

本研究旨在透過多元線性回歸模型，建立健身進度與卡路里消耗的預測模型，並嚴格遵循 **CRISP-DM 方法論** 完成八個階段的實作。共通流程始於資料載入、EDA 與系統化的**特徵工程**，並實施相關性檢查以**避免資料洩漏**。

在特徵選擇環節，專案必須執行此步驟，並採用三種方法（如 RFE、SelectKBest）的綜合投票機制，最終選定了 16 個特徵。模型比較部分，基於 **Linear Regression 基線**，納入 **Ridge/Lasso/ElasticNetCV** 等正則化模型進行訓練與交叉驗證（CV）。

最佳實務要求提供 CV 與測試集的 **RMSE/MAE/ R^2** 評估，其中最佳模型（Linear Regression）在測試集上達到 $R^2 = 0.9006$ 。為了**殘差診斷**，專案提供了殘差圖與 Q-Q 圖，並通過 Shapiro-Wilk 正態性檢驗 ($p = 0.9141$)，驗證了模型的嚴謹性。繪製 **95% 預測區間帶** 也是關鍵成果，以量化預測的不確定性。此流程確保了模型的穩定性與可解釋性，避免了因診斷不足而產生的常見統計陷阱。 (199 words)