

運用機械學習模型預測廢水處理效率

指導教授:黃皓 學生:陳品樺 謝竹筠 張語軒 2022.12.29

研究動機及目的

隨著科技發展和工業化使得廢水大量增加，工廠生產過程中會產生含有雜質的廢水，若將未經妥善處理的有毒物化學物質與危險材料排放至河流，將造成嚴重的環境污染，人類也會遭受影響。為了能夠更有效且合法地處理含有雜質的廢水，本組將會使用多項數據進行測試與分析。

我們希望可以準確預測出廢水廠處理廢水的效率，比較我們得到的數據和之前老數據的差異，最後提出我們的觀點並總結結論。



實驗設備

測試設備 : ASUSPRO P1440

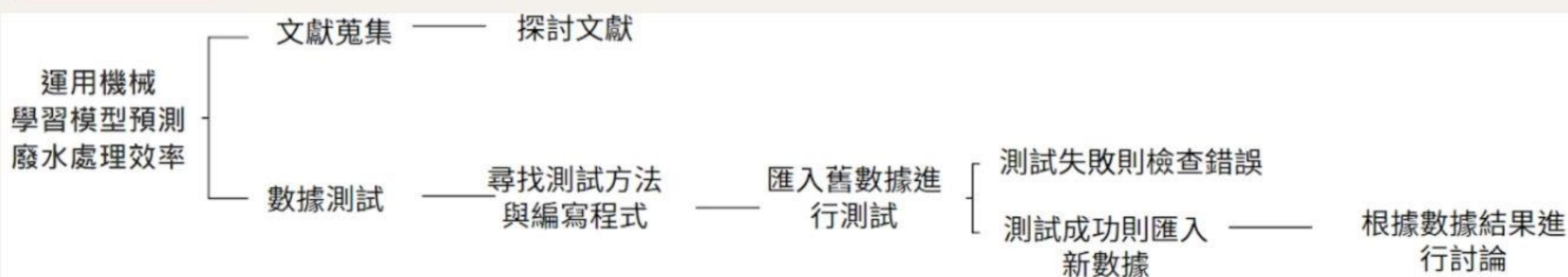
測試軟體 : Anaconda 的 spyder

撰寫程式時所使用的語言 : Python

研究工具

- OLS:全名為Ordinary Least Square，中文名為最小平方法回歸模型，是使散點圖上的所有觀測值到回歸直線距離的平方最小。
- SVR:全名為Support Vector Regression，中文名為支持向量回歸模型，SVR能在多種可能中找出最佳解，讓我們可以靈活地定義模型中可接受的誤差。

操作流程



研究方法

1. 計算工具

均方誤差 (MSE) 是最常用的回歸損失函數，計算方法是求預測值與真實值之間差距的平方和，可以評價數據的變化程度，數學特性很好，使計算梯度變得更容易。

均方根誤差 (RMSE) 是迴歸預測模型的兩種主要效能指標之一，可衡量預測值和實際值之間的平均差異，藉此估計預測模型預測目標值的準確度。

梯度增強樹(GBT)是一種迭代的決策樹算法，這種算法由多棵決策樹組成，每一棵樹學習的都是之前所有樹結論的殘差，就是真實值和預測值之間的差值，以此類推直到殘差小於某個接近0的閾值或回歸樹數目達到某一閾值，不斷擬合殘差，使殘差不斷減少。產出的結果是RMSE。

投票回歸器(Voting Regressor)是將不同的機械學習回歸器組合起來，並返回平均預測值，對於一組同樣表現良好的模型可以起到平衡各自弱點的作用，其預測結果是所有投票結果的平均值。產出的結果是MSE。

數據結果

RMSE	新數據 OLS	新數據 SVR	新數據 ANN	新數據 Gradient Tree Boosting	新數據 Voting regressor
Original	41.6077	41.6961	38.3153	2.2381	41.6077
t-1	39.4837	40.5592	37.6838	2.2381	39.4837
t-2	39.3169	40.3895	37.2015	2.2381	39.3169
t-3	39.2053	40.5993	37.8305	2.2381	39.2053
t-4	39.1062	40.7999	36.1902	2.2381	39.1062
t-5	39.0062	41.1088	36.7490	2.2381	39.0062

結論

- 舊數據不能代表目前廢水處理的狀況，故專注於新數據的測試結果。
- 透過研究時使用不同方法測試新數據的結果得知出 Gradient Tree Boosting 的 RMSE 值最小。
- Gradient Tree Boosting 是最適合這個問題的方法。