

三、研究計畫內容（以中文或英文撰寫）：

- (一) 研究計畫之背景。請詳述本研究計畫所要探討或解決的問題、研究原創性、重要性、預期影響性及國內外有關本計畫之研究情況、重要參考文獻之評述等。如為連續性計畫應說明上年度研究進度。

1. 摘要

本計畫旨在開發一套智慧化水質資料分析與評估系統，以解決過去在環境數據分析上缺乏高效率方法的問題。儘管於現今資料蒐集已可透過物聯網（IoT）等方式以提升工作效率，但數據分析仍需耗費大量的人力時間。傳統統計工具如 Microsoft Excel、SAS 或 SPSS 雖然能用以資料分析，但在操作上卻須高度仰賴使用者的專業知識。因此本計畫基於人工智慧（AI）技術，運用 Scikit-learn 自由軟體機器學習（ML）框架，建構適用於水質分析之多元多項式迴歸（MPR）模型，並部署於 Google 雲端平臺（GCP），以提供物聯網裝置遠端呼叫 MPR 模型以進行分析運算的能力。本計畫的核心在於研究如何提升機器學習技術於水質資料分析，藉由文獻回顧可知，使用 MPR 模型能夠深入探討各式水質資料之間的交互關係，找出潛在的水質健康風險因子，並能給予明確的改善建議。本計畫將使用 Python 程式語言進行機器學習模型開發，其主要最佳化過程包括資料前處理、特徵工程與模型訓練等。實作時將結合使用最小平方方法和梯度下降法，以提高模型訓練的準確性與泛化能力。此外搭配使用 R^2 和 MAE 等性能指標對 MPR 模型進行研究評估，以確保所訓練之模型能夠有效達成數據分析上的效能需求，最終結合建構完成的 MPR 模型，開發一套完整的智慧化水質資料分析與評估系統。水資源管理是環境生態保育重要的一環，本計畫跨領域結合資訊與環境工程技術，預期對於人工智慧與生態保育的研究發展具有實質效益。

關鍵詞：人工智慧、機器學習、多元多項式迴歸、Scikit-learn、水質分析

Keywords：Artificial Intelligence、Machine Learning、Multivariate Polynomial Regression、Scikit-learn、Water Quality Analysis

2. 研究動機

如圖 1 所示，本計畫旨在開發一套具機器學習（Machine Learning, ML）能力的智慧化水質資料分析程式，以做為目前在水質資料分析上缺乏高效率方法的解決方案之一。計畫提案人曾擔任科技部計畫兼任研究助理，協助執行「運用智慧型綠能永續景觀噴泉促進環境保育效益之研究」（計畫編號：MOST 110-2410-H025-024），運用物聯網（Internet of Things, IoT）技術蒐集來自南投麒麟潭的水質資料，其自動化運行的結果相較於人工記錄更有效率，然而後續分析資料時卻仍需耗費大量的工作時間。

關於水質資料分析，除了使用 Microsoft Excel 直接進行計算外，通常亦會使用如 Statistics Analysis System（SAS）或 Statistical Product and Service Solutions（SPSS）等統計專有用軟體（Proprietary Software）。然而此類套裝程式並非開源自由軟體，需花費購買且需投入額外的時間成本學習才能熟練使用，其操作高度仰賴使用者的專業知識，因此分析結果可能會因人而異。且因程式碼無法修改與物聯網設備直接進行資料串聯傳輸，造成整合實現自動化資料收集與數據分析的障礙。為了解決以上問題，本計畫將運用機器學習技術，研究開發一套智慧化水質分析程式，實現更有效率且易於操作的創新水質分析方法，為人工智慧物聯網（Artificial Intelligence of Things, AIoT）在水資源管理和環境生態保育方面的應用，帶來技術上實質的創新與進步。

3. 研究問題

機器學習被視為統計學的現代延伸，有助於快速地整理資料並進行分析，可準確且有效率地應用於水質情況的評估。據我們的了解，如本計畫結合物聯網技術與人工智慧方法進行水質資料分析的研究在目前仍是相對少見。除了原有運用物聯網自動運作的優勢之外，預期經由機器學習深度分析水質資料（含水溫、酸鹼值、溶氧量、導電度、氧化還原電位，與各項微量元素含量如總氮、總磷、氨氮、亞硝酸鹽、硝酸鹽、磷酸鹽等），還能進一步探討各項數據間的複雜交互作用關係，找出潛在的水質健康風險因子。水資源的管理是環境生態保育重要的一環，本計畫的目標除了能夠對水質的健康狀況進行綜合評估之外，並能給予明確的改善建議。我們藉由文獻回顧與技術探討，評估以建構基於多元多項式迴歸（Multiple Polynomial Regression, MPR）機器學習模型進行水質資料分析具有可行性。為此我們將基於 Scikit-learn 自由軟體機

器學習庫，建構一個適合應用於分析水質資料的 MPR 模型，相較於一般的線性迴歸（Linear Regression, LR）與多項式迴歸（Polynomial Regression, PR）模型，MPR 能夠更完整與深入地分析各項水質資料與其之間的交互關係。綜合以上所述，本研究跨領域結合資訊與環境工程，預計對於包含人工智慧技術與水資源生態保育的研究發展具有實質效益。

4. 研究方法

4.1 程式開發架構

4.2 水質研究工具

4.3 水質研究方法

4.3.1 資料獲取

4.3.2 資料匯入

4.3.3 資料前處理

4.3.3.1 遺失值處理

4.3.3.2 型態調整

4.3.3.3 資料拆分

4.3.3.4 特徵工程

4.4 多元多項式迴歸模型之建構

4.4.1 模型訓練

4.4.1.1 最小平方法

4.4.1.2 梯度下降法

4.4.2 模型評估與測試

4.4.2.1 決定係數

4.4.2.2 均方根誤差

4.4.2.3 平均絕對誤差

4.5 部署及應用

4.5.1 雲端平台部署

4.5.2 前端程式開發

5. 研究步驟

(三) 預期完成之工作項目及成果。請分年列述：1. 預期完成之工作項目。2. 對於參與之工作人員，預期可獲之訓練。3. 預期完成之研究成果（如實務應用績效、期刊論文、研討會論文、專書、技術報告、專利或技術移轉等質與量之預期成果）。4. 學術研究、國家發展及其他應用方面預期之貢獻。

(四) 整合型研究計畫說明。如為整合型研究計畫請就以上各點分別說明與其他子計畫之相關性。