

機器學習之超參數優化-以化纖廠之共用設施為例

學生：江孟蒨、吳若琦、鍾惠捷 指導老師：梁韻嘉

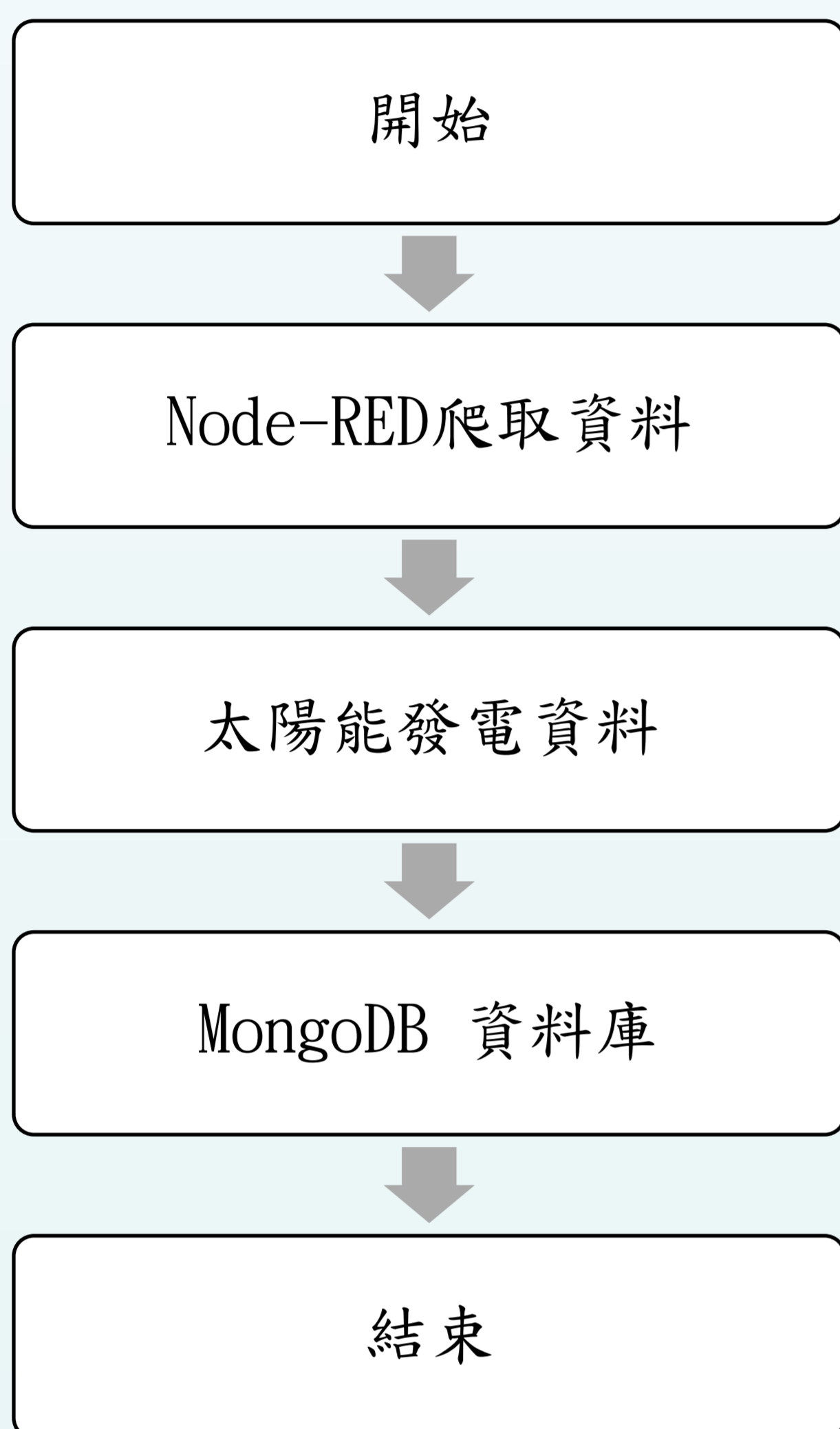
研究動機與目的

本研究是利用以化纖廠之公用設施太陽能發電系統的發電資料來進行訓練及預測，使用病毒最佳化演算法結合機器學習方法藉以進行超參數優化之分析建立日射量預測模型，利用啟發式的方法去設定參數範圍，根據資料量的大小及搭配的機器學習方法不同再來進行比較，得知哪個方法預測結果較準確、較合適，使人員能夠更有效率的監控路燈設施系統。

研究方法

資料蒐集及系統後端

本研究爬取的資料為個案公司內太陽能發電系統的網頁，所以利用網路爬蟲的方式爬取每10分鐘一筆的太陽能發電資料。



資料拆分

完整資料使用的資料是從2020/05/27—2021/05/27

完整資料量

- 完整資料26986筆
- 春季資料6031筆
- 夏季資料7427筆
- 秋季資料6817筆
- 冬季資料6711筆

平衡資料:刪除中間數據，留前後各兩個小時。

平衡資料量

- 完整資料11321筆
- 春季資料2529筆
- 夏季資料3313筆
- 秋季資料2855筆
- 冬季資料2804筆

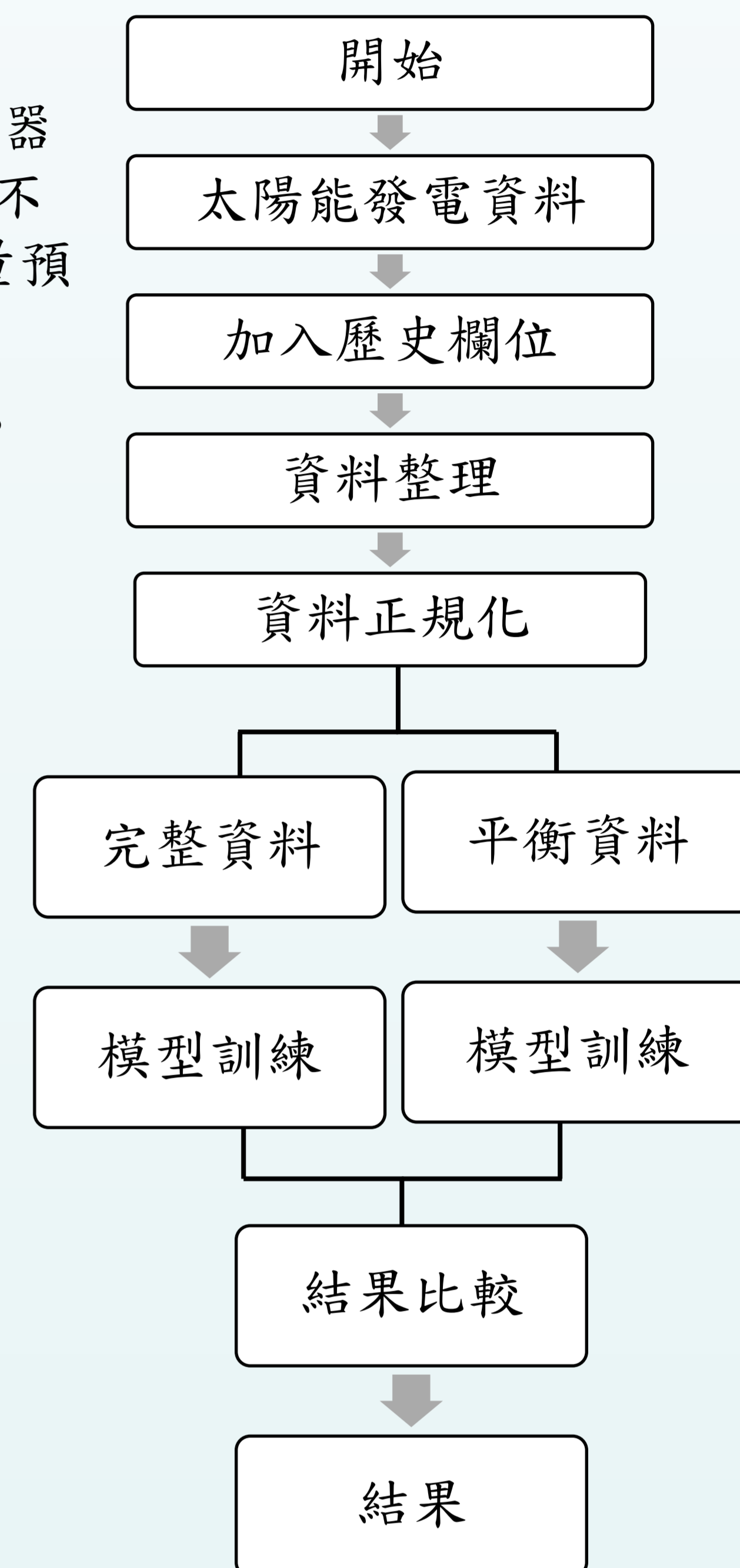
模型訓練方法

預測訓練流程圖(如右圖)，在使用機器學習預測模型部分，本研究提出5種不同機器學習分類模型進行太陽日射量預測的研究，分別為

- K-近鄰演算法(K-nearest neighbors algorithm)
- 隨機森林(Random Forest)
- 自適應增強(AdaBoost)
- 支援向量機(Support Vector Machine)
- 多層感知器(Multilayer perceptron)

資料彙整

| 特徵選擇 |
|--------|
| 變流器轉換率 |
| 降雨機率 |
| 即時發電量 |
| 日射量 |
| 太陽能板溫度 |



研究成果

此研究有三種指標，Recall代表實際應該要關燈但預測沒關燈，Precision代表預測認為要關燈但實際不關燈，以公司層面來考慮不管是安全還是省電方面都很重要，所以此研究已Accuracy為基準去選擇最佳模型。進行比對後認為完整資料ADA的模型最為穩定，以下為ADA所產生的結果。

Tab1.完整資料ADA Accuracy最佳參數

| Test | n_estimators | learning_rate | algorithm | criterion | max_depth | min_samples_split | min_samples_leaf | max_features |
|----------|--------------|---------------|-----------|-----------|-----------|-------------------|------------------|--------------|
| 0.956904 | 65 | 0.108676 | SAMME.R | Entropy | 357 | 2 | 0.004344 | log2 |

Tab2.完整資料五種機器學習方法之分數比較

| 模型方法 | 資料使用方法 | 績效指標 | All | SPRING | SUMMER | FALL | WINTER |
|------|----------|-----------|--------|--------|--------|--------|--------|
| KNN | Accuracy | Accuracy | 0.9547 | 0.9567 | 0.9679 | 0.9501 | 0.9496 |
| | | Recall | 0.8684 | 0.8572 | 0.7571 | 0.8921 | 0.9097 |
| | | Precision | 0.8840 | 0.8743 | 0.8596 | 0.8834 | 0.9116 |
| SVM | Accuracy | Accuracy | 0.9543 | 0.9577 | 0.9655 | 0.9489 | 0.9497 |
| | | Recall | 0.8723 | 0.8696 | 0.8446 | 0.7605 | 0.8977 |
| | | Precision | 0.8838 | 0.8687 | 0.7789 | 0.8118 | 0.9192 |
| RF | Accuracy | Accuracy | 0.9567 | 0.9608 | 0.9690 | 0.8751 | 0.9520 |
| | | Recall | 0.8886 | 0.8789 | 0.7888 | 0.9060 | 0.9245 |
| | | Precision | 0.8796 | 0.8744 | 0.8463 | 0.8754 | 0.9063 |
| ADA | Accuracy | Accuracy | 0.9569 | 0.9583 | 0.9686 | 0.9514 | 0.9533 |
| | | Recall | 0.8965 | 0.8789 | 0.7919 | 0.9072 | 0.9242 |
| | | Precision | 0.8718 | 0.8543 | 0.8307 | 0.8756 | 0.9105 |
| MLP | Accuracy | Accuracy | 0.9482 | 0.9600 | 0.9678 | 0.9512 | 0.9085 |
| | | Recall | 0.8091 | 0.8758 | 0.8462 | 0.9092 | 0.9283 |
| | | Precision | 0.9059 | 0.8616 | 0.8336 | 0.8648 | 0.9132 |

Tab3.完整資料ADA的混淆矩陣

| | | Prediction | | |
|-----------|--------------|------------|-------|------|
| accuracy | Actual Value | 0(關燈) | 1(開燈) | |
| | | 0(關燈) | 5,660 | 664 |
| | 1(開燈) | 451 | 4,546 | |
| | | Prediction | | |
| recall | Actual Value | 0(關燈) | 1(開燈) | |
| | | 0(關燈) | 5,061 | 1263 |
| | 1(開燈) | 3,996 | 1,001 | |
| | | Prediction | | |
| precision | Actual Value | 0(關燈) | 1(開燈) | |
| | | 0(關燈) | 5,674 | 650 |
| | 1(開燈) | 462 | 4,535 | |

Tab4.完整資料五種機器學習方法之時間比較

| 模型方法 | 資料使用方法 | All | SPRING | SUMMER | FALL | WINTER |
|------|----------|-----------------|--------------|-----------------|----------------|-----------------|
| KNN | Accuracy | 1:03:14 | 0:14:56 | 0:09:34 | 0:07:41 | 0:08:36 |
| SVM | Accuracy | 0:59:43 | 0:07:41 | 0:06:06 | 0:05:45 | 0:07:30 |
| ADA | Accuracy | 5:06:14 | 1:25:32 | 1:37:03 | 0:59:50 | 1:05:43 |
| RF | Accuracy | 4:01:12 | 1:11:33 | 1:56:20 | 1:46:36 | 1:52:22 |
| MLP | Accuracy | 4 days, 6:42:30 | 2days, 11:17 | 6 days, 5:14:09 | 1 day, 1:01:26 | 1 day, 14:40:47 |

結論

本研究未來展望為嘗試更多不同的訓練模型方法，來找出最適合廠內路燈開關的模型，提升穩定性及準確度，建立更精確的日射量預測不只能更有效的控制路燈的開關，也可以降低投資建設的風險，協助評估建設太陽能發電系統，推動再生能源的使用。

路燈的開關決策，利用工業工程的最佳化方法來考量電力成本以及天氣狀況等因素，期許未來能往自動化邁進，能夠更有效率的監控路燈系統，並且達到節約能源的永續發展，也使廠內因路燈照明造成的危險降到最低，找出最佳化路燈開啟的方式。