

讓綠能觸手可及-根據區域微氣候資料預測發電量競賽

下載與上傳檔案格式

比賽簡介：

太陽能為全國主要再生能源供電來源，為了因應太陽能的不穩定，需要智慧電網的協助，提供更為準確的預測。本次比賽即是利用花東地區之微氣候數據資料來預測太陽能發電瓦數，並在比賽過程中增加運用 AI 的能力和對太陽能的認識。

檔案格式皆為 csv，請詳閱以下內容：

(1) 訓練資料，共有 17 份，檔名如下：

L1_Train.csv, L2_Train.csv, L3_Train.csv, L4_Train.csv
L5_Train.csv, L6_Train.csv, L7_Train.csv, L8_Train.csv
L9_Train.csv, L10_Train.csv, L11_Train.csv, L12_Train.csv
L13_Train.csv, L14_Train.csv, L15_Train.csv, L16_Train.csv
L17_Train.csv

(2) 上傳結果檔案，共有 1 份，檔名如下：

upload.csv

(說明：競賽期間之實際上傳結果檔案命名，可由參賽者自行定義，不需與本說明文件舉例之檔案名稱相同)

以下將依序講解：

- 1.訓練資料
- 2.比賽方式
- 3.補充事項
- 4.上傳結果檔案
- 5.答案算法
- 6.評分方式

1. 訓練資料

資料包含大約 2000 天的數據，提供 17 個太陽能監測地點的太陽能板設備附近之微氣候數據(分開成 17 個 csv 檔，即 L1_Train.csv~L17_Train.csv)，除了時間與地點之外還包含下列 5 個微氣候特徵與 1 個標籤。總共 7 個特徵，一個標籤。

- 7 個特徵：地點代號 LocationCode(1-17)、時間 DateTime(Y/M/D/hour/minute/second)、風速 WindSpeed(m/s)、大氣壓力 Pressure(hPa)、溫度 Temperature(°C)、濕度 Humidity(%)、亮度 Sunlight(Lux)

- 1 個標籤：太陽能板每分鐘平均發電量 Power(mW)

如圖 1 所示，訓練資料的時間序列基本上為**每分鐘**一行，但有時會因感測器異常、電力中斷或是人為因素，致使資料缺失。某些天 9:00 之後的數據資料將作為題目答案之用，所以也會移除。

LocationCode	DateTime	WindSpeed(m/s)	Pressure(hpa)	Temperature(°C)	Humidity(%)	Sunlight(Lux)	Power(mW)
Example File	2024-01-01 6:29:01	0	1022.91	17.9	63.1	22.5	0
	2024-01-01 6:30:01	1.7	1023.01	17.9	63.1	29.17	0
	2024-01-01 6:31:01	1.5	1023.06	17.9	63.2	35	0.01
	2024-01-01 6:32:02	0.69	1022.97	17.9	63.4	38.33	0
	2024-01-01 6:33:02	0	1022.95	17.9	63.3	42.5	0.01
	2024-01-01 6:34:08	0	1016.58	17.5	86.7	39.17	0
	2024-01-01 6:35:01	0	1022.97	17.9	63.4	52.5	0.01
	2024-01-01 6:36:01	0	1022.97	17.9	63.4	56.67	0.01
	2024-01-01 6:37:01	0	1023.03	17.9	63.4	65	0.01
	2024-01-01 6:38:08	0	1016.56	17.5	86.9	74.17	0.01
	2024-01-01 6:39:01	0	1023.07	17.9	63.4	84.17	0.01
	2024-01-01 6:40:01	0.37	1023	17.9	63.6	96.67	0.01
	2024-01-01 6:41:08	0	1016.51	17.5	86.9	115	0.01
	2024-01-01 6:42:08	0	1016.53	17.4	86.7	127.5	0

圖 1、範例氣象數據(僅供格式參考)

2. 比賽方式

每一題為預測某一天 9:00(含)之後的每一筆(每 10 分鐘)的平均發電量(mW)，持續預測直至當天 16:59(含)，共有 48 筆預測資料，圖 2 為比賽方式示意圖。可以參照氣象局公布的相關天氣數據來推算預測值。此次比賽公開測試題目與正式測試題目將會混和給出，總共 200 題，等同需要預測 9600 筆(200*48)，題目為指定裝置需預測的時間段，將在 upload.csv 內提供，參賽者須將答案(即該時段的預測發電量)填寫後上傳。

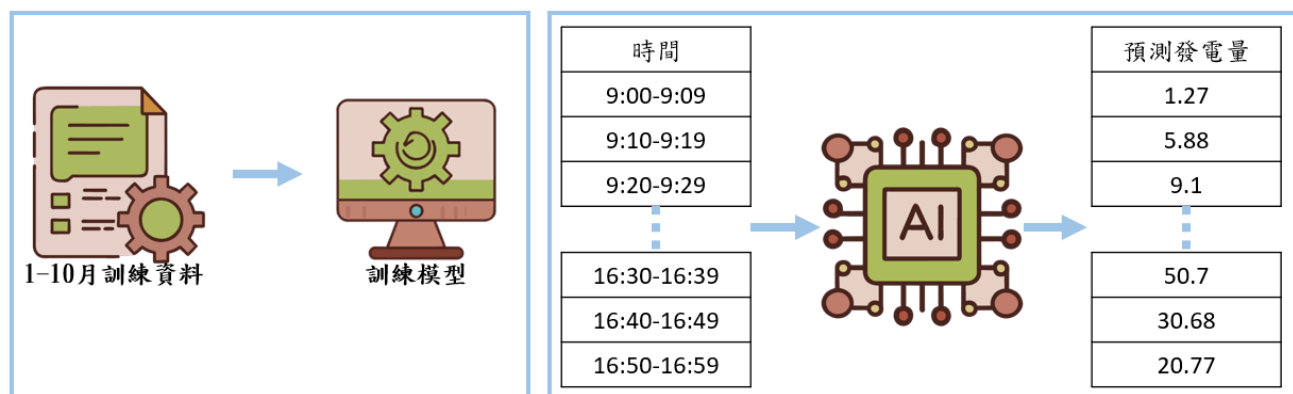


圖 2、訓練與預測示意圖(左圖為訓練、右圖為預測)

3. 補充事項

3.1. 感測計缺陷補償

因光照度感測計最大數值為 117758.2 lux，當陽光十分充足時感測計數值會停留在此大值，但太陽能板所提供的發電量還在持續上升。每個地點所收集到的光照度資料使用迴歸分析後的 R2(可解釋度)皆大於 70%，代表光照度特徵對標籤還具有極高的解釋力，因此參賽者可自行決定是否針對此缺陷進行數值上的修正，以增加預測發電量準度。因本次比賽主要是預測發電量，可以利用迴歸分析先分析光照度計數值尚未達到最大值前的光照度對發電量的比率，再回推發電量產生的光照度。

3.2. 風速計異常

因部分地點風速計時好時壞，有異常情況時，風速的數值會直接補 0，參賽者可以自行決定是否要參考此特徵。

4. 上傳結果檔案格式

上傳結果檔案為 upload.csv，格式為：

序號西元年(4 碼)+月(2 碼)+日(2 碼)+預測時間(4 碼)+裝置代號 (2 碼)，總共 14 碼，其中預測時間代表要預測往後十分鐘的平均發電瓦數，例如 0900，表示要預測 09:00 至 09:09(含)間的平均發電瓦數，發電單位為毫瓦特(mW)。請參考圖 3 提供的範例結果檔案。

序號	答案
20240218090001	
20240218091001	
20240218092001	
20240218093001	
20240218094001	
20240218095001	
20240218100001	
20240218101001	
20240218102001	
20240218103001	

圖 3、範例結果檔案(僅供格式參考)

範例：

- 序號 20240218090001 意即需要預測裝置 01 在 2024/02/18 09:00~09:09 的平均發電瓦數
- 序號 20240220165010 意即需要預測裝置 10 在 2024/02/20 16:50~16:59 的平均發電瓦數

上傳時請將預測答案填至對應之答案欄位，並請注意答案數值應**取四捨五入至小數點後兩位**，例如預測答案若為 18.3789 請填寫 18.38。

5. 答案算法

本次比賽的實際答案算法請參考以下邏輯：

將實際資料每 10 分鐘的數據取平均值，並以四捨五入法取至小數點後 2 位。

平均方式：

為同個小時內區分成 6 個時段，每個時段內的數值都必須取平均

第一時段-->0 分 00 秒 - 9 分 59 秒

第二時段-->10 分 00 秒 - 19 分 59 秒

第三時段-->20 分 00 秒 - 29 分 59 秒

第四時段-->30 分 00 秒 - 39 分 59 秒

第五時段-->40 分 00 秒 - 49 分 59 秒

第六時段-->50 分 00 秒 - 59 分 59 秒

答案算法的示範如圖 4，9:00 至 9:09 為同一時段，取平均後四捨五入取至小數點後兩位，此時段之 10 分鐘平均發電瓦數(答案)為 35.06，第二個時段以此類推，10 分鐘平均發電瓦數(答案)為 129.85。

時間	實際答案
09:00	10.5
09:01	12.67
09:02	30.88
09:03	40
09:06	50.5
09:09	65.78
09:10	90.9
09:11	100
09:12	111.1
09:13	111.5
09:15	111.9
09:16	130
09:17	135.5
09:18	177.57
09:19	200.2

第一時段	35.06
第二時段	129.85

圖 4、實際答案算法(僅供格式參考)

6. 評分方式

將每一筆預測資料與正確資料相比並計算誤差(取絕對值)，並將所有誤差加總在一起，以加總誤差值為排名依據，加總誤差越少排名越高。

計分公式如下：

假設總共有 n 筆預測資料，每筆資料的正確答案為 S_i ，參賽者的答案為 A_i (其中 $i=1,2,\dots,n$)。

總分數= $\sum_{i=1}^n |A_i - S_i|$ 。

總分數越低者則排名越高。

範例：

共有 5 筆預測資料，正確發電量為 $S = [5,10,15,20,25]$ ，參賽者 1 的答案為 $A_1 = [1,2,3,4,5]$ 參賽者 2 的答案為 $A_2 = [30,40,50,60,70]$

則 參賽者 1 的總分數為 $4+8+12+16+20=60$

參賽者 2 的總分數為 $25+30+35+40+45=175$

參賽者 1 的排名會高於參賽者 2。