**基於機器學習的**

**混凝土性能預測報告**

第二組

S0822034吳翊瑄 S0822126呂映萱

**目錄**

**簡介**

**資料探索**

**資料分析**

**模型選擇**

**模型評估**

**找出最適模型**

**最終模型評估**

**最終模型選擇**

**參考資料**

**簡介**

1. 建築結構和基礎：

混凝土是建築結構和基礎的主要組成部分。它提供了結構的強度和穩定性，能夠承受負載和抗震能力。混凝土構件如柱、樑、牆壁和地基等都是建築物的重要組成部分，確保了建築物的安全和耐久性。

1. 道路和橋梁建設：

混凝土在道路和橋梁建設中起著重要的作用。它能夠承受交通負載，提供平穩的行駛表面，並具有耐久性和長壽命。混凝土道路和橋梁可以改善交通流量，連接城市和地區，促進經濟發展。

1. 水利工程：

混凝土在水利工程中被廣泛應用，如水壩、堤壩、河道結構、水池和排水系統等。混凝土的高強度和耐久性能使其能夠抵抗水壓和環境侵蝕，確保水利設施的穩定運行和安全性。

1. 工業設施和基礎設施：

混凝土在工業設施和基礎設施建設中扮演重要角色。例如，混凝土被用於建造工廠、倉庫、停車場、機場跑道等，提供了穩定的基礎和耐久的結構。

1. 可持續建築：

混凝土可以被製作成綠色和可持續的建築材料。它的生產過程可以利用再生資源和工業副產品，減少對自然資源的消耗。混凝土的長壽命和低維護成本使其成為可持續建築的理想選擇。

綜上所述，混凝土的重要性體現在它在建築、基礎設施和工程領域的廣泛應用。它提供了結構的強度、耐久性。

**資料探索**

可能影響混擬土強度(Strength)的變數：

Cement（水泥）、Blast Furnace Slag（高爐礦渣）、Fly Ash（飛灰）、Water（水 ）、Super-plasticizer（減水劑）、Coarse Aggregate（粗骨料）、Fine Aggregate（細骨料）、Age（年齡）。

觀察每個變數與"混擬土強度"的高低關係：

Cement（水泥）

一張含有 螢幕擷取畫面 的圖片

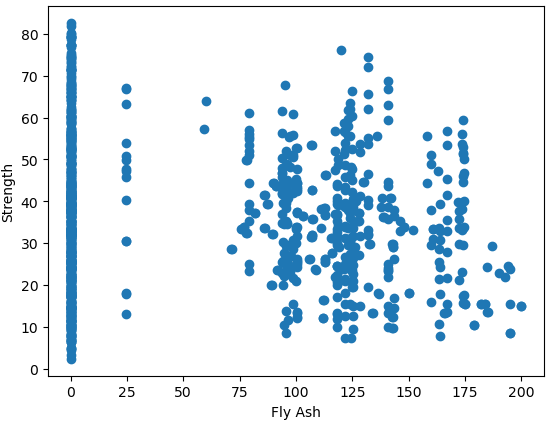
自動產生的描述

Blast Furnace Slag（高爐礦渣）

一張含有 螢幕擷取畫面 的圖片

自動產生的描述

Fly Ash（飛灰）



Water（水 ）

一張含有 螢幕擷取畫面, 圖表 的圖片

自動產生的描述

Super-plasticizer（減水劑）

一張含有 螢幕擷取畫面, 文字 的圖片

自動產生的描述

Coarse Aggregate（粗骨料）

一張含有 文字, 螢幕擷取畫面, 雪花 的圖片

自動產生的描述

Fine Aggregate（細骨料）

一張含有 文字, 螢幕擷取畫面, 雪花 的圖片

自動產生的描述

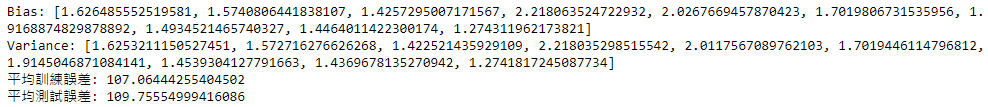
Age（年齡）

一張含有 文字, 螢幕擷取畫面, 圖表, 繪圖 的圖片

自動產生的描述

**模型選擇**

LinearRegression + K-fold交叉驗證



平均偏差介於1.27到2.22之間：

表示模型在不同訓練集上的預測結果與實際值之間的平均差異。

平均變異介於1.27到2.22之間 ：

表示模型在不同訓練集上的預測結果的平均波動性。

平均訓練誤差: 107.06。

平均測試誤差: 109.76。

**模型評估**

損失函數（Loss function）



均方誤差（MSE）為 105.69057015686639：

表示模型的預測值與實際值的平方差的平均值較大。

平均絕對誤差（MAE）為 8.080208581793617：

表示模型的預測值與實際值的絕對差的平均值較小。

以殘差圖來分析預測結果

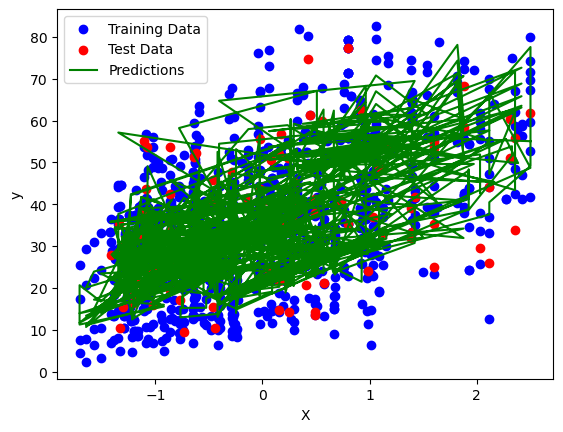
一張含有 螢幕擷取畫面, 文字, 圖表 的圖片

自動產生的描述

此殘差圖大致呈現常態分布無明顯趨勢，表示模型可能沒有存在太大系統性的偏差或誤差。

**找出最適模型**

梯度下降法



圖中藍色的點表示訓練數據，紅色的點表示測試數據，綠色的線條表示使用訓練得到的模型對訓練數據進行預測得到的結果。

繪製的圖像看起來很複雜，可能是因為數據集的特徵較多或者模型的複雜度較高。在這種情況下，可能需要進一步分析和處理數據，或者選擇更簡單的模型出出來進行回歸分析。

**最終模型評估**

分別以R-squared、MAE、MSE進行評估



R-squared的值為0.5399表示模型可以解釋目標變量的53.99%的變異量。



MSE的值為104.4507。



MAE的值為8.1157。

**模型選擇**

SVM + K-fold交叉驗證

一張含有 文字, 字型, 螢幕擷取畫面, 資訊 的圖片

自動產生的描述

平均偏差介於0.15到0.28之間：

表示模型在不同訓練集上的預測結果與實際值之間的平均差異。

平均變異介於0.15到0.21之間 ：

表示模型在不同訓練集上的預測結果的平均波動性。

平均訓練誤差: 205.68。

平均測試誤差: 207.14。

**模型評估**

損失函數（Loss function）

一張含有 文字, 字型, 白色, 行 的圖片

自動產生的描述

均方誤差（MSE）為 175.26 ：

表示模型的預測值與實際值的平方差的平均值較大。

平均絕對誤差（MAE）為 10.74：

表示模型的預測值與實際值的絕對差的平均值較小。

以殘差圖來分析預測結果

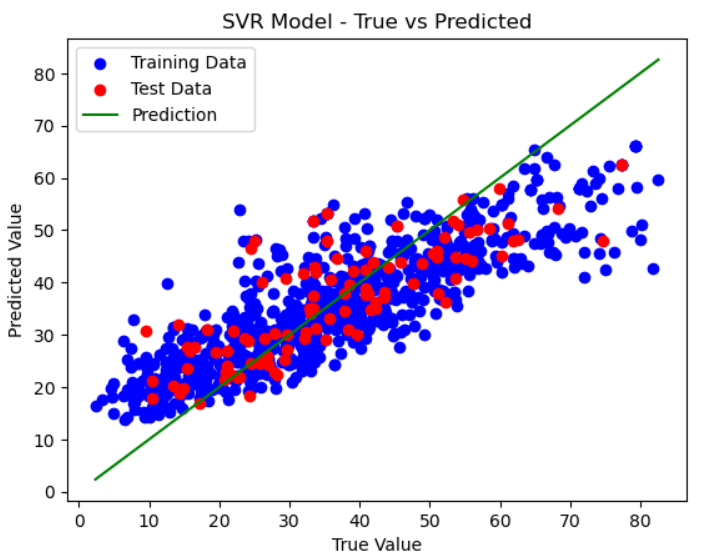
一張含有 螢幕擷取畫面, 圖表 的圖片

自動產生的描述

此殘差圖大致呈現常態分布無明顯趨勢，表示模型可能沒有存在太大系統性的偏差或誤差。

**找出最適模型**

序列最小優化(SMO)



圖中藍色的點表示訓練集中的數據點 紅色的點表示測試集中的數據點 綠色線表示理想的預測結果，即真實值等於預測值的情況。該線是一條斜率為1的直線，代表完美的預測情況。

點分散在綠色線附近並且沒有太明顯的趨勢偏差 ，表示模型的預測能力較好。

**最終模型評估**

分別以R-squared、MAE、MSE進行評估



R-squared的值為0.6503表示模型可以解釋目標變量的65.03%的變異量。



MSE的值為79.389。



MAE的值為6.903。

**最終模型選擇**

以解釋性、預測準確度和穩定性來進行兩種模型的比較

以解釋性來說：

LinearRegression作為模型的R-squared值為0.5399。

SVM作為模型的R-squared的值為0.6503。

就解釋性來說,我們選擇SVM作為模型。

以預測準確度來說：

LinearRegression作為模型的MSE和MAE分別為

"104.4507"和"8.1157"。

SVM作為模型的MSE和MAE分別為

"79.3894"和"6.9027"。

就預測準確度來說,我們選擇SVM作為模型。

以穩定性來說：

LinearRegression作為模型的變異（Variance）的值

介於1.27到2.21之間。

SVM作為模型的變異（Variance）的值

介於0.15到0.21之間。

就穩定性來說,我們選擇SVM作為模型。

因此最終模型選擇為SVM

**參考資料**

<https://www.jmp.com/zh_tw/statistics-knowledge-portal/what-is-regression/simple-linear-regression-assumptions.html>

<https://www.coursera.org/learn/machine-learning>

<https://en.wikipedia.org/wiki/Linear_regression>

<https://en.wikipedia.org/wiki/Gradient_descent>

<https://scikit-learn.org/stable/modules/linear_model.html#ordinary-least-squares>

<https://www.datacamp.com/community/tutorials/understanding-model-performance-python>

[https://machinelearningmastery.com/model-evaluation-model-selection-and-algorithm-selection-in-machine-learning](https://machinelearningmastery.com/model-evaluation-model-selection-and-algorit)