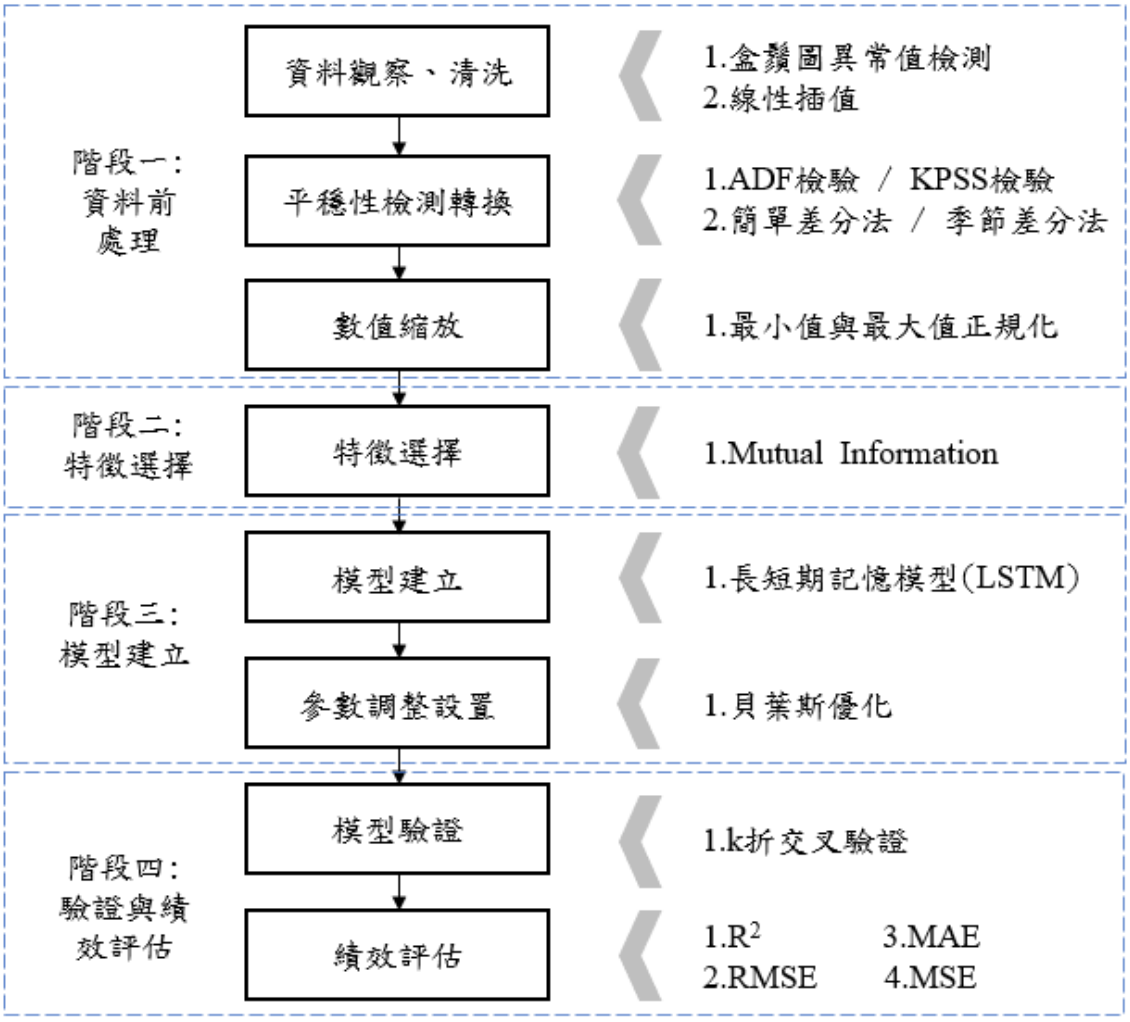
階段一：
資料前
處理
資料觀察、清洗
1.盒鬚圖異常值檢測
2.線性插值
平穩性檢測轉換
1.ADF檢驗 / KPSS檢驗
2.簡單差分法 / 季節差分法
數值縮放
1.最小值與最大值正規化
階段二：
特徵選擇
特徵選擇
1.Mutual Information
階段三：
模型建立
模型建立
1.長短期記憶模型(LSTM)
參數調整設置
1.貝葉斯優化
階段四：
驗證與績
效評估
模型驗證
1.k折交叉驗證
績效評估
1.R^2
2.RMSE
3.MAE
4.MSE

圖 3 資料分析流程圖

3.2 資料集介紹

本研究使用的實驗資料為 Kaggle 資料科學平台提供的風力渦輪機狀態數據，數據集主要由安裝在風力渦輪機上的 SCADA 資料採集與監視系統所蒐集，時間範圍為 2019 年至 2021 年且時間間隔為每 10 分鐘感測一次，包含 76 個特徵值(機組狀態)、1 個目標值(發電功率)，共有 136,731 筆實際資料，其相關統計資料如表 1 所示。

表 1 各特徵與目標值的統計資料

變數名稱	單位	最小值	最大值	平均值
特徵值				
變速箱 T1 高速軸溫度	°C	-273	99999	912.59
變速箱 T3 高速軸溫度	°C	-273	99999	891.24
變速箱 T1 中間高速軸溫度	°C	-273	99999	980.84
變速箱空心軸承溫度	°C	16	99999	946.43
塔台正常加速度	mm/s	0	99999	931.59
變速箱 2 油溫	°C	-273	99999	924.2
塔台側向加速度	mm/s	0	99999	935.31
軸承-A 溫度	°C	-273	99999	865.81
變壓器-3 溫度	°C	13.85	99999	916.45
變速箱 T3 中間高速軸溫度	°C	-273	99999	907.96
變速箱 1 油溫	°C	-273	99999	936.49
變速箱油溫	°C	-260.71	99999	852.88
力矩	%	-327.68	99999	966.49
無功功率控制轉換器	kVAr	-1303.37	99999	884.99
變壓器 2 溫度	°C	15	99999	922.53
無功功率	kVAr	-1307.99	99999	882.43
軸承 1 溫度	°C	-273	99999	872.71
變速箱分配器溫度	°C	-273	99999	905.47
力矩 D 過濾	kNm	-3001.27	99999	1021.96

力矩 D 方向	kNm	-3001.36	99999	979.7
機組 N 轉速	rpm	0	99999	2234.78
運作狀態	無	6	99999	837.24
功率因數	無	-1	99999	921.94
軸承 2 溫度	°C	-273	99999	912.11
機艙溫度	°C	-273	99999	866.78
電壓 A-N	V	0	99999	1279.56
軸箱 3 溫度	°C	-83.67	99999	908.95
電壓 C-N	V	0	99999	1223.29
軸箱 2 溫度	°C	-103.89	99999	927.69
軸箱 3 溫度	°C	-42.99	99999	932.41
電壓 B-N	V	0	99999	1267.58
機艙位置度數	度	0	99999	1121.69
電壓控制轉換器	V	0	99999	1515.8
電池箱 3 溫度	°C	-110.63	99999	894.48
電池箱 2 溫度	°C	-273	99999	912.85
電池箱 1 溫度	°C	-150.36	99999	951.65
液壓壓力	bar	-25	99999	976.75
角轉子位置	度	0.02	99999	1071.34
塔基底溫度	°C	4	99999	854.82
俯仰偏移非對稱負載控制器 2	度	-0.05	99999	866.35
俯仰偏移塔反饋	度	-0.02	99999	874.45
電源頻率	Hz	0	99999	975.13
內部功率限制	kW	1595.17	99999	3618.59
斷路器切入	無	395	99999	1452.23
粒子計數器	無	0	99999	902.67
塔台原始正常加速度	mm/s	-15832.72	99999	838.86

扭矩偏移塔反饋	Nm	-9.27	99999	828.78
外部功率限制	kW	0	99999	4108.58
葉片 2 實際角度 B	度	-1081.24	99999	866.85
葉片 1 實際角度 B	度	-1020.98	99999	913.8
葉片 3 實際角度 B	度	-2116.34	99999	844.43
熱交換控制轉換器溫度	°C	1	99999	839.48
塔台原始側向加速度	mm/s	-8.81	99999	829.75
環境溫度	°C	-273	99999	890.14
機艙迴轉	無	-1.81	99999	804.03
俯仰偏移非對稱負載控制器 1	度	-0.92	99999	809.43
塔偏轉	ms	2952	99999	3999.82
俯仰偏移非對稱負載控制器 3	度	-0.54	99999	894.27
每秒風向偏差	度	-180	99999	891.17
每 10 秒風向偏差	度	-180	99999	837.62
鄰近感測器 135 度	mm	-0.24	99999	837.79
狀態和故障	無	1	99999	919.29
鄰近感測器 225 度	mm	-0.24	99999	852.38
葉片 3 實際角度 A	度	-408.78	99999	831.67
CH4 範圍	無	-408.78	99999	837.29
葉片 2 實際角度 A	度	-408.78	99999	839.5
葉片 1 實際角度 A	度	-408.78	99999	956.68
葉片 2 設置度數	度	-0.13	99999	882.6
俯仰需求基線度	度	0	99999	934.69
葉片 1 設置度數	度	-0.77	99999	873.61
葉片 3 設置度數	度	-1.18	99999	845.04
力矩 Q 方向	kNm	-1287.39	99999	884.32
力矩 Q 過濾	kNm	-1286.71	99999	824.71

鄰近感測器 45 度	mm	-0.24	99999	817.36
渦輪機狀態	無	1	99999	885.7
鄰近感測器 315 度	無	-0.24	99999	869.65
目標值				
發電功率	kw	-48.597	2779.423	1138.556

3.3 資料前處理

在資料取得之後必須對資料進行前處理，消除資料不平衡的狀態，以便於後續的資料分析，也確保模型的預測績效可以保持良好。資料前處理主要分為四個部分，包含異常值的檢測與缺失值的插補、平穩性檢驗以及資料正規化，以下將針對各步驟進行說明。

3.3.1 盒鬚圖識別異常值與線性插值法

在資料分析中，資料時常會存在極度偏離平均值的離群值，而若將存有極小值或極大值的資料當作輸入時，則會影響後續模型訓練時的績效，進而導致預測準確度不佳。而本研究採用的異常值檢測方法為盒鬚圖，此法是將一組數據顯示資料分佈的統計圖，主要由最大值、最小值、中位數、第一四分位數($Q1$)、第三四分位數($Q3$)組成，在計算第一四分位數前會先將n個資料依大小作排序，接著會透過公式(1)進行計算，若i為整數，則第一四分位數的值為第i個與第$i+1$個觀察值之平均；若i為非整數，則第一四分位數的值為大於i的下一個整數的觀察值，反之，第三四分位數計算方式則為公式(2)。i為第一四分位數或第三四分位數中的第幾筆觀察值，n為資料的總筆數。

$$i = (25/100) \times n \tag{1}$$

$$i = (75/100) \times n \tag{2}$$

而其中第三四分位數與第一四分位數的差值稱為四分位距(Interquartile range, *IQR*)，1.5 倍的四分位距也是組成盒鬚圖中籬笆(fence)的部分，是判斷資料是否為異常值的重要依據，當資料中的值超出$[Q1 - 1.5 \times IQR\,, Q3 + 1.5 \times IQR]$這個範圍時，則斷定該值為離群值(outlier)，實際的盒鬚圖表示方法如圖 4 所示。

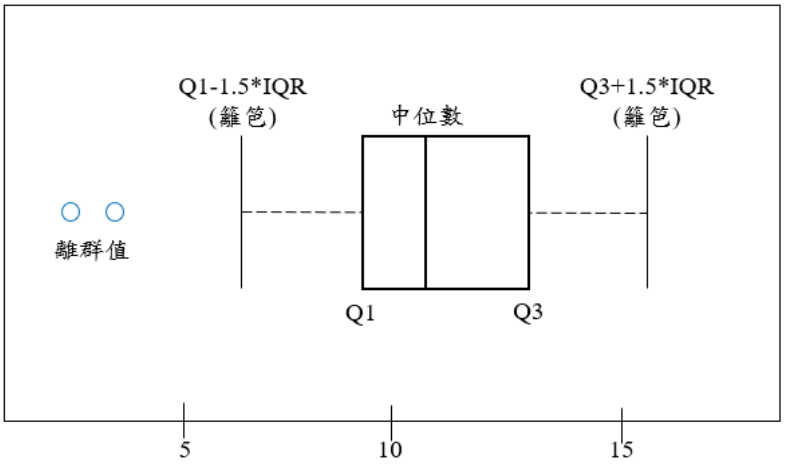

圖 4 盒鬚圖示意圖

當資料做完異常值檢測之後，必須將識別出的異常值進行處理，由於本研究使用的資料為風力渦輪機機組狀態的數據，屬於時間序列的資料，而時間序列的資料存在先後順序的特性，故不適將異常值直接剔除，若直接剔除資料會造成資料產生不連續的問題，本研究使用線性插值法對異常值進行處理，其計算方式為計算任兩個鄰近的表列點之間的數值，依據兩點與所求點的比例計算而得，其計算方式如公式(3)、(4)所示。本研究之各機組狀態資料數值為$Y = [y_1 \text{、} y_t \dots y_T]$、時間日期序列為$X = [x_1 \text{、} x_t \dots x_T]$、$y_t$為所求的機組狀態數值、$y_{t-i}$及$y_{t+k}$為鄰近時間點的機組狀態數值、$t$為第幾項資料、$t-i$與$t+k$為鄰近的時間點。

$$\frac{y_t - y_{t-i}}{x_t - x_{t-i}} = \frac{y_{t+k} - y_{t-i}}{x_{t+k} - x_{t-i}} \tag{3}$$

經整理後得:

$$y_t = \frac{y_{t+k} - y_{t-i}}{x_{t+k} - x_{t-i}}(x_t - x_{t-i}) + y_{t-i} \tag{4}$$

3.3.2 平穩性檢驗

在許多時間序列資料分析的研究中，一般會假設時序數據皆保有平穩性的性質，但是這個法則不一定會有效，尤其是在真實數據的分析中(Salles et al., 2019)。平穩性是指在時序數據中，資料的平均數、共變異數以及變異數並不會隨著時間的推移而有所改變，因此本研究會利用平穩性檢驗對蒐集到的數據進行辨別，若結果為非平穩的狀態時，將透過差分法轉換為平穩的狀態。以下將詳細說明平穩性的種類、檢驗方法與轉換方法。

1. 平穩性的檢驗方法

（1）ADF 檢驗(Augmented Dickey-Fuller test)

主要檢查時間序列數據中有無單位根(unit root)的存在，故又稱為單位根檢驗，當時序數據存在單位根時，則為非平穩性的數據，反之當不存在單位根時，則為平穩性數據。在 ADF 檢驗中，首先會設立虛無假設(H0)與對立假設(H1)，H0 為存在單位根，反之 H1 則為不存在單位根，其判斷方式為檢驗統計量 t 值是否小於在不同信賴區間下的臨界值，可分為 1%、5%與 10%的信賴區間，若檢驗統計量 t 值小於臨界值，則可拒絕 H0，也代表時序為平穩的，而若檢驗統計量 t 值大於臨界值，則不可拒絕 H0，並表示時序是非平穩的。

（2）KPSS 檢驗(Kwiatkowski Phillips-Schmidt-Shin test)

在 KPSS 檢驗中則和 ADF 檢驗相反，其 H0 假設為時序是平穩的，而 H1 假設為時序是非平穩的，也就是時序存在單位根。當檢驗統計量 t 值小於不同信賴區間下的臨界值時，則不可拒絕 H0 假設，也就是時序是平穩的，反之，當檢驗統計量 t 值大於臨界值時，則拒絕 H0 假設，表示時序是非平穩的。

2. 平穩性的種類

當時序數據在經過 KPSS 檢驗與 ADF 檢驗後，需要將判別後的結果進行分類，以利後續對其進行差分法的轉換，而平穩性的種類主要可分為三個不同類別，詳細說明如表 2 所示。

表 2 平穩性種類的說明

平穩性種類	ADF 檢驗結果	KPSS 檢驗結果	轉換方法
嚴格平穩	平穩	平穩	無
趨勢平穩	非平穩	平穩	季節差分法
差分平穩	平穩	非平穩	簡單差分法

3. 非平穩性的轉換方法

在經過上述表格的呈現，可以清楚看到當 ADF 檢驗為非平穩，而 KPSS 檢驗為平穩時，該時序數據被視為趨勢平穩。反之，當 ADF 檢驗為平穩，KPSS 檢驗為非平穩時，則稱作差分平穩。接著需要對非平穩的時序數據進行轉換，若未對非平穩性的數據進行轉換，則資料本身可能會因為時間的增加而產生季節性或趨勢性，進而影響後續資料的分析。非平穩性的轉換方法又可分為季節差分法與簡單差分法，以下將對兩方法進行說明。

（1） 簡單差分法

藉由計算當前時間點的數值與上一個時間點的數值之間連續項的差值，來消除該時序中的平均數變化。其計算方式如公式(5)所示，y_t為時間點t的觀察值、$t-1$為上一個時間點的觀察值、y_t'為兩觀察值的差值。

$$y_t' = y_t - y_{t-1} \tag{5}$$

（2） 季節差分法

此方法不計算連續數值間的差異，而是計算當前觀察值與上一個相同季節的觀察值之間的差值，以此消除數據中的趨勢性，例如星期三的觀察值會與上一個星期三的觀察值進行相減。其計算方式如公式(6)所示，y_t為時間點t的觀察值、$t-n$為上一季時間點的觀察值、y_t'為兩觀察值的差值。

$$y_t' = y_t - y_{t-n} \tag{6}$$

3.3.3 數值縮放

進行資料分析時，通常會因為數據中單位的不同或數字大小代表性的不同，而影響統計分析的過程，最後導致模型的準確度降低。因此，本研究利用最小值與大最大值正規化(Min-Max Normalization)將原始資料進行等比例縮放至 0 到 1 區間，藉此來解決資料內單位不同的問題，並提高模型的準確度與降低模型訓練的時間。其計算方式如公式(7)所示。$X_{(t)max}$與$X_{(t)min}$分別為資料中的最大值與最小值，t為時間點，$X_{(t)nom}$為正規化後的值。

$$X_{(t)nom} = \frac{X_{(t)} - X_{(t)min}}{X_{(t)max} - X_{(t)min}} \in [0,1] \tag{7}$$

3.4 特徵選擇

在機器學習中，特徵選擇(Feature Selection)又稱作變量選擇或屬性選擇，主要過程為將訓練數據中冗餘或無關的特徵移除，並選出與目標值高度相關的特徵變量，其目的是為了簡化模型，使研究人員能更好地瞭解模型結構，並且也能降低模型的訓練時間以及發生過擬合的可能性。

3.4.1 相互資訊(Mutual Information, MI)

本研究使用的特徵選擇方法為相互資訊(Mutual Information, MI)，其主要計算方式可參考公式(8)，藉由計算兩個隨機變數間的相互依賴性，來量化變數中所含的資訊量。$H(X)$為變數所包含的平均信息量，$H(X|Y)$為在變數Y下，變數X剩餘的信息量。透過圖 5 可清楚瞭解其之間的關係。

$$I(X;Y) = H(X) - H(X|Y) \tag{8}$$

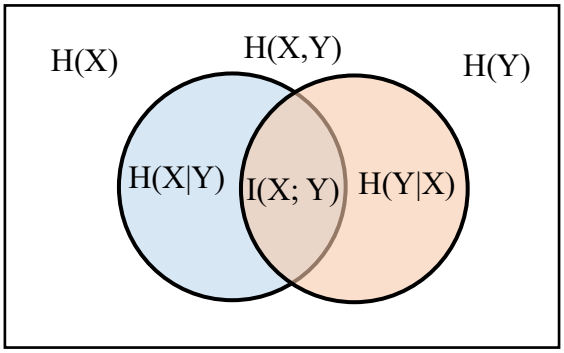

圖 5 變數之間的關係圖

3.5 模型建立

在經過上述異常值的識別與處理、平穩性檢驗、數值縮放等前處理步驟與特徵選擇後，將建立能夠預測時間序列資料的模型。而本研究所使用的資料為風力渦輪機機組狀態的數據，其具有時間先後順序的特性，也就是數據會隨著時間的推移而產生變化。因此，本研究將利用 LSTM 模型進行風力渦輪機發電量的預測，以下將詳細介紹模型與參數設置。

3.5.1 長短期記憶(Long Short-Term Memory, LSTM)

長短期記憶模型(LSTM)是一種能夠預測時間序列資料的神經網路，能夠改善循環神經網路(Recurrent Neural Network, RNN)中，因長時間累積而產生的梯度爆炸或梯度消失的問題，其結構主要由三個閥門所組成，分別為遺忘閥門(Forget Gate)、輸入閥門(Input Gate)與輸出閥門(Output Gate)，如圖 6 所示。

首先遺忘閥門會對當前輸入資料(X_t)與前一個時間點的隱藏狀態(h_{t-1})做選擇性刪除的動作，刪除對模型不重要的資訊，並保留較為重要的資料。而在輸入閥門(i_t)會決定輸入資料是否能夠加入到記憶單元，其目的為辨別資料的重要性，並透過激活函數(tanh)的計算，獲得臨時的單元狀態(c)，隨後將前一個時間點的單元狀態(C_{t-1})更新為新的單元狀態(C_t)。最後，輸出閥門(o_t)會決定下個隱藏狀態的值(h_t)，在過程中單元狀態(C_t)會與激活函數(tanh)做運算，並與激活函數(Sigmoid)的輸出進行相乘。

在 LSTM 中三個閥門的權重，都會透過輸入的資料自行學習與控制，而模型中的記憶單元可以持續地攜帶訊息並保留較早的訊息，藉此將重要的訊息傳遞下去。各閥門的權重係數為 W，各閥門的常數為 b。各函數的計算方式如公式(9)、(10)、(11)、(12)、(13)、(14)所示。

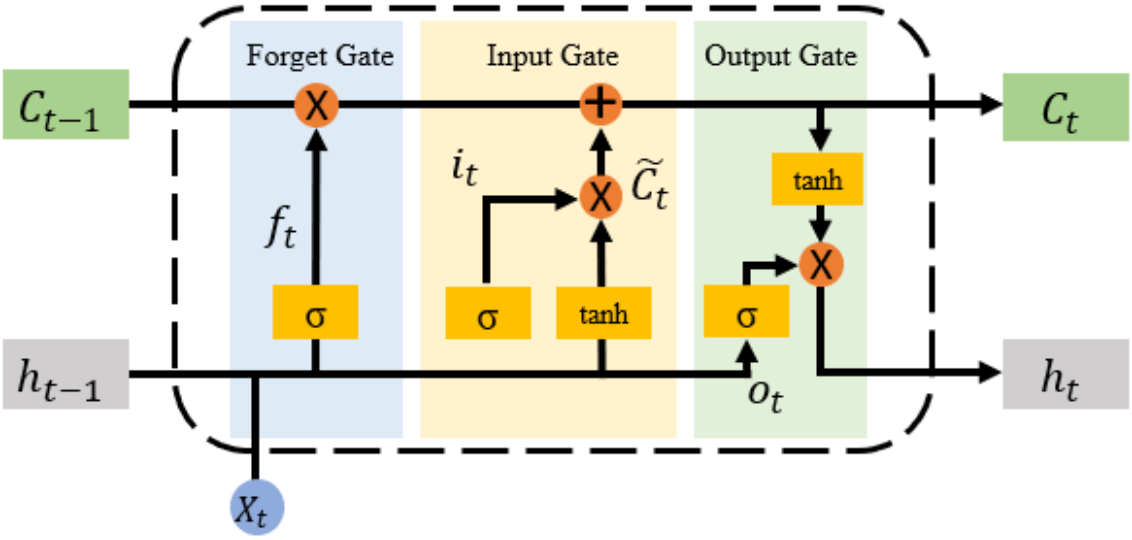

圖 6 LSTM 內部結構示意圖

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f) \quad (9)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i) \quad (10)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o) \quad (11)$$

$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, X_t] + b_C) \quad (12)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (13)$$

$$h_t = \sigma \times \tanh(C_t) \quad (14)$$

3.5.2 參數調整

在模型建立後，需要對模型內的參數進行設置，在機器學習中許多使用者都會藉由學習模型預設的參數，直接進行模型的訓練，雖然參數調整並非必要的步驟，但模型中預設的參數，往往並不是最佳的參數配置。然而，好的參數配置是能夠大幅提升模型整體的效能，並提高模型的績效指標。

目前常見的參數調整方法主要有兩種，分別為網格搜尋(Grid Search)與隨機搜尋(Random Search)，主要目的都是決定訓練模型中超參數的組合。網格搜尋是透過給定的參數範圍，並將範圍內的所有參數組合進行訓練，也就是當模型的參數組合共有 n 個時，模型進行訓練的次數也是 n 次，並在訓練結束時回傳最佳的參數組合。而隨機搜尋是為每個參數選擇一定數量的隨機值組合，舉例來說，當 A 參數有 2 種選擇、B 有 4 種、C 有 3 種，接著會在每次訓練時隨機抽取參數中的值，進行搭配組合然後訓練。雖然以上兩種方法會計算給定範圍內所有的值，但會花費較多時間與成本在運算的時間上。

而本研究將使用貝葉斯優化(Bayesian Optimization)進行參數的設置，假設有一個超參數組合為$X = \{x_1, x_2, x_{3,\ldots\ldots} x_n\}$($x_n$為某個超參數的值)，其目的為假設會有一個函數關係，存在於超參數與最後優化的損失函數中，且相異的超參數會獲得不同的效益，如公式(15)所示。在模型的參數調整上，貝葉斯優化是能夠有效提升模型整體的效能(Boelrijk et al., 2021)。

$$x^* = arg\min_{x \epsilon X} f(x) \tag{15}$$

其運算步驟如下所示:

步驟 1.定義輸入空間X的邊界，也就是要優化的參數上限及下限

步驟 2.選擇初始的方法參數，可透過均勻或隨機分布在輸入空間中，並在空間中的點執行實驗

步驟 3.使用可提供預測平均值與變異數的概率模型，來擬合目標中的函數$f(x)$，例如高斯過程、隨機森林等模型

步驟 4.依據擬合的模型，來最大化蒐集函數，並在下一次的運算空間中為參數提供

參考

步驟 5.選定空間中的一個點開始實驗

步驟 6.若滿足實驗的條件，則停止計算，否則重返步驟 3

3.6 驗證與績效評估

當模型透過訓練資料集訓練完成時，需要對其做驗證的動作，接著再使用測試資料集進行模型誤差值的分析，也就是利用評估指標來衡量模型的預測準確度，本研究使用四種不同的評估指標，分別為 R^2(R-squared)、RMSE(Root Mean Square Error)、MAE(Mean Absolute Error)與 MSE(Mean Square Error)，以下將說明本研究的模型驗證方法與績效評估指標的計算公式。

3.6.1 模型驗證流程

本研究中的風力渦輪機機組狀態資料，在經過資料前處理後，將其分成訓練資料與測試資料，再從訓練資料中切割出部分資料當作驗證資料，如圖 7 所示。訓練資料主要用來訓練模型中的各參數。驗證資料會透過訓練完的模型來檢驗模型的狀況，也就是進行模型的參數調整，以此提高模型的準確度。測試資料將被用做評估模型最終的表現，也就是用不同的評估指標來探討模型的預測能力。

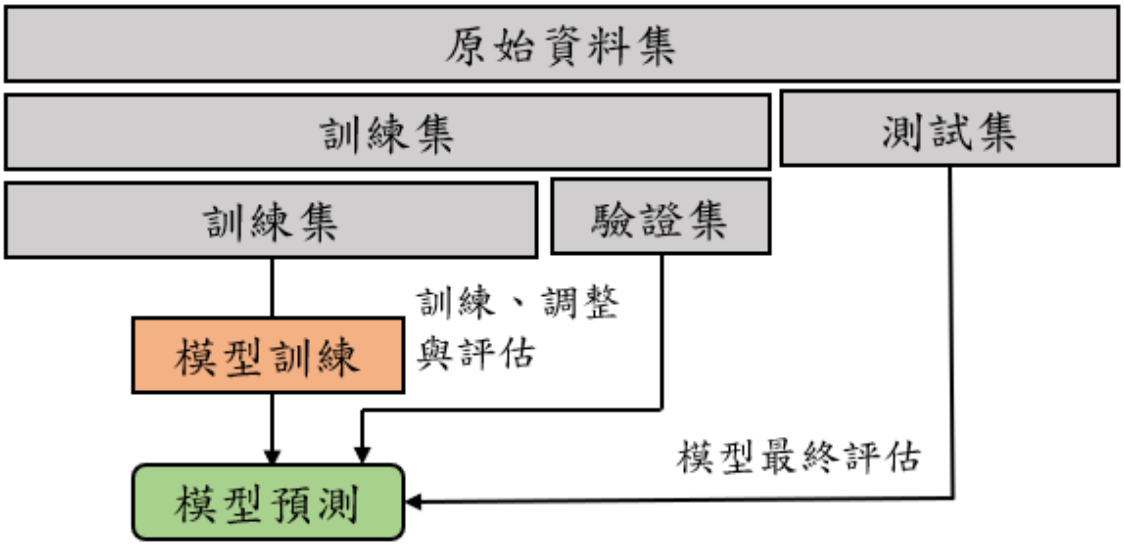

圖 7 模型驗證表示圖

3.6.2 驗證機制

由於機器學習模型可能存在的偽隨機性，如果在訓練模型時，只透過特定部分的訓練資料與測試資料，可能會產生偏差與方差導致模型的績效受影響。因此本研究使用 k 折交叉驗證(K-fold Cross-Validation)將訓練集切割成 k 個等份，並將 k-1 個的份作為訓練資料，剩餘的一份當作驗證資料，在每一次的計算會產生一個正確率，最後經過 k 次的反覆計算會產 k 個正確率，將其加總取平均值為最後的正確率。其計算流程如圖 8 所示。

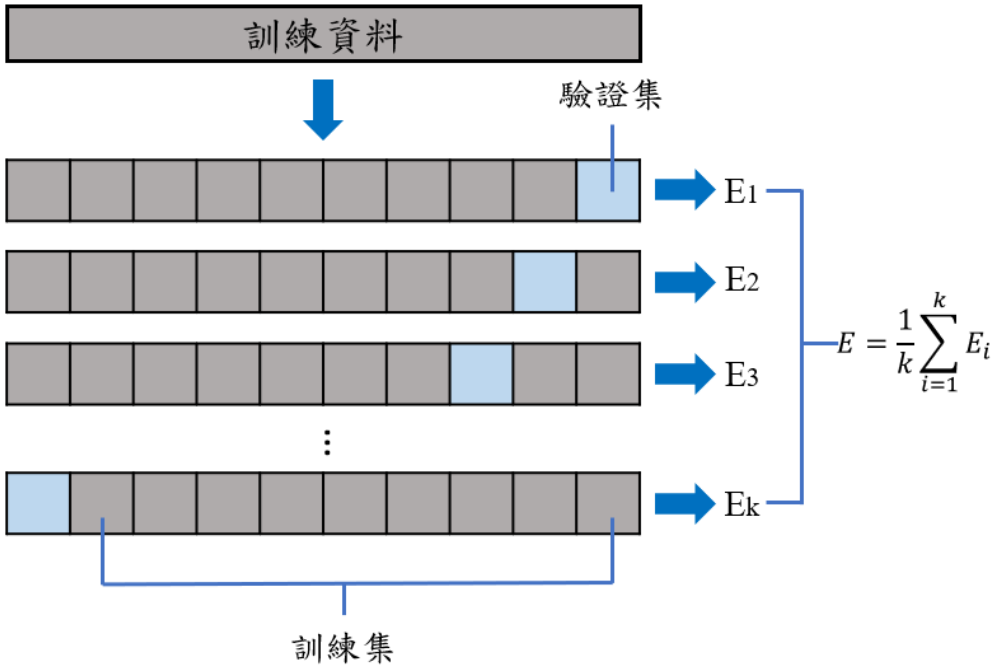

圖 8 k 折交叉驗證示意圖

3.6.3 評估指標

當模型驗證結束之後，需要利用測試資料進行模型的最終評估，由於本研究使用的資料為風力渦輪機機組狀態資料，其資料型態為數值型態，模型預測的結果為一個數值，並非分類標籤的類別型態，故本研究選擇 R^2、RMSE、MAE 與 RMSE 作為模型的評估指標方法，以下將對四種方法的計算方式進行詳細的說明。

1. R 平方(R-squared, R^2)

在統計學中 R 平方為常見的一種回歸模型評估指標，其所計算出的值通常介於 0 至 1 區間，主要用來衡量模型的可解釋性，也就是模型所解釋的依變量的變異性在總變異性中所佔的比例。當 R 平方的值越大，表示模型的可解釋越好，反之則越不好。R 平方的值並不代表模型的好壞能力，需要綜合其他回歸類型的評估指標結果。其計算方式如公式(16)所示，F_i為第i期的預測值，A_i為第i期的實際值，$\bar{A}$為實際值加總的平均，n為預測的總次數。

$$R^2 = 1 - \frac{1}{n}\frac{\sum_{i=1}^{n}(A_i - F_i)^2}{\sum_{i=1}^{n}(A_i - \bar{A})^2} \tag{16}$$

2. 均方根誤差(Root Mean Square Error, RMSE)

在回歸模型中，均方根誤差是一種很常用來衡量模型表現的評估指標，其主要探討的是預測值與觀察值之間的差異，並加總平方取根號。如公式(17)所示，是將均方誤差(Mean Squared Error, MSE)開根號後求得的，其目的為使數據值的單位能夠一致，對於模型結果的解釋也能夠更為直觀。當 RMSE 越小代表模型的預測能力越強，反之，當 RMSE 越大表示模型的預測能力越弱。F_i為第i期的預測值，A_i為第i期的實際值，n為預測的總次數。

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(F_i - A_i)^2}{n}} \tag{17}$$

3. 平均絕對誤差(Mean Absolute Error, MAE)

在回歸模型中，平均絕對誤差也是一種很常見的評估指標方式，主要是將每次觀察到的誤差取絕對值後平均，而 MAE 相較於 RMSE 下，對於數據中的極值較不敏感。其計算方式如公式(18)所示，F_i為第i期的預測值，A_i為第i期的實際值，n為預測的總次數。

$$MAE = \sum_{i=1}^{n} \frac{1}{n} |F_i - A_i| \tag{18}$$

4. 均方誤差(Mean Square Error, MSE)

均方誤差與均方根誤差一樣，是一種很常用來衡量回歸模型表現的評估指標，其主要是計算預測值與觀察值距離之間的平方和，MSE 可以評估資料的變化程度，MSE 的值越小代表模型的預測能力越佳，反之則代表預測能力較差。計算方式如公式(19)所示，F_i為第i期的預測值，A_i為第i期的實際值，n為預測的總次數。

$$MSE = \frac{\sum_{i=1}^{n} (F_i - A_i)^2}{n} \tag{19}$$

第四章、研究結果

在本研究中使用前 30 分鐘的風力渦輪機機組狀態數據來預測下一個 10 分鐘風力渦輪機發電量的目標值，以達到模型預測結果具有較佳的時效性，在本章節將詳細說明資料集、資料前處理結果、模型參數設定與目標值預測的績效。

4.1 資料集說明

本研究的資料集總共包含 76 個特徵值與 1 個目標值，特徵值主要為風力渦輪機機組狀態的數據，目標值則為風力渦輪機的發電量。而由於各特徵的名稱欄位長短不一，因此為了方便觀察與分析將各特徵設定編號，各特徵目標的編號、名稱與單位如表 3、表 4 所示。

表 3 特徵值欄位說明

編號	變數名稱	單位
0	變速箱 T1 高速軸溫度	°C
1	變速箱 T3 高速軸溫度	°C
2	變速箱 T1 中間高速軸溫度	°C
3	變速箱空心軸承溫度	°C
4	塔台正常加速度	mm/s
5	變速箱 2 油溫	°C
6	塔台側向加速度	mm/s
7	軸承-A 溫度	°C
8	變壓器-3 溫度	°C
9	變速箱 T3 中間高速軸溫度	°C
10	變速箱 1 油溫	°C
11	變速箱油溫	°C
12	力矩	%
13	無功功率控制轉換器	kVAr
14	變壓器 2 溫度	°C
15	無功功率	kVAr
16	軸承 1 溫度	°C
17	變速箱分配器溫度	°C

18	力矩 D 過濾	kNm
19	力矩 D 方向	kNm
20	機組 N 轉速	rpm
21	運作狀態	無
22	功率因數	無
23	軸承 2 溫度	°C
24	機艙溫度	°C
25	電壓 A-N	V
26	軸箱 3 溫度	°C
27	電壓 C-N	V
28	軸箱 2 溫度	°C
29	軸箱 3 溫度	°C
30	電壓 B-N	V
31	機艙位置度數	度
32	電壓控制轉換器	V
33	電池箱 3 溫度	°C
34	電池箱 2 溫度	°C
35	電池箱 1 溫度	°C
36	液壓壓力	bar
37	角轉子位置	度
38	塔基底溫度	°C
39	俯仰偏移非對稱負載控制器 2	度
40	俯仰偏移塔反饋	度
41	電源頻率	Hz
42	內部功率限制	kW
43	斷路器切入	無
44	粒子計數器	無
45	塔台原始正常加速度	mm/s
46	扭矩偏移塔反饋	Nm
47	外部功率限制	kW
48	葉片 2 實際角度 B	度
49	葉片 1 實際角度 B	度

50	葉片 3 實際角度 B	度
51	熱交換控制轉換器溫度	°C
52	塔台原始側向加速度	mm/s
53	環境溫度	°C
54	機艙迴轉	無
55	俯仰偏移非對稱負載控制器 1	度
56	塔偏轉	ms
57	俯仰偏移非對稱負載控制器 3	度
58	每秒風向偏差	度
59	每 10 秒風向偏差	度
60	鄰近感測器 135 度	mm
61	狀態和故障	無
62	鄰近感測器 225 度	mm
63	葉片 3 實際角度 A	度
64	CH4 範圍	無
65	葉片 2 實際角度 A	度
66	葉片 1 實際角度 A	度
67	葉片 2 設置度數	度
68	俯仰需求基線度	度
69	葉片 1 設置度數	度
70	葉片 3 設置度數	度
71	力矩 Q 方向	kNm
72	力矩 Q 過濾	kNm
73	鄰近感測器 45 度	mm
74	渦輪機狀態	無
75	鄰近感測器 315 度	mm

表 4 目標值欄位說明

編號	變數名稱	單位
76	發電量	(kW)

4.2 資料前處理

本研究在取得資料數據後，會先觀察資料的分布以及型態，並透過各種前處理方法清洗資料，前處理共包含三個部份，依序為盒鬚圖異常值的判斷、平穩性檢驗與數值縮放。首先將風力渦輪機機組狀態數據進行異常值的檢測，藉由盒鬚圖標示出數據中的離群值，以消除資料不平衡的狀態，接著使用線性插值法填補空值，藉此完整數據的連續性，完成後需使用平穩性檢驗來判別資料有無單位根的存在，最後使用最小值與最大值正規化，將數據中的值調整為[0,1]區間，以降低模型訓練的時間與單位不相同的問題。

4.2.1 盒鬚圖識別異常值與線性插值法

由於風力渦輪機機組狀態數據是由感測器蒐集的，因此可能會受環境因素或感應不良的影響，導致數據產生極值的問題。本研究在觀察資料後，發現各特徵欄位中的值都包含 99,999 數值，推測為感測器感應不良導致的問題，所以首先將 99,999 的值全部移除並填補為空值後，再進行第一次盒鬚圖異常值檢測，在此以變速箱 T1 高速軸溫度欄位為例，可以發現在籬笆外包含數個離群值，如圖 9 所示。

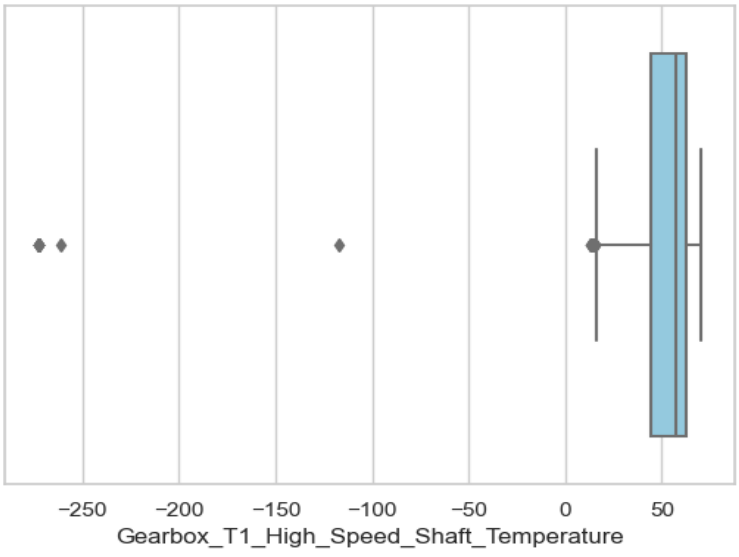

圖 9 盒鬚圖異常值檢測圖

透過上述盒鬚圖的呈現，可以發現將 99,999 數值移除後，仍存在離群值，因此需先將剩餘離群值進行排除，並使用線性插值法進行填補。如圖 10 所示，經過線性插值後可以發現數據中的資料皆符合正常區間內，本小節以變速箱 T1 高速軸

溫度欄位為例，其結果顯示只要經過一次異常值檢測與線性插值，即可達到本研究理想的效果，但若是其餘欄位無法在第一次達到預期效果，則會進行第二次異常值的檢測與線性插值，直到該欄位未存在異常值。

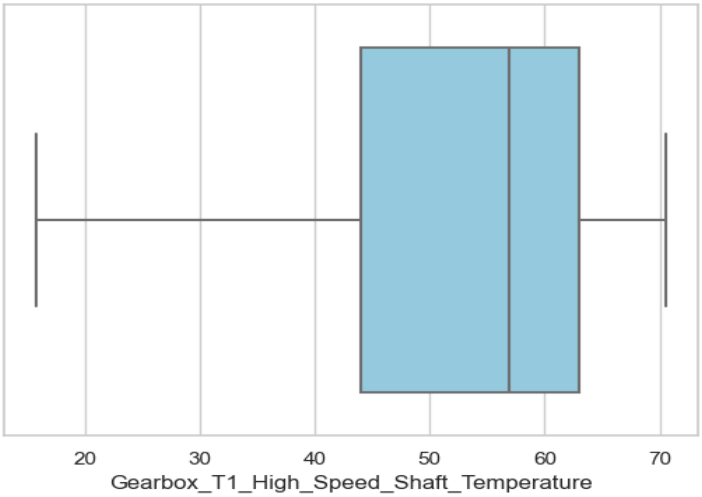

圖 10 線性補值後結果示意圖

由於塔台原始正常加速度(編號 45)、塔偏轉(編號 56)與渦輪機狀態(編號 74)等三個欄位，皆因過多的遺漏值導致無法正常的進行異常值檢測與補值，因此本研究將這三個欄位完全刪除，並重新設定剩餘特徵值欄位的編號，其結果如表 5 所示。

表 5 重新設定後的特徵值欄位說明

編號	變數名稱	單位
0	變速箱 T1 高速軸溫度	°C
1	變速箱 T3 高速軸溫度	°C
2	變速箱 T1 中間高速軸溫度	°C
3	變速箱空心軸承溫度	°C
4	塔台正常加速度	mm/s
5	變速箱 2 油溫	°C
6	塔台側向加速度	mm/s
7	軸承-A 溫度	°C
8	變壓器-3 溫度	°C
9	變速箱 T3 中間高速軸溫度	°C
10	變速箱 1 油溫	°C

11	變速箱油溫	°C
12	力矩	%
13	無功功率控制轉換器	kVAr
14	變壓器 2 溫度	°C
15	無功功率	kVAr
16	軸承 1 溫度	°C
17	變速箱分配器溫度	°C
18	力矩 D 過濾	kNm
19	力矩 D 方向	kNm
20	機組 N 轉速	rpm
21	運作狀態	無
22	功率因數	無
23	軸承 2 溫度	°C
24	機艙溫度	°C
25	電壓 A-N	V
26	軸箱 3 溫度	°C
27	電壓 C-N	V
28	軸箱 2 溫度	°C
29	軸箱 3 溫度	°C
30	電壓 B-N	V
31	機艙位置度數	度
32	電壓控制轉換器	V
33	電池箱 3 溫度	°C
34	電池箱 2 溫度	°C
35	電池箱 1 溫度	°C
36	液壓壓力	bar
37	角轉子位置	度
38	塔基底溫度	°C
39	俯仰偏移非對稱負載控制器 2	度
40	俯仰偏移塔反饋	度
41	電源頻率	Hz
42	內部功率限制	kW

43	斷路器切入	無
44	粒子計數器	無
45	扭矩偏移塔反饋	Nm
46	外部功率限制	kW
47	葉片 2 實際角度 B	度
48	葉片 1 實際角度 B	度
49	葉片 3 實際角度 B	度
50	熱交換控制轉換器溫度	°C
51	塔台原始側向加速度	mm/s
52	環境溫度	°C
53	機艙迴轉	無
54	俯仰偏移非對稱負載控制器 1	度
55	俯仰偏移非對稱負載控制器 3	度
56	每秒風向偏差	度
57	每 10 秒風向偏差	度
58	鄰近感測器 135 度	mm
59	狀態和故障	無
60	鄰近感測器 225 度	mm
61	葉片 3 實際角度 A	度
62	CH4 範圍	無
63	葉片 2 實際角度 A	度
64	葉片 1 實際角度 A	度
65	葉片 2 設置度數	度
66	俯仰需求基線度	度
67	葉片 1 設置度數	度
68	葉片 3 設置度數	度
69	力矩 Q 方向	kNm
70	力矩 Q 過濾	kNm
71	鄰近感測器 45 度	mm
72	鄰近感測器 315 度	mm

4.2.2 平穩性檢測與轉換

在經過異常值檢測與補值後，需要對目標值發電量進行 ADF 與 KPSS 檢驗，以評估資料是否為平穩狀態。透過表 6 的 ADF 檢驗結果可以觀察到 p-value 值為 0.000 小於 0.05，而檢驗統計量 t 值為-20.243 明顯小於任何信賴區間下的臨界值，因此在 ADF 檢驗中可以拒絕 H0 原假設，即發電量時間序列為平穩的數據。

表 6 發電量 ADF 檢驗結果

Test Statistic	-20.243
p-value	0.000
Critical Value (1%)	-3.43
Critical Value (5%)	-2.862
Critical Value (10%)	-2.567

透過表 7 的 KPSS 檢驗可以發現 p-value 值為 0.1 大於 0.05，而檢驗統計量 t 值為 0.187 明顯小於任何信賴區下的臨界值，因此在 KPSS 檢驗中不可拒絕 H0 原假設，即發電量時間序列為平穩的數據。

表 7 發電量 KPSS 檢驗結果

Test Statistic	0.187
p-value	0.1
Critical Value (1%)	0.739
Critical Value (5%)	0.463
Critical Value (10%)	0.347

經由上述 ADF 檢驗與 KPSS 兩種檢驗的結果，可以得知 ADF 與 KPSS 檢驗結果皆為平穩性的數據，因此不須透過季節差分法或簡單差分法的轉換即可進行後續的分析預測。

4.2.3 數值縮放

在經由平穩性的檢測後，需要先將資料縮放至 0 到 1 區間，以解決單位與區間大小的不同，藉此提高模型準確度與降低訓練的時間。在本小節以發電量目標值的前 100 筆資料進行範例展示，結果如圖 11、圖 12 所示，可以看到在未進行數值縮放前資料的範圍非常大，而經過最小值與最大值正規化後，資料的範圍已被縮至 0 到 1 區間，並且也能發現資料的趨勢結構未受影響。

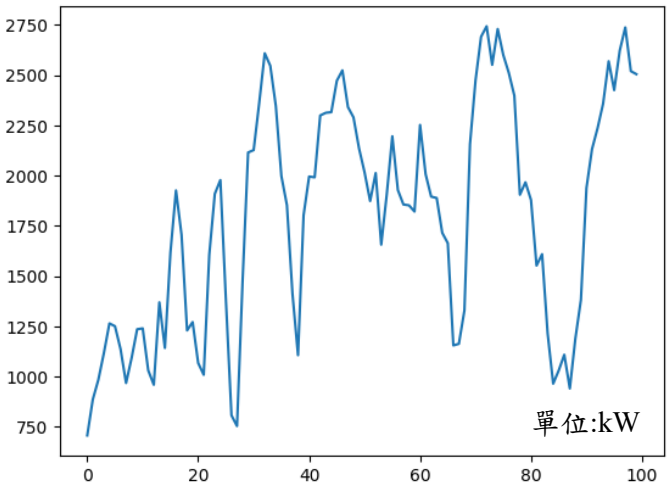

圖 11 正規化前的發電量趨勢圖

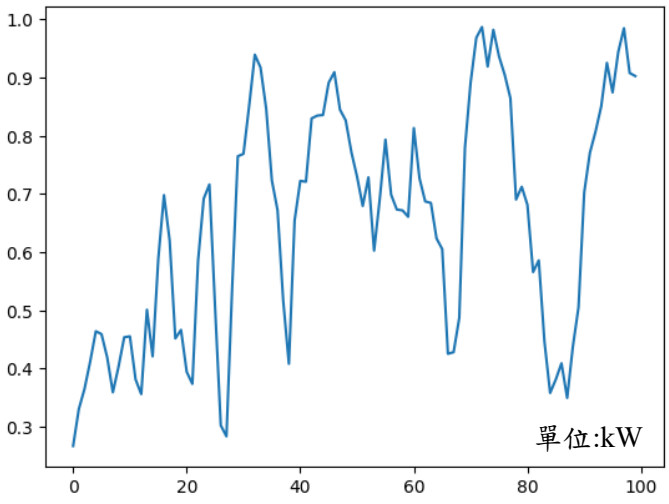

圖 12 正規化後的發電量趨勢圖

4.3 特徵選擇

在將資料匯入到模型前，需先進行特徵選擇來篩選出重要的特徵，以減少資料中冗餘的雜訊，藉此來優化模型的整體效能。本研究所使用的特徵選擇方法為 MI，因此主要是透過 MI 計算出特徵值與目標值的相互依賴程度，來判斷是否為重要的特徵，而本研究主要根據 6 個不同的分數標準來篩選並比較不同特徵數對於預測結果的影響，分數的區間分別為大於 0.5、大於 0.6、大於 0.7、大於 0.8、大於 0.9 以及大於 1.0，其結果如圖 13 至圖 18 所示。

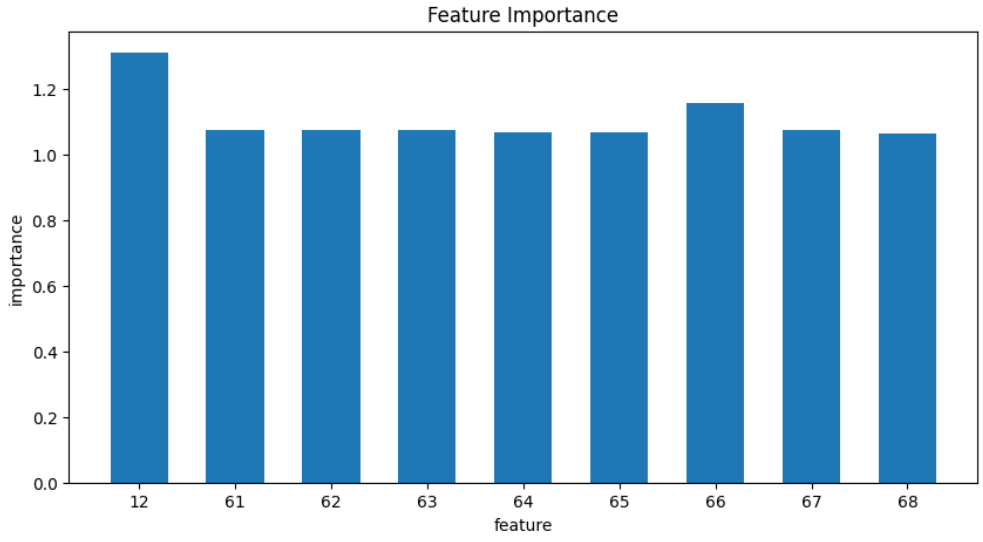

圖 13 重要性大於 1.0 的特徵圖

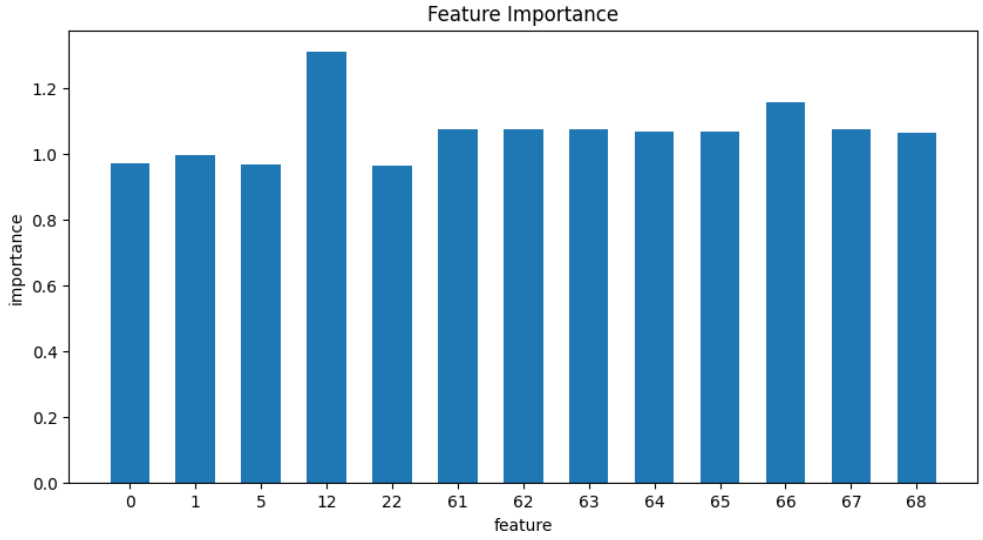

圖 14 重要性大於 0.9 的特徵圖

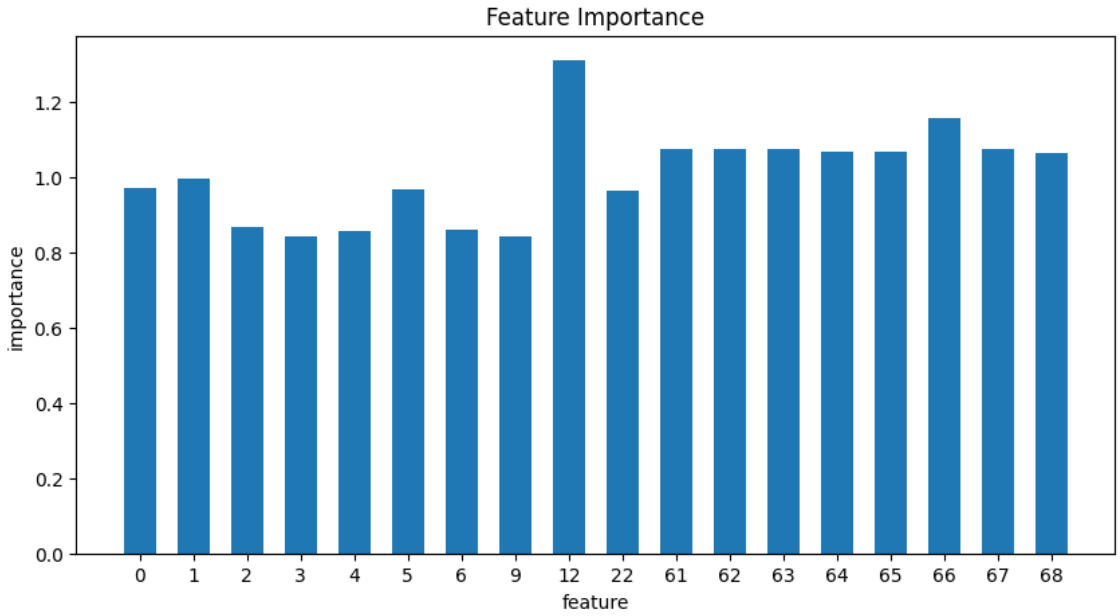

圖 15 重要性大於 0.8 的特徵圖

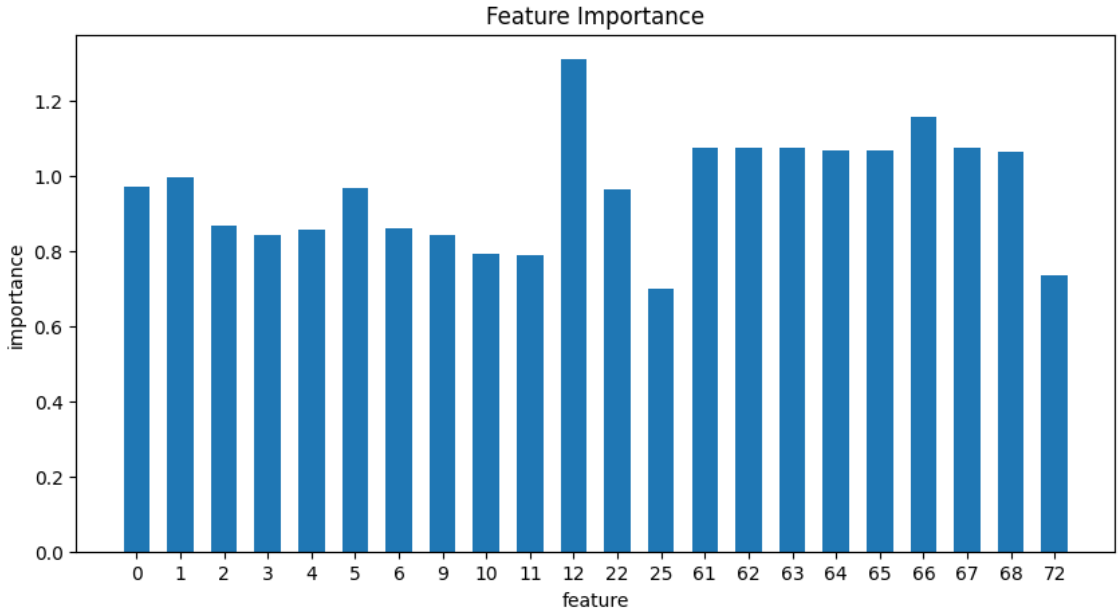

圖 16 重要性大於 0.7 的特徵圖

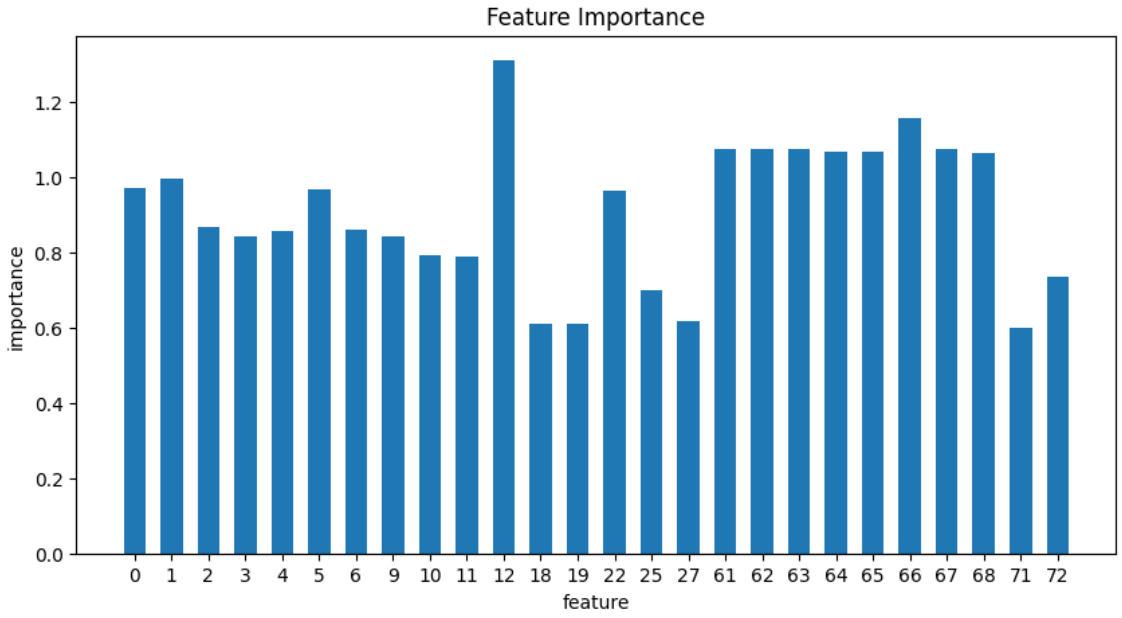

圖 17 重要性大於 0.6 的特徵圖

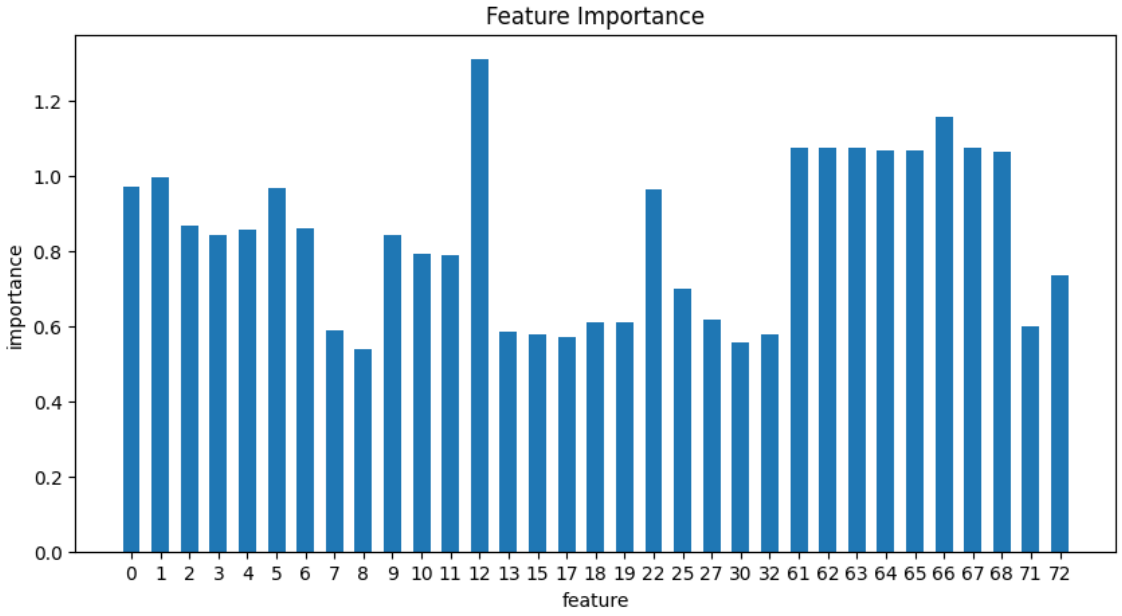

圖 18 重要性大於 0.5 的特徵圖

4.4 模型建立

進行完特徵選擇後，接著需要建立時序模型並輸入重要的特徵，以利後續的分析預測。而在模型建立前需先將輸入的資料拆分為訓練集、驗證集與測試集，而本研究的輸入資料筆數為 136,731 筆，資料集切割按照比例來劃分，訓練集為 82,039 筆(60%)主要作為模型的訓練，驗證集為 27,346 筆(20%)用來進行參數的調整與初步評估，最後測試集為 27,346 筆(20%)用來進行最終模型的績效評估，資料集拆分結果如表 8 所示。

表 8 資料集拆分結果

總資料集(100%)	136,731 筆
訓練集(60%)	82,039 筆
驗證集(20%)	27,346 筆
測試集(20%)	27,346 筆

由於本研究有使用貝氏優化方法進行參數的最佳化，因此在模型建立時需設定需要最佳化的參數與範圍，以便後續模型在訓練時能夠產生最佳的參數結果，參數的設定與範圍如表 9 所示。

表 9 參數的設定與範圍

參數名稱	參數範圍
Memory Cells	(64, 256)
Batch size	(8, 64)
Epochs	(10, 40)
Learning rate	(0.0001, 0.01)

4.5 績效評估

在使用訓練集進行完模型訓練後，將透過驗證集與測試集對模型進行績效評估。本小節將詳細說明不同特徵組合的績效，對於不同特徵數的貝氏優化參數範圍為相同，且訓練集、驗證集與測試集的資料也相同。

4.5.1 發電量預測結果與分析(重要性大於 1.0 特徵)

在重要性大於 1.0 的特徵中，總共包含力矩(編號 12)、葉片 3 實際角度 A(編號 61)、CH4 範圍(編號 62)、葉片 2 實際角度 A(編號 63)、葉片 1 實際角度 A(編號 64)、葉片 2 設置度數(編號 65)、俯仰需求基線度(編號 66)、葉片 1 設置度數(編號 67)與葉片 3 設置度數(編號 68)等 9 個特徵。

本研究經過貝氏優化的調整後，參數設定：Memory Cells 為 230，Batch size 為 8，Epochs 為 39，Learning rate 為 0.00898，其結果如表 10 所示。

表 10 重要性大於 1.0 特徵之參數調整結果

參數名稱	參數設定值
Memory Cells	230
Batch size	8
Epochs	39
Learning rate	0.00898

經過貝氏優化參數調整後，獲得 4 種不同模型的衡量標準，分別為 R^2、RMSE、MAE 與 MSE，其值如表 11 所示，R^2 為 0.95，RMSE 為 157.58，MAE 為 84.63，MSE 為 24,831.46。

表 11 重要性大於 1.0 特徵之績效表

績效指標	績效值
R^2	0.95
RMSE	157.58
MAE	84.63
MSE	24,831.46

4.5.2 發電量預測結果與分析(重要性大於 0.9 特徵)

在重要性大於 0.9 的特徵中，總共包含變速箱 T1 高速軸溫度(編號 0)、變速箱 T3 高速軸溫度(編號 1)、變速箱 2 油溫(編號 5)、力矩(編號 12)、功率因素(編號 22)、葉片 3 實際角度 A(編號 61)、CH4 範圍(編號 62)、葉片 2 實際角度 A(編號 63)、葉片 1 實際角度 A(編號 64)、葉片 2 設置度數(編號 65)、俯仰需求基線度(編號 66)、葉片 1 設置度數(編號 67)與葉片 3 設置度數(編號 68)等 13 個特徵。

本研究經過貝氏優化的調整後，參數設定：Memory Cells 為 75，Batch size 為 17，Epochs 為 16，Learning rate 為 0.00117，其結果如表 12 所示。

表 12 重要性大於 0.9 特徵之參數調整結果

參數名稱	參數設定值
Memory Cells	78
Batch size	17
Epochs	16
Learning rate	0.00117

經過貝氏優化參數調整後，獲得 4 種不同模型的衡量標準，分別為 R^2、RMSE、MAE 與 MSE，其值如表 13 所示，R^2 為 0.95，RMSE 為 122.26，MAE 為 69.29，MSE 為 14,947.51。

表 13 重要性大於 0.9 特徵之績效表

績效指標	績效值
R^2	0.95
RMSE	122.26
MAE	69.29
MSE	14,947.51

4.5.3 發電量預測結果與分析(重要性大於 0.8 特徵)

在重要性大於 0.8 的特徵中，總共包含變速箱 T1 高速軸溫度(編號 0)、變速箱 T3 高速軸溫度(編號 1)、變速箱 T1 中間高速軸溫度(編號 2)、變速箱空心軸承溫度(編號 3)、塔台正常加速度(編號 4)、變速箱 2 油溫(編號 5)、塔台側向加速度(編號 6)、變速箱 T3 中間高速軸溫度(編號 9)、力矩(編號 12)、功率因素(編號 22)、葉片 3 實際角度 A(編號 61)、CH4 範圍(編號 62)、葉片 2 實際角度 A(編號 63)、葉片 1 實際角度 A(編號 64)、葉片 2 設置度數(編號 65)、俯仰需求基線度(編號 66)、葉片 1 設置度數(編號 67)與葉片 3 設置度數(編號 68)等 18 個特徵。

本研究經過貝氏優化的調整後，參數設定：Memory Cells 為 192，Batch size 為 45，Epochs 為 39，Learning rate 為 0.00695，其結果如表 14 所示。

表 14 重要性大於 0.8 特徵之參數調整結果

參數名稱	參數設定值
Memory Cells	192
Batch size	45
Epochs	39
Learning rate	0.00695

經過貝氏優化參數調整後，獲得 4 種不同模型的衡量標準，分別為 R^2、RMSE、MAE 與 MSE，其值如表 15 所示，R^2 為 0.96，RMSE 為 118.71，MAE 為 70.37，MSE 為 14,092.06。

表 15 重要性大於 0.8 特徵之績效表

績效指標	績效值
R^2	0.96
RMSE	118.71
MAE	70.37
MSE	14,092.06

4.5.4 發電量預測結果與分析(重要性大於 0.7 特徵)

在重要性大於 0.7 的特徵中，總共包含變速箱 T1 高速軸溫度(編號 0)、變速箱 T3 高速軸溫度(編號 1)、變速箱 T1 中間高速軸溫度(編號 2)、變速箱空心軸承溫度(編號 3)、塔台正常加速度(編號 4)、變速箱 2 油溫(編號 5)、塔台側向加速度(編號 6)、變速箱 T3 中間高速軸溫度(編號 9)、變速箱 1 油溫(編號 10)、變速箱油溫(編號 11)、力矩(編號 12)、功率因素(編號 22)、電壓 A-N(編號 25)、葉片 3 實際角度 A(編號 61)、CH4 範圍(編號 62)、葉片 2 實際角度 A(編號 63)、葉片 1 實際角度 A(編號 64)、葉片 2 設置度數(編號 65)、俯仰需求基線度(編號 66)、葉片 1 設置度數(編號 67)、葉片 3 設置度數(編號 68)與鄰近感測器 315 度(編號 72)等 22 個特徵。

本研究經過貝氏優化的調整後，參數設定：Memory Cells 為 65，Batch size 為 47，Epochs 為 15，Learning rate 為 0.0062，其結果如表 16 所示。

表 16 重要性大於 0.7 特徵之參數調整結果

參數名稱	參數設定值
Memory Cells	65
Batch size	47
Epochs	15
Learning rate	0.0062

經過貝氏優化參數調整後，獲得 4 種不同模型的衡量標準，分別為 R^2、RMSE、MAE 與 MSE，其值如表 17 所示，R^2 為 0.96，RMSE 為 119.26，MAE 為 78.12，MSE 為 14,222.95。

表 17 重要性大於 0.7 特徵之績效表

績效指標	績效值
R^2	0.96
RMSE	119.26
MAE	78.12
MSE	14222.95

4.5.5 發電量預測結果與分析(重要性大於 0.6 特徵)

在重要性大於 0.6 的特徵中，總共包含變速箱 T1 高速軸溫度(編號 0)、變速箱 T3 高速軸溫度(編號 1)、變速箱 T1 中間高速軸溫度(編號 2)、變速箱空心軸承溫度(編號 3)、塔台正常加速度(編號 4)、變速箱 2 油溫(編號 5)、塔台側向加速度(編號 6)、變速箱 T3 中間高速軸溫度(編號 9)、變速箱 1 油溫(編號 10)、變速箱油溫(編號 11)、力矩(編號 12)、力矩 D 過濾(編號 18)、力矩 D 方向(編號 19)、功率因素(編號 22)、電壓 A-N(編號 25)、電壓 C-N(編號 27)、葉片 3 實際角度 A(編號 61)、CH4 範圍(編號 62)、葉片 2 實際角度 A(編號 63)、葉片 1 實際角度 A(編號 64)、葉片 2 設置度數(編號 65)、俯仰需求基線度(編號 66)、葉片 1 設置度數(編號 67)、葉片 3 設置度數(編號 68)、鄰近感測器 45 度(編號 71)與鄰近感測器 315 度(編號 72)等 26 個特徵。

本研究經過貝氏優化的調整後，參數設定：Memory Cells 為 201，Batch size 為 62，Epochs 為 26，Learning rate 為 0.00973，其結果如表 18 所示。

表 18 重要性大於 0.6 特徵之參數調整結果

參數名稱	參數設定值
Memory Cells	201
Batch size	62
Epochs	26
Learning rate	0.00973

經過貝氏優化參數調整後，獲得 4 種不同模型的衡量標準，分別為 R^2、RMSE、MAE 與 MSE，其值如表 19 所示，R^2 為 0.96，RMSE 為 108.76，MAE 為 62.98，MSE 為 11,828.74。

表 19 重要性大於 0.6 特徵之績效表

績效指標	績效值
R^2	0.96
RMSE	108.76
MAE	62.98
MSE	11,828.74

4.5.6 發電量預測結果與分析(重要性大於 0.5 特徵)

在重要性大於 0.6 的特徵中，總共包含變速箱 T1 高速軸溫度(編號 0)、變速箱 T3 高速軸溫度(編號 1)、變速箱 T1 中間高速軸溫度(編號 2)、變速箱空心軸承溫度(編號 3)、塔台正常加速度(編號 4)、變速箱 2 油溫(編號 5)、塔台側向加速度(編號 6)、軸承-A 溫度(編號 7)、變壓器-3 溫度(編號 8)、變速箱 T3 中間高速軸溫度(編號 9)、變速箱 1 油溫(編號 10)、變速箱油溫(編號 11)、力矩(編號 12)、無功功率控制轉換器(編號 13)、無功功率(編號 15)、變速箱分配器溫度(編號 17)、力矩 D 過濾(編號 18)、力矩 D 方向(編號 19)、功率因素(編號 22)、電壓 A-N(編號 25)、電壓 C-N(編號 27)、電壓 B-N(編號 30)、電壓控制轉換器(編號 32)、葉片 3 實際角度 A(編號 61)、CH4 範圍(編號 62)、葉片 2 實際角度 A(編號 63)、葉片 1 實際角度 A(編號 64)、葉片 2 設置度數(編號 65)、俯仰需求基線度(編號 66)、葉片 1 設置度數(編號 67)、葉片 3 設置度數(編號 68)、鄰近感測器 45 度(編號 71)與鄰近感測器 315 度(編號 72)等 33 個特徵。

本研究經過貝氏優化的調整後，參數設定：Memory Cells 為 110，Batch size 為 14，Epochs 為 17，Learning rate 為 0.00059，其結果如表 20 所示。

表 20 重要性大於 0.5 特徵之參數調整結果

參數名稱	參數設定值
Memory Cells	110
Batch size	14
Epochs	17
Learning rate	0.00059

經過貝氏優化參數調整後，獲得 4 種不同模型的衡量標準，分別為 R^2、RMSE、MAE 與 MSE，其值如表 21 所示，R^2 為 0.96，RMSE 為 112.91，MAE 為 68.92，MSE 為 12,748.67。

表 21 重要性大於 0.5 特徵之績效表

績效指標	績效值
R^2	0.96
RMSE	112.91

MAE	68.92
MSE	12,748.67

第五章、討論

本研究藉由特徵選取中的相互資訊方法快速得從繁多的特徵中選出對模型訓練最有效的輸入，挑選出的特徵如下所示，風力渦輪機機組狀態數據中的變速箱 T1 高速軸溫度(編號 0)、變速箱 T3 高速軸溫度(編號 1)、變速箱 T1 中間高速軸溫度(編號 2)、變速箱空心軸承溫度(編號 3)、塔台正常加速度(編號 4)、變速箱 2 油溫(編號 5)、塔台側向加速度(編號 6)、變速箱 T3 中間高速軸溫度(編號 9)、變速箱 1 油溫(編號 10)、變速箱油溫(編號 11)、力矩(編號 12)、力矩 D 過濾(編號 18)、力矩 D 方向(編號 19)、功率因素(編號 22)、電壓 A-N(編號 25)、電壓 C-N(編號 27)、葉片 3 實際角度 A(編號 61)、CH4 範圍(編號 62)、葉片 2 實際角度 A(編號 63)、葉片 1 實際角度 A(編號 64)、葉片 2 設置度數(編號 65)、俯仰需求基線度(編號 66)、葉片 1 設置度數(編號 67)、葉片 3 設置度數(編號 68)、鄰近感測器 45 度(編號 71)與鄰近感測器 315 度(編號 72)這些特徵對於目標值發電量的預測是最有相關的。透過上述的特徵可以發現大部分多為各機組件的溫度、力矩大小與葉片角度，本研究推測當葉片受到風的帶動，進而導致渦輪機中的各機組件的溫度也會隨之提高，葉片角度則是會受到風流動的方向，並使葉片角度也會跟著做角度的調整，也因此才會篩選出與目標值發電量高度相關的關鍵因素。

以實務層面進行探討如何將研究的成果落實應用在其中也是至關重要的。對於電廠人員來說，確保發電機機組狀況以提供穩定供電是首要目標，過往或許會以一個固定的時程進行檢查及保養，當中仰賴的通常是在電廠內資深的工程師以經驗法則進行處理，當中若是處理不當就可能會有停電或是機台受損的風險，對此本研究期望若可以將研究結果提供給電廠人員作為一個參考依據，來判斷機組狀態並更有效地安排後續的作業排程，藉此來達到電廠整體運作效率的提升。而對於本研究所預測出的結果誤差是否為電廠可接受的範圍，也需要多與專業的電廠工程師與作業人員來進行討論來進行更深入的研究。而在後面也可以將此研究的概念投射進不同的能源產業，像是陽能發電、水力發電、地熱發電等，增加產業在落實方面的意願及提供更好的專業保障。

第六章、結論與未來研究

6.1 結論

本研究利用公開的風力渦輪機機組狀態數據來分析預測發電量，首先前處理使用盒鬚圖與線性補值方法進行異常值的檢測與填值，使數據盡可能保持連續性與完整性，接著透過 ADF 檢驗與 KPSS 檢驗得知目標值發電量為平穩的數據，接著利用最小值與最大值正規化讓數據縮小至[0,1]區間，以提升模型訓練的效率。最後利用相互資訊(Mutual Information)從 76 個特徵篩選出 26 個影響目標值發電量的關鍵因素。

本研究比較了 6 個不同特徵組合的績效結果，並得到 26 個特徵作為輸入可以取得最佳的預測結果，其模型的 R^2 為 0.96，RMSE 為 108.76，MAE 為 62.98，其中 R^2 可以看出本研究在預測上與實際值的表現是良好的，但是，預測與實際值的平均差異很大的 RMSE 表現卻不佳，本研究推論原因可能與資料集本身具有一定相關性，在資料集內，有些資料或許在前處理過後仍為極值，所以間接導致了 R^2 的表現良好，而 RMSE 的績效不佳的情形。目前本研究認為此次研究結果可以間接證明文獻中所提到的(Su et al., 2019)，當模型考慮更多風機狀態的數據確實可以提高模型預測的績效。

6.2 未來研究

對於未來研究可以建構發電量預測系統並建置人機介面供電廠作業人員使用，前台使用者透過人機介面輸入過往的歷史數據來預測下一個時間點的發電量，藉此來達到更即時的預測分析，以利安排後續的作業排程，而後台管理者也能藉由大量的歷史數據不斷更新模型保持最佳的預測效果。本研究透過 Mockplus 軟體進行人機介面塑模，使用者可以輸入歷史數據並選擇預測時段，最後點選分析結果輸出就能產生發電量預測趨勢圖與報表。人機介面示意圖及模擬結果如圖 19、圖 20 所示。

圖 19 人機介面模擬圖

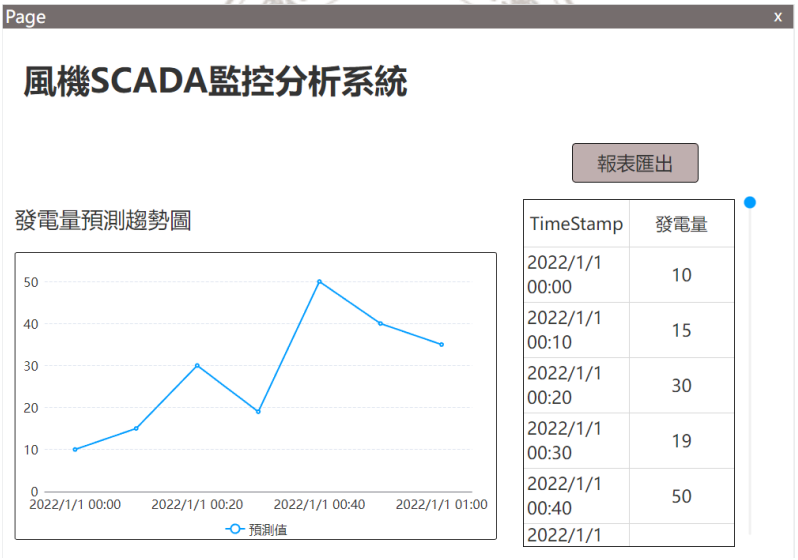

圖 20 人機介面結果產出模擬圖

參考文獻

Aly, H. H. (2020). A novel deep learning intelligent clustered hybrid models for wind speed and power forecasting. Energy, 213, 118773.

Boelrijk, J., Pirok, B., Ensing, B., & Forré, P. (2021). Bayesian optimization of comprehensive two-dimensional liquid chromatography separations. Journal of Chromatography A, 1659, 462628.

Castellani, F., Astolfi, D., & Natili, F. (2021). SCADA data analysis methods for diagnosis of electrical faults to wind turbine generators. Applied Sciences, 11(8), 3307.

Celik, S. (2020). The effects of climate change on human behaviors. In Environment, climate, plant and vegetation growth (pp. 577-589). Springer, Cham.

Chelgani, S. C., Shahbazi, B., & Hadavandi, E. (2018). Support vector regression modeling of coal flotation based on variable importance measurements by mutual information method. Measurement, 114, 102-108.

Eseye, A. T., Zhang, J., & Zheng, D. (2018). Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information. Renewable energy, 118, 357-367.

Evans, A., Strezov, V., & Evans, T. J. (2009). Assessment of sustainability indicators for renewable energy technologies. Renewable and sustainable energy reviews, 13(5), 1082-1088

Ferguson, D., McDonald, A., Carroll, J., & Lee, H. (2019). Standardisation of wind turbine SCADA data for gearbox fault detection. The Journal of Engineering, 2019(18), 5147-5151.

Garan, M., Tidriri, K., & Kovalenko, I. (2022). A Data-Centric Machine Learning Methodology: Application on Predictive Maintenance of Wind Turbines. Energies, 15(3), 826.

Gazafroudi, A. S. (2015). Assessing the impact of load and renewable energies' uncertainty on a hybrid system. International Journal of Energy and Power Engineering, 5(2–2), 1-8.

Hanifi, S., Liu, X., Lin, Z., & Lotfian, S. (2020). A critical review of wind power forecasting methods—past, present and future. Energies, 13(15), 3764.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

Hong, Y. Y., & Rioflorido, C. L. P. P. (2019). A hybrid deep learning-based neural network for 24-h ahead wind power forecasting. Applied Energy, 250, 530-539.

Hossain, M. S., & Mahmood, H. (2020, February). Short-term photovoltaic power forecasting using an LSTM neural network. In 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT) (pp. 1-5). IEEE.

Jaseena, K. U., & Kovoor, B. C. (2019, December). Deep learning based multi-step short term wind speed forecasts with LSTM. In DATA (pp. 7-1).

Jung, J., & Broadwater, R. P. (2014). Current status and future advances for wind speed and power forecasting. Renewable and Sustainable Energy Reviews, 31, 762-777.

Leng, H., Li, X., Zhu, J., Tang, H., Zhang, Z., & Ghadimi, N. (2018). A new wind power prediction method based on ridgelet transforms, hybrid feature selection and closed-loop forecasting. Advanced Engineering Informatics, 36, 20-30.

Lin, Z., & Liu, X. (2020). Assessment of wind turbine aero-hydro-servo-elastic modelling on the effects of mooring line tension via deep learning. Energies, 13(9), 2264.

Lin, Z., & Liu, X. (2020). Wind power forecasting of an offshore wind turbine based on high-frequency SCADA data and deep learning neural network. Energy, 201, 117693.

Lin, Z., Liu, X., & Collu, M. (2020). Wind power prediction based on high-frequency SCADA data along with isolation forest and deep learning neural networks. International Journal of Electrical Power & Energy Systems, 118, 105835.

Liu, X., Zhang, H., Kong, X., & Lee, K. Y. (2020). Wind speed forecasting using deep neural network with feature selection. Neurocomputing, 397, 393-403.

Maseda, F. J., López, I., Martija, I., Alkorta, P., Garrido, A. J., & Garrido, I. (2021). Sensors data analysis in supervisory control and data acquisition (SCADA) systems to foresee failures with an undetermined origin. Sensors, 21(8), 2762.

Salles, R., Belloze, K., Porto, F., Gonzalez, P. H., & Ogasawara, E. (2019). Nonstationary time series transformation methods: An experimental review. Knowledge-Based Systems, 164, 274-291.

Su, Y., Yu, J., Tan, M., Wu, Z., Xiao, Z., & Hu, J. (2019, August). A LSTM based wind power forecasting method considering wind frequency components and the wind turbine states. In 2019 22nd International Conference on Electrical Machines and Systems (ICEMS) (pp. 1-6). IEEE.

Wang, G., Awad, O. I., Liu, S., Shuai, S., & Wang, Z. (2020). NOx emissions prediction based on mutual information and back propagation neural network using correlation quantitative analysis. Energy, 198, 117286.

Wang, Y., Zou, R., Liu, F., Zhang, L., & Liu, Q. (2021). A review of wind speed and wind power forecasting with deep neural networks. Applied Energy, 304, 117766.

Xiang, L., Wang, P., Yang, X., Hu, A., & Su, H. (2021). Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism. Measurement, 175, 109094.

Xiong, B., Meng, X., Wang, R., Wang, X., & Wang, Z. (2021). Combined Model for Short-term Wind Power Prediction Based on Deep Neural Network and Long Short-Term Memory. In Journal of Physics: Conference Series (Vol. 1757, No. 1, p. 012095). IOP Publishing.

Zhang, J., Jiang, X., Chen, X., Li, X., Guo, D., & Cui, L. (2019, April). Wind power generation prediction based on LSTM. In Proceedings of the 2019 4th International Conference on Mathematics and Artificial Intelligence (pp. 85-89).

Zhang, J., Yan, J., Infield, D., Liu, Y., & Lien, F. S. (2019). Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model. Applied Energy, 241, 229-244.

Zhao, Y., Ye, L., Li, Z., Song, X., Lang, Y., & Su, J. (2016). A novel bidirectional mechanism based on time series model for wind power forecasting. Applied energy, 177, 793-803.