|  |
| --- |
| 第一組 |
| **進階程式設計期末報告Word檔案** |
| 預測性維護資料集(AI4I 2020) |

|  |
| --- |
| **工二乙 S1231132 許芸禎**  **工二乙 S1231136 黃崇岫**  2025/6/20 |

**一、清晰的問題定義（含實際舉例）**

在傳統金屬加工產線中，例如使用銑床進行鋁合金零件切削製造時，現場常面臨以下幾項實務困境：

* 突發性設備故障導致整條產線停擺：

某廠在例行接單生產過程中，因機台主軸軸承過熱導致臨時停機，不僅影響交期，還造成原材料浪費與人力閒置。事後才發現設備早有異常振動與扭力異常跡象，卻未被即時監控與回報。

* 刀具磨耗無法即時監測，影響加工品質：

現場操作員依經驗每週固定更換刀具，結果某次因過早更換造成刀具資源浪費，而另一次則因過度使用導致工件毛邊與尺寸誤差增加，產品須報廢處理。

* 維修排程僅依靠固定週期或人員經驗：

一家中小型製造廠仍仰賴人工登記機台運轉狀態與每月巡檢，無法精準掌握每台設備真實負載與健康狀況，導致有些機台提前維修、有些機台則錯過最佳保養時機。

這些問題凸顯了傳統製造業缺乏數據驅動維修決策機制，加上現場感測器數據雖已逐步建立，但尚未發揮其潛在價值。

**本研究欲解決的核心問題：**

如何透過現有的簡單感測資料（如轉速、扭力、溫度、刀具磨耗等），建立能提早預測故障風險與設備磨耗程度的機器學習模型，進而讓傳統加工產線邁向數位化與智慧維護。

**具體任務目標：**

**1. 分類任務 – Machine Failure（機器故障）預測：**

根據即時感測數據判斷某筆資料是否代表即將發生故障，協助管理者提早進行預防性維修。

**2. 回歸任務 – Tool Wear（刀具磨耗時間）預測：**

根據設備的運作參數預測刀具已使用多久、何時應更換，作為動態排程依據，降低品質風險與資源浪費。

藉由 AI4I 2020 資料集建模，本研究希望驗證：在未改裝設備、只依靠原有感測數據的情況下，是否可發展出具實用價值的預測性維護模型，進一步促進中小企業在智慧製造的轉型中也能有所著力。

**二、數據前處理與處理方式**

AI4I 2020 資料集包含10,000筆由約2,000台設備所蒐集的運作紀錄，是典型具時間序列性與多機台特徵的工業資料。

1. **主要變數涵蓋：**

* 運作條件：扭力（Torque）、轉速（Rotational Speed）、空氣與製程溫度（Air/Process Temperature）
* 結果指標：刀具磨耗（Tool Wear）、是否故障（Machine Failure）

一張含有 文字, 螢幕擷取畫面, 陳列, Rectangle 的圖片

AI 產生的內容可能不正確。

一張含有 圖表, 螢幕擷取畫面, Rectangle, 行 的圖片

AI 產生的內容可能不正確。

* 而L型刀具最多，其次是M，最後是H，針對刀具磨耗進行預測，三類（L、M、H）均約 100 分鐘，平均磨損時間穩定。
* 鬚範圍：

M：約 0 至 250 分鐘。

L：約 0 至 250 分鐘。

H：約 0 至 240 分鐘，變異性最小。

無異常值：三類無黑點，顯示磨損時間分佈正常。

**2. 類別型特徵：設備型號（Type）、五種子故障類別（TWF~RNF）**

一張含有 圖表, 圖畫, 地圖 的圖片

AI 產生的內容可能不正確。

**3. 前處理步驟詳述：**

* 缺失值與異常值處理：資料完整無缺失；對極端值採保留策略，考量其可能反映異常運作模式，有助於模型學習異常行為的邊界。
* 資料型態轉換：針對設備型號（Type）進行 One-Hot Encoding 以利模型處理。
* 類別不平衡處理：故障樣本僅佔3%，高度不平衡。此種分佈若僅使用 Accuracy 評估，將導致模型偏向預測「非故障」類別，無法有效辨識真正的故障風險。
* 特徵標準化策略：保留物理單位與實際量綱，避免對模型解釋性造成干擾，並有助於後續部署與實務解釋。
* 資料集切分：使用8:2比例分割為訓練集與測試集，保持模型泛化能力與效能評估信度。

**三、可解釋的洞察與視覺化分析**

變數間的物理與統計關聯

* Torque 與 Power：高度正相關（r=0.98），反映出能量傳輸過程中，功率受扭力影響為主，驗證機械物理特性。
* Rotational Speed 與 Torque：強烈負相關（r=-0.88），顯示負載與轉速間的能量平衡。
* Air Temperature 與 Process Temperature：呈現穩定線性正相關，可作為加工環境熱穩定性參考依據。一張含有 螢幕擷取畫面, 正方形, Rectangle, 樣式 的圖片

  AI 產生的內容可能不正確。

故障樣本洞察

* 故障樣本明顯集中在高轉速與高扭力區域，表示複合高負載情境是潛在故障熱區。
* 多變量交互作用顯示：高功率 + 高磨耗 → 故障風險顯著升高。

刀具磨耗觀察

* 故障刀具的磨耗時間明顯偏高，顯示「過度使用」是主因。
* 建議模型加入「Torque × 使用時間」交互項，有望強化非線性磨耗預測能力。

**四、模型簡明解釋與可信的效能評估**

**1.分類模型（Machine Failure 預測）**

在預測設備是否發生故障的任務中，本研究測試以下六種常見的分類模型，並以 Precision、Recall、F1-score 與 ROC AUC 作為主要比較依據：

**分類任務指標（以 Decision Tree 為例）**

* Accuracy：0.981
* Precision：0.677
* Recall：0.721（主要關注指標）
* F1-score：0.698
* ROC-AUC：0.855一張含有 文字, 螢幕擷取畫面, 字型 的圖片

  AI 產生的內容可能不正確。

此結果顯示模型能在保持整體高正確率的同時，具備優異的「召回潛在故障」能力，適用於預警導向的實務場景。

**（1）Decision Tree（決策樹）**

擁有最高的 Recall（0.7213），在本研究以「找出最多故障」為首要目標下，表現最佳。其模型結構清晰、可視化程度高，適合應用於需要可解釋性與現場快速部署的狀況。F1-score（0.6984）也為所有模型中最佳，顯示其在 precision 與 recall 間取得良好平衡。

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

AI 產生的內容可能不正確。

一張含有 文字, 螢幕擷取畫面, 行, 繪圖 的圖片

AI 產生的內容可能不正確。

**（2）Random Forest（隨機森林）**

雖然整體 Accuracy 與 ROC AUC 最高（0.9835 和 0.964），但 Recall 僅為 0.5901，略低於決策樹。若在實務應用中重視「不要誤報太多」，可考慮調整其閾值進行部署，作為輔助模型。

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

AI 產生的內容可能不正確。  
一張含有 文字, 螢幕擷取畫面, 行, 圖表 的圖片

AI 產生的內容可能不正確。

**2.回歸模型（Tool Wear、Power、Process Temperature 等預測）**

針對連續變數預測任務（如Power、Process temperature、Tool wear），本研究針對數值型變數間的關聯性進行建模分析，測試了以下七種回歸模型，並以R²、RMSE、MAE、MAPE 四項誤差指標進行模型效能評估：

一張含有 文字, 字型, 螢幕擷取畫面, 平行 的圖片

AI 產生的內容可能不正確。

以下提供的分析模型回最佳回歸模型，分別針對**扭力(Torque)與功率(Power)、功率 (Power)與轉速(Rotational speed)和空氣溫度(Air temperature)與製成溫度 (Process temperature)**，三種進行數值型變數間關聯性分析。

**（1）KNearest Neighbors Regression（KNN 回歸）**

雖在 Torque → Power 任務中的 R² 表現最佳，但 RMSE、MAE、MAPE 指標略高，顯示其在部分資料區段仍有擬合誤差，泛化能力略低於線性與正則化模型。

一張含有 螢幕擷取畫面, 繪圖, 行, 文字 的圖片

AI 產生的內容可能不正確。

（上圖為扭矩與功率KNN回歸圖）

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

AI 產生的內容可能不正確。

(上圖為扭矩與功率誤差指標與單位)

**（2）Rotational Speed → Power-Polynomial Regression（二次回歸）**

在處理 Rotational Speed → Power 等具明顯非線性趨勢的變數對時，表現最為優異，R²、MAE、MAPE 皆為最佳。可有效擬合彎曲趨勢，適合進行故障發展曲線建模與刀具磨耗趨勢預測。

一張含有 螢幕擷取畫面, 文字, 繪圖, 行 的圖片

AI 產生的內容可能不正確。

（上圖為轉速與功率二次回歸圖）

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

AI 產生的內容可能不正確。

(上圖為轉速與功率誤差指標與單位)

**（3）Support Vector Regression（SVR）**

整體表現不佳，在 Torque、Rotational Speed 與 Power、Tool Wear 等模型中，R² 偏低、誤差指標普遍偏高，擬合曲線亦出現異常趨勢。雖於 Air Temperature → Process Temperature 中表現良好（四項指標最佳），但在其他變數對中不建議使用。

一張含有 螢幕擷取畫面, 文字, 繪圖, 行 的圖片

AI 產生的內容可能不正確。

（上圖為製程溫度與空氣溫度KNN回歸圖）

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

AI 產生的內容可能不正確。

(上圖為製程溫度與空氣溫度誤差指標與單位)

**回歸任務指標補充**

* R²：解釋變異能力（越接近 1 越佳）
* RMSE / MAE / MAPE：衡量預測值與實際值間的誤差大小與穩定性
* 交叉驗證（K-fold CV）：避免過擬合，提高泛化能力

**六、明確的工業價值與建議**

1.分類模型（Decision Tree）應用價值

* 風險預警：高召回率有助於提早發現異常，降低停機風險。
* 現場可視化訓練：結構直觀，易於培訓現場技術人員。
* 資源分配：可結合風險評級，彈性調整維修優先順序。

2.回歸模型（Polynomial Regression）應用價值

* 磨耗趨勢預測：可建立刀具剩餘壽命模型，優化更換時機。
* 節能監控：以功耗曲線輔助參數最佳化，提升能效。
* 異常偵測：提早識別連續數值異常，補足分類模型盲點。

3.綜合應用建議

建議採雙模型架構：

* 回歸模型監控參數變化趨勢
* 分類模型判斷是否進入風險區域
* 結合結果用於排程、警示與維修判斷

**七、模型的局限性與下一步**

1.目前限制

* 資料極度不平衡造成少數類辨識困難，易漏檢潛在風險設備。
* 缺乏時間序列與過程特徵，無法精確建模磨耗進程。
* 子故障（TWF ~ RNF）尚未建模，難以支援細緻化維護策略。
* 感測器維度受限，未涵蓋如振動、電流、噪音等多源訊號。

2.未來發展方向

1. 引入過採樣技術（如 SMOTE）：平衡類別分佈，提升召回能力。

2. 導入時間序列模型（如 LSTM）：強化 Tool Wear 進程建模。

3. 子故障類型建模：發展多分類策略，細緻推估不同故障機制。

4. 拓展感測器維度：納入多源異質資料進行多模態融合。

5. 規劃場域部署測試：實地驗證模型可行性與回饋機制。