

1. 問題目標：

以加工機台完整的「加工參數」和「加工品質」作為訓練資料，於預測階段預測 20 項重點參數。是屬於迴歸預測型的問題。

2. 資料描述

讀取原始資料集（包含訓練集與測試集）之後發現有以下問題需要處理：

1. 原始資料集當中存在文字（字串）類型的資料，無法直接使用
2. 原始資料當中包含遺失值（空白的儲存格）

3. 解題流程：

3-1 資料清理與特徵工程（包含訓練集與測試集）

1. 文字類型的資料，轉換成數值類型的資料（完成轉換後的資料欄位會倍增）
2. 填補遺失值（本模型使用 [MICE](#) 方法來填補所有資料當中的遺失值）
3. 以六台鑽孔機台加工參數、共同設定參數、以及個別機台的加工品質數據，作為分組依據，分別對上述各組原始資料計算其 rolling mean、rolling standard deviation、以及 rolling range 等人工特徵。

3-2 模型訓練與結果預測

1. 使用清理過後的原始資料搭配人工特徵資料作為模型的輸入項、以各組的目標參數作為預測目標，先訓練一個模型，作為產生「預測目標的合成資料」(synthetic data) 的模型使用。
2. 再以步驟 1 的模型所產生出來的合成資料，以「解題流程：步驟 3」的人工特徵產生方式，產生對應的人工特徵
3. 再以步驟 2 所產生出來的合成資料、合成資料之人工特徵、與步驟 1 所使用的訓練資料，整併在一起作為最終的訓練資料集，並以該資料集訓練第二階段的模型。第二階段的模型即用於產生最終預測值使用。

舉例：以預測「Input_A2_016」、「Input_A2_017」、「Input_A2_024」這三項參數為例。

1. 以「Output_A2」、「Output_A2 的人工特徵」、「Output_A2」、「扣除掉 016、017、024 的其他 21 項 Input_A2_參數資料」以及其人工特徵、「所有的共同項參數 C」。以上述資料作為輸入項，並以「Input_A2_016」、「Input_A2_017」、「Input_A2_024」作為模型預測目標，先訓練第一階段的模型，以產生出「Input_A2_016」、「Input_A2_017」、

「Input_A2_024」的合成資料。

2. 使用第一階段模型所產生出來的三個欄位的合成資料，分別建構其人工特徵。
3. 於步驟 3，將步驟 1 所使用的輸入項、與第一階段模型所產生出來的合成資料以及其人工特徵，交互使用以訓練第二階段的模型。

例如，當我們要訓練一個用來產生「Input_A2_016」最終預測結果的第二階段模型時，其模型的輸入項，除了第一階段所使用到的輸入項資料之外、還包含於步驟 1 所產生出來的「Input_A2_017」、「Input_A2_024」的合成資料、以及步驟 2 所產生出來的「Input_A2_017」、「Input_A2_024」合成資料的人工特徵，依此類推。

3-3 其他說明

1. 在預測階段時的流程，與上述 3-2 的流程相同，僅差別在跳過訓練階段，皆使用已訓練好的第一階段與第二階段模型，搭配清理完畢之測試資料集，來預測最終結果。
2. 由於在訓練 A1、A2、A3 與 A6 的模型時，會使用到共同項參數 C，但是部分共同項參數 C，卻也同時是本題目當中所指定的預測目標。故在建模流程上，會先針對要預測的共同項參數 C，以所述的模型訓練流程，訓練出相對應的模型，並取得最終的預測結果。再使用該最終預測結果，以所述訓練流程，分別建構 A1、A2、A3 與 A6 的模型。
3. 所有在本專案當中的模型參數，均已使用訓練資料集，搭配人工進行調教。