

# 機器學習－期末報告

林薇如 313652026

December , 2025

## Contents

1	AI 的未來能力	2
2	實現未來能力所需的成分 (Ingredients)	2
3	涉及的機器學習類型	3
4	簡化模型	3
4.1	模型一：監督式學習預測年度電力需求 . . . . .	3
4.1.1	模型目的 . . . . .	3
4.1.2	資料集 . . . . .	4
4.1.3	模型選擇：隨機森林 (Random Forest) . . . . .	4
4.2	模型二：能源決策最佳化模型 . . . . .	5
4.2.1	決策變數與目標 . . . . .	5
4.2.2	經濟成本 . . . . .	5
4.2.3	氣候成本與損失權重 . . . . .	5
4.2.4	總損失函數 . . . . .	5
4.2.5	最佳化求解 . . . . .	6
5	討論與結論	6
5.1	模型整體預測 . . . . .	6
5.2	特徵重要性 (Feature Importance) . . . . .	7
5.3	模型限制 . . . . .	7
5.4	整體結論 . . . . .	7

# 1 AI 的未來能力

目前的人工智慧雖能協助進行氣候預測與能源分析，但仍屬於被動的輔助工具。它可以告訴我們氣溫將上升、能源需求將增加，卻無法主動提出全球層級的永續解決方案。我認為在未來二十年內，AI 有可能達成一項具有重大意義的能力——自主發現並實施全球能源與氣候最佳化策略。

這項能力的核心在於整合多層次的資訊，包括氣候變化模型、能源供需、經濟活動以及人類行為。未來的 AI 不僅能預測氣候趨勢，更能在模擬環境中試驗政策：例如在虛擬城市中模擬不同碳稅制度與政策的長期影響，尋找環境與經濟的平衡點。如此一來，AI 不僅僅是分析工具，而是進化為決策者，在全球層面協調人類行為以減緩氣候變遷。

## 2 實現未來能力所需的成分 (Ingredients)

**資料 (Data)** AI 必須同時掌握氣候、能源、經濟與政策資訊等大規模多模態資料，以支撐政策模擬及長期預測。

### 建模與推理工具 (Tools)

- PDE：描述氣候與能源的動態行為
- Graph Neural Networks：處理電力網、交通流等網路結構

**硬體與環境 (Hardware / Environment)** 需要高效能運算 (HPC) 來運行地球系統模擬，以及能回饋即時資料的感測器。

### 學習架構 (Learning Setup)

- self-supervised：學習氣候與社會資料的結構
- reinforcement learning：在模擬環境中學習政策策略
- meta-learning：快速適應不同區域與情境
- RLHF：結合人類對政策偏好的評價

這些學習方式相互補充，讓系統能夠在複雜且開放的環境中持續改善策略。

### 3 涉及的機器學習類型

- **監督式學習 (Supervised Learning)**：可用來建立氣候與能源預測模型，從歷史資料中學習氣溫、降雨量、碳排放與能源使用等之間的關聯。資料來源包括氣象觀測、衛星影像與能源市場資料及產業活動紀錄；目標訊號則是實際測量到的氣候或能源指標。
- **非監督式學習 (Unsupervised Learning)**：在複雜的多維資料中自動辨識具有相似能源結構或氣候特徵的地區，發現潛在模式。
- **強化學習 (Reinforcement Learning)**：讓 AI 能在模擬的氣候—經濟環境中自主嘗試政策行動，根據獎勵訊號（如溫度穩定、碳排減少、經濟成長）學習長期最優策略。

## 4 簡化模型

本研究提出兩個互相銜接的模型：

1. **監督式學習模型**：預測年度電力需求；
2. **能源決策最佳化模型**：決定最佳再生能源比例。

兩者結合後，AI 即可根據未來情境（GDP、氣溫、人口等）自動推估能源需求，再進行最佳政策規劃。

### 4.1 模型一：監督式學習預測年度電力需求

#### 4.1.1 模型目的

**年度電力需求**  $D$  是能源政策評估中的關鍵變數，受到多種因素影響，包含工業生產指數 IPI  $I$ 、氣溫  $T$ 、人口  $P$ 、匯率 (元/美元)  $r$ 、國內生產毛額 GDP  $G$ 、國民所得毛額 GNI  $N$  等。為了在後續的能源最佳化模型中進行政策推估，我們首先建立一個監督式學習模型，以歷史資料訓練年度需求的預測函數：

$$\hat{D} = f_{\theta}(I, T, P, r, G, N),$$

其中  $f_{\theta}$  是具有可學習參數  $\theta$  的回歸模型。

#### 4.1.2 資料集

**資料來源** 中華民國統計資訊網、交通部統計查詢網、經濟部能源署能源統計專區。

**資料範圍** 民國 90 年至民國 113 年。

**資料格式處理** 所有資料經過適當轉換，統一整理為「以季為時間間隔」的格式。

**訓練集與測試集切分** 本研究將民國 110 年以前的觀測值作為訓練集，111 年以後的觀測值作為測試集。

#### 4.1.3 模型選擇：隨機森林（Random Forest）

本研究採用**隨機森林（Random Forest）** 作為需求預測模型。隨機森林透過建立多棵回歸樹並取平均，能有效降低單棵決策樹的高變異性，提升模型穩定度與預測準確度。

**模型數學形式** 令輸入特徵為

$$x = (I, T, P, r, G, N),$$

隨機森林由  $M$  棵回歸樹  $\{T_m\}_{m=1}^M$  組成，其預測函數可寫為

$$\hat{D} = f_{\theta}(x) = \frac{1}{M} \sum_{m=1}^M T_m(x),$$

其中  $\theta$  表示所有樹的參數集合， $T_m(x)$  為第  $m$  棵樹對輸入  $x$  的預測。

每一棵回歸樹會將特徵空間以遞迴切分成多個區域  $R_{mj}$ ，並在該區域中給出一個常數預測值  $c_{mj}$ ，因此可寫成：

$$T_m(x) = \sum_{j=1}^{J_m} c_{mj} \mathbf{1}(x \in R_{mj}),$$

其中  $\mathbf{1}(x \in R_{mj})$  為指示函數（indicator function）。

#### 為何適用於能源需求預測

- 能捕捉非線性，例如溫度上升可能以非線性方式推動電力需求；
- 能處理特徵交互作用，例如人口與氣溫同時上升時的合併效果；
- 能提供特徵重要性（feature importance），有助於解釋哪些因素影響需求最多。

## 4.2 模型二：能源決策最佳化模型

### 4.2.1 決策變數與目標

在已知年度能源需求為  $\hat{D}$  的前提下，政府需決定再生能源比例  $x \in [0, 1]$ ，目標為最小化總社會成本：

$$J(x) = C_{\text{econ}}(x) + C_{\text{clim}}(x)$$

其中  $C_{\text{econ}}$  為再生能源投資成本， $C_{\text{clim}}$  為化石燃料排放造成的氣候成本。

### 4.2.2 經濟成本

將經濟成本設計為凸成本，以二次函數建模：

$$C_{\text{econ}}(x) = ax^2, \quad a > 0.$$

反映「比例愈高愈困難」的經濟成本特性。當再生能源比例提高時，會因最佳場址先被使用、需更大儲能與電網升級、發電調度更困難等原因使邊際成本遞增。

### 4.2.3 氣候成本與損失權重

化石燃料供應比例為  $(1 - x)$ ，排放量為  $E(x) = (1 - x)e_F \hat{D}$ ，其中  $e_F$  為每單位化石能源的排放係數。氣候成本為：

$$C_{\text{clim}}(x) = \lambda e_F \hat{D}(1 - x).$$

其中  $\lambda$  為氣候損失權重。

**氣候損失權重  $\lambda$  的意義**  $\lambda$  是一個參數，將排放量轉換成對應的損失，用來衡量排放一單位  $\text{CO}_2$  所造成的氣候與經濟損害在決策中的重要性。

- $\lambda$  大：強調減碳  $\rightarrow$  最佳解  $x^*$  高；
- $\lambda$  小：偏重經濟  $\rightarrow$  最佳解  $x^*$  低。

### 4.2.4 總損失函數

令  $B = \lambda e_F \hat{D}$ ，則總損失為：

$$J(x) = C_{\text{econ}}(x) + C_{\text{clim}}(x) = ax^2 + B(1 - x).$$

#### 4.2.5 最佳化求解

對  $x$  求導數並令其為零：

$$\frac{dJ}{dx} = 2ax - B, \quad x^* = \frac{B}{2a} = \frac{\lambda e_F \hat{D}}{2a}.$$

考慮  $x \in [0, 1]$ ：

$$x^* = \min \left\{ 1, \max \left( 0, \frac{\lambda e_F \hat{D}}{2a} \right) \right\}.$$

## 5 討論與結論

以下結論與討論針對模型一：

### 5.1 模型整體預測

隨機森林模型在季資料下達到：

- $R^2 = 0.8295$
- $MAE = 1,880,939$  千度
- $MAPE = 2.64\%$

代表模型對整體趨勢掌握程度高，能捕捉季節性與長期成長。但季節資料常帶有高變異，而年資料經過季節加總後噪音降低，能預測更準確的年度總需求。

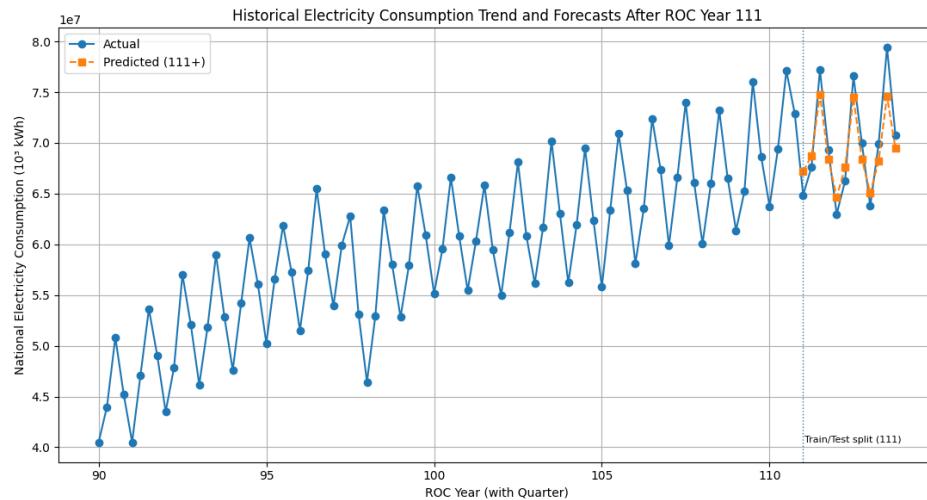


Figure 1: 模型一預測季用電量

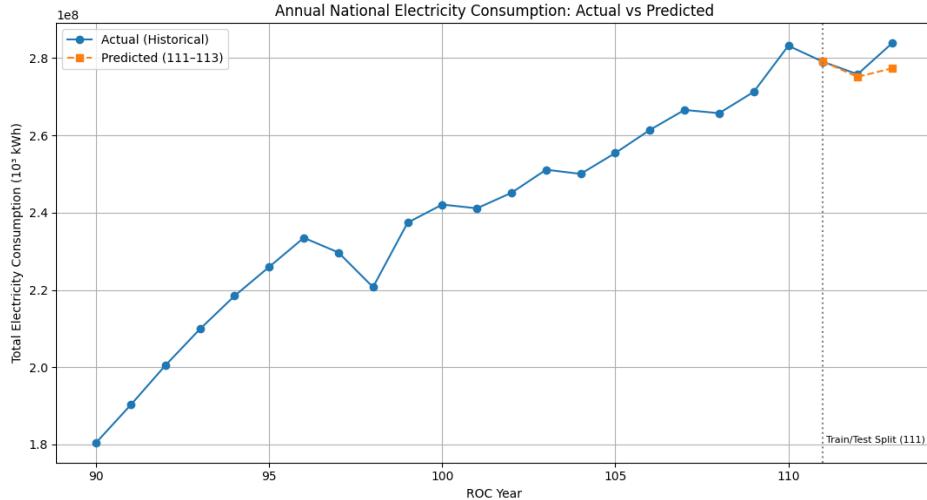


Figure 2: 模型一預測季用電量加總得年用電量

## 5.2 特徵重要性 (Feature Importance)

依據 feature importance，主要影響預測的特徵為：

- 工業生產指數 IPI (63%)：顯示工業活動是用電主因，也反映真實經濟——台灣用電量主要由製造業與工業支撐
- 氣溫 (15%)：顯示夏季冷氣負載明顯
- 國內生產毛額 GDP (11%)

## 5.3 模型限制

- 對極端高溫、特殊產業政策、突發性經濟波動較不敏感
- 隨機森林模型本身不擅長外推極端事件 (extrapolation)
- 夏季用電預測偏差較大
- 特徵未包含更多能源政策與價格資訊

## 5.4 整體結論

整體而言，模型一能有效掌握台灣用電需求的主要變動因素，並提供穩定且可信的季與年需求預測。雖然在極端氣候與政策性事件上的反應仍有限，但模型已展現監督式學習於能源預測中的實用性。未來若加入更多政策與價格相關特徵，預測能力將能進一步提升。