分頁 1

**AI分析嬰兒動作年齡**

作者姓名 陳柏夆 孫浩文 王淳毅

**摘要**

本研究旨在探討幼兒的動作行為，並運用資料分析技術識別影響其發展的關鍵因素。我們採用了 SRDA 幼兒發展資料庫建置計畫中的 33,474 份問卷數據，結合回歸分析與機器學習方法進行數據建模。研究結果顯示，幼兒動作表現與其年齡具有顯著相關性。此外，我們提出了數個模型，能根據幼兒對不同動作的熟悉程度預測其年齡，並針對結果提供相關建議。

**壹、研究目的與動機**

研究問題:

幼兒動作發展程度與年齡的相關性，進而透過動作靈敏度預測幼兒的年齡。

研究動機:

在當前少子化的社會背景下，每一個孩子都是家長的心肝寶貝。對於家長而言，了解孩子在不同年齡階段應該達成的動作標準，不僅是確認其是否正常發展的重要指標，更有可能揭示孩子的潛在天賦，從而為家長提供有價值的參考，協助評估孩子的早期發展潛力。

基於此，我們計劃運用數據分析技術，打造一個以幼兒動作為基礎，準確預測其發展年齡的智能模型。這不僅能幫助家長及早識別孩子的成長趨勢，還能為幼兒教育和早期干預提供科學依據，進一步提升孩子的潛能發揮與發展速度。我們相信，這項創新技術將在未來成為每位家長和教育工作者的得力助手，為幼兒的全面發展提供更精準的指導。

**貳、科學方法與工具使用**

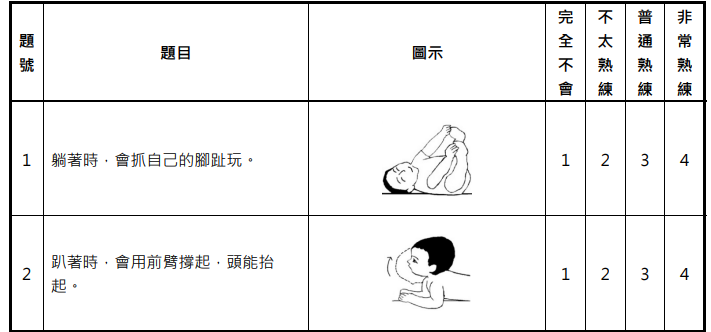
在資料處理過程中，我們主要使用了 Python 數據分析工具，如 Pandas 和 NumPy，進行資料前處理，包括資料篩選、分類與加總等操作，以確保資料品質並提升分析準確性。為了進行機器學習模型訓練，我們將數據隨機抽樣並分割為 8:2 的比例，其中 80% 作為訓練集（training set），用於模型的學習，剩餘 20% 作為測試集（testing set），用於模型評估與預測性能的驗證。

我們選取了家長問卷中的 44 項有關幼兒動作熟練度的問題作為分析對象，並應用多種機器學習模型進行建模與比較，包括隨機森林（Random Forest，n\_estimators=100）、XGBoost（n\_estimators=300）、支持向量回歸（SVR）以及類神經網絡等。通過不同模型的比較與驗證，旨在找出最具準確性與穩定性的預測模型。

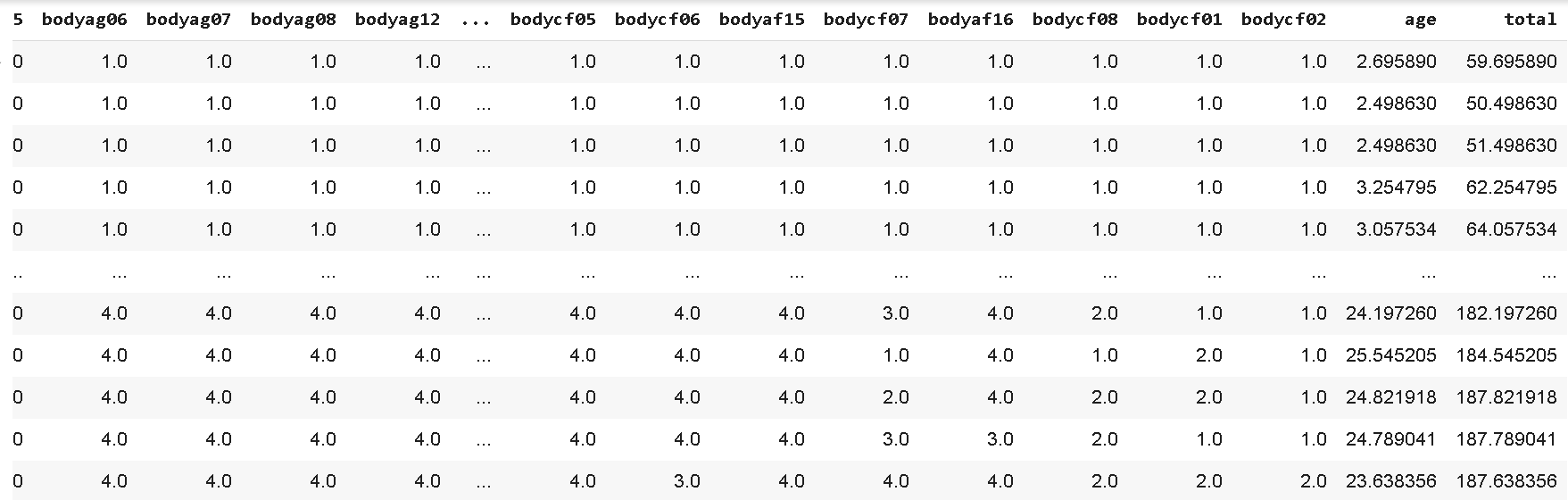
此外，我們利用 Matplotlib 將結果進行可視化展示，生成直觀的圖表，清晰呈現數據趨勢與分析結果，為後續的討論與決策提供支持。

**參、成果展現與報告**

我們使用了3-24月齡的家長面談問卷題本，此次的分析我們聚焦在身體動作發展問卷進行討論與預測，下圖為關於此問卷中與身體動作相關題目的型式



關於身體動作發展問卷共有44題，以下使用pandas匯入我們要進行資料分析的表格

****

**變數介紹:**

自變數:

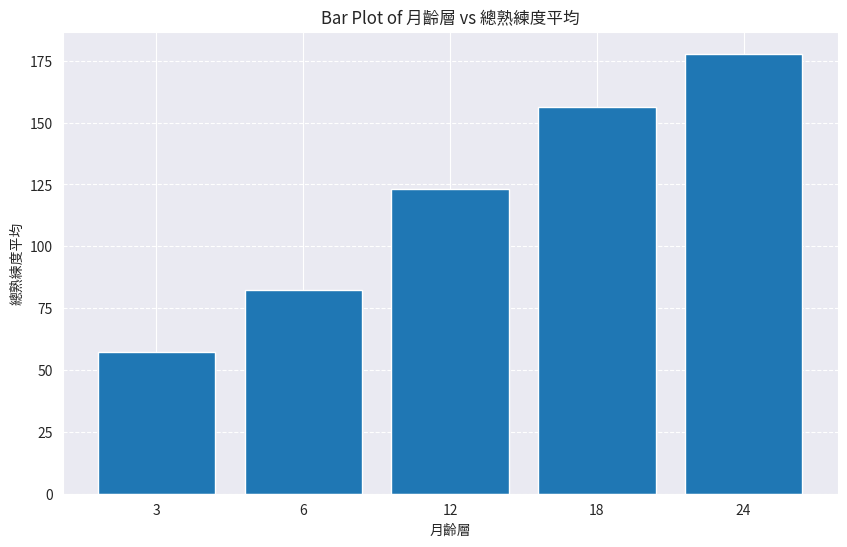
共有44個不同動作，我們拆解各動作各自成一欄，1到4為其熟練度，我們也新增了一個欄位為其所有動作熟練度的加總。

應變數:

這次的主題中，我們想要預測年齡(幼兒之月齡)。

**結果分析:**

以下為我們分析在個月齡層的總熟練度平均，根據此圖，明顯的發現月齡越大，動作熟練度總和越高。



各月齡組總熟練度的平均值

**機器學習模型:(皆以8:2的比例拆分為訓練集與測試集)**

在以下的各種模型中，我們皆分為兩個區塊，

1.分類問題，將畫分為3,6,12,18,24月齡組，共五種分類進行分析，每個月齡組

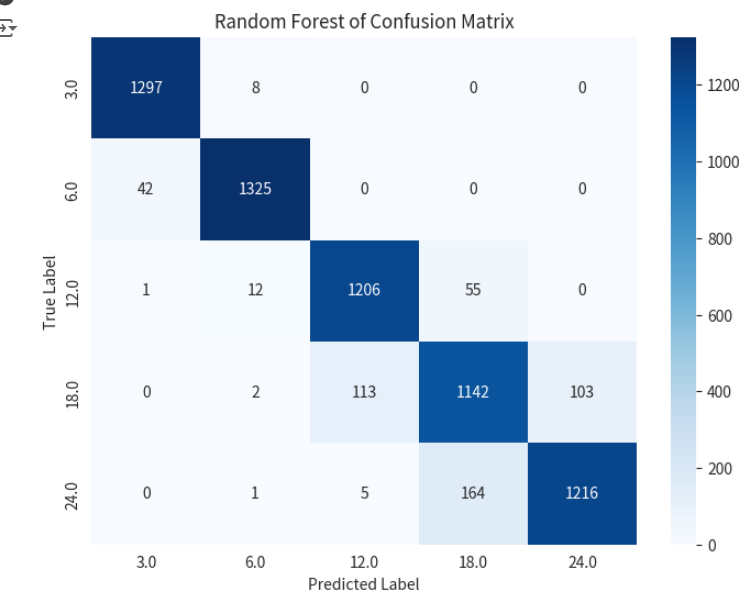
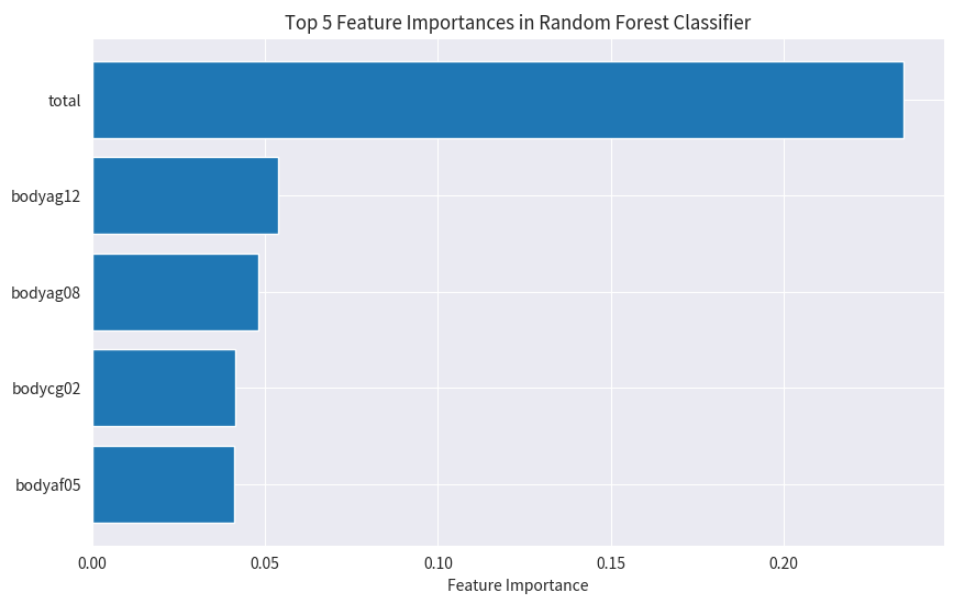
，正負差不超過一月齡

2.迴歸問題，使用每個樣本的實際年齡

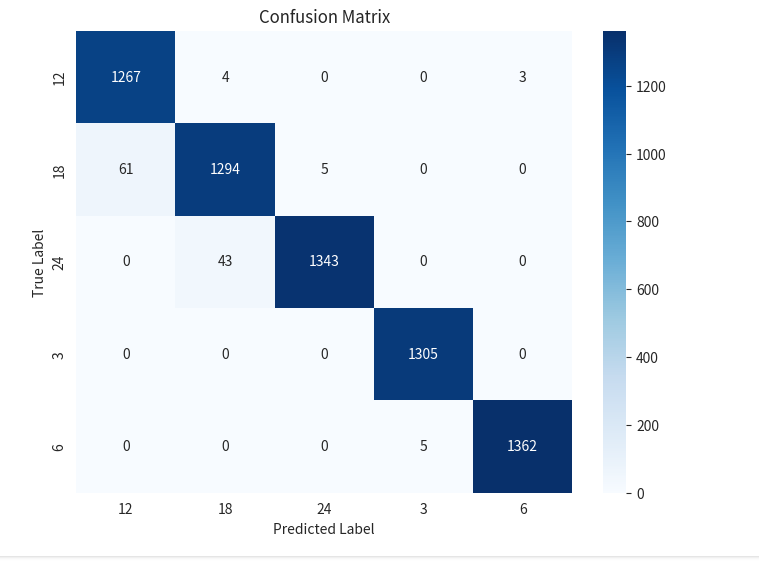
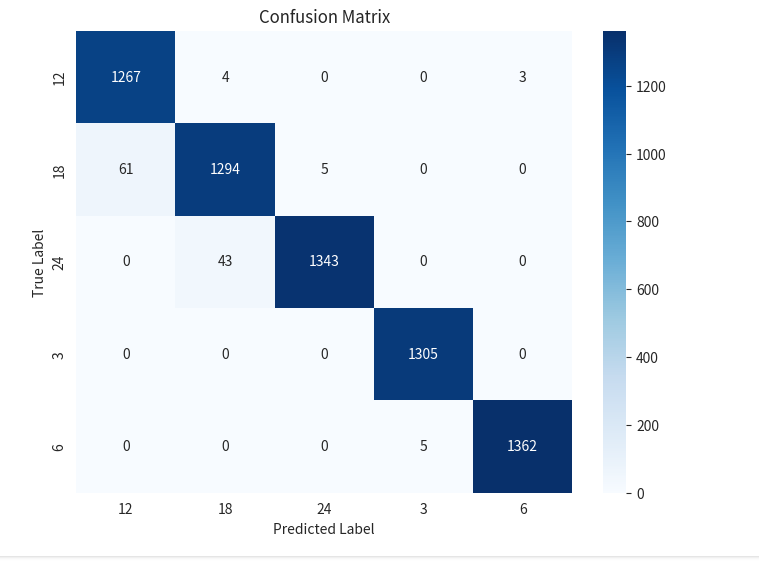
在回歸模型的準確值的估算中，我們以總樣本的最大與最小月齡相減\*0.05當作誤差值，其值接近1，以下的模型預測準確性評估我們皆以正負1月齡為誤差。

(一)分類問題

我們的模型展示結果如下:

Random Forest模型混淆矩陣 Random Forest模型五大動作重要性指標

SVM模型混淆矩陣 多分類神經網路模資料

1.Random Forest模型中達到了92%的準確率

Support Vector Machine模型中達到97%的準確率

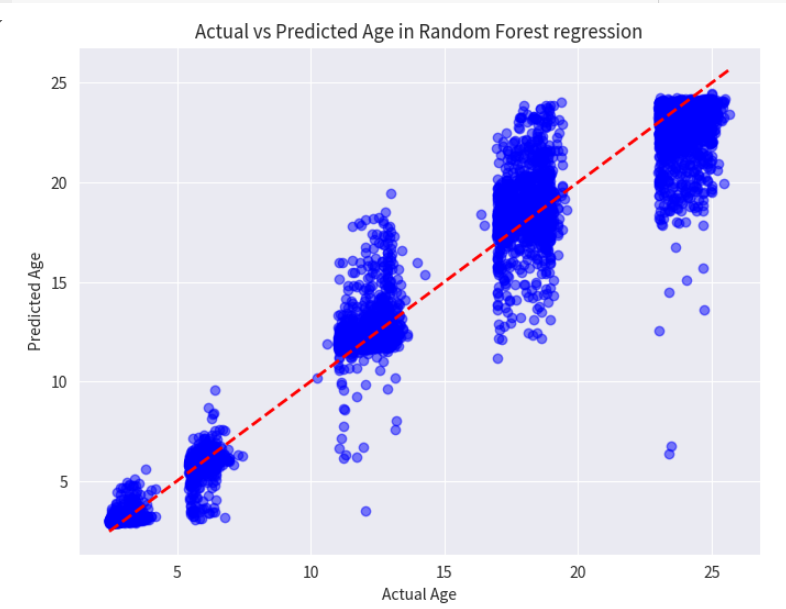
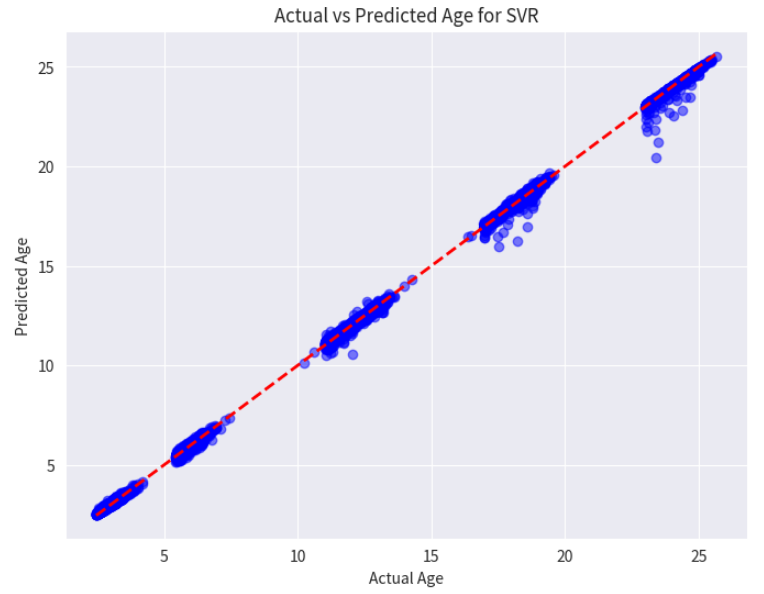
多分類神經網路模型 達到98%的準確率

相較其他模型 Random Forest的準確率較低

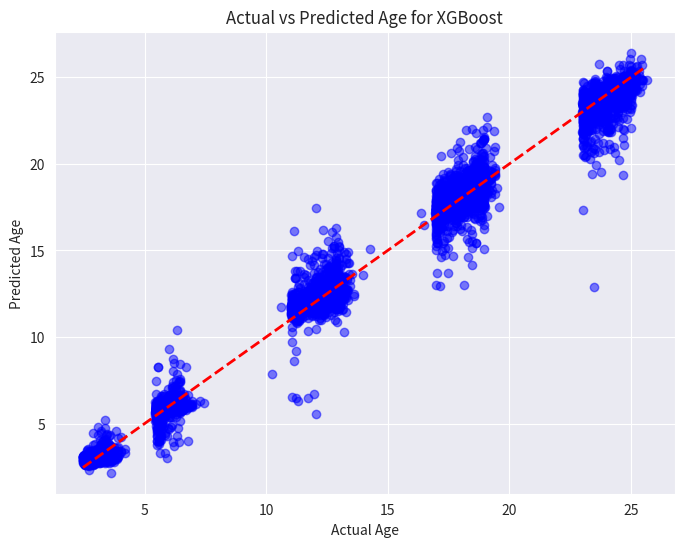
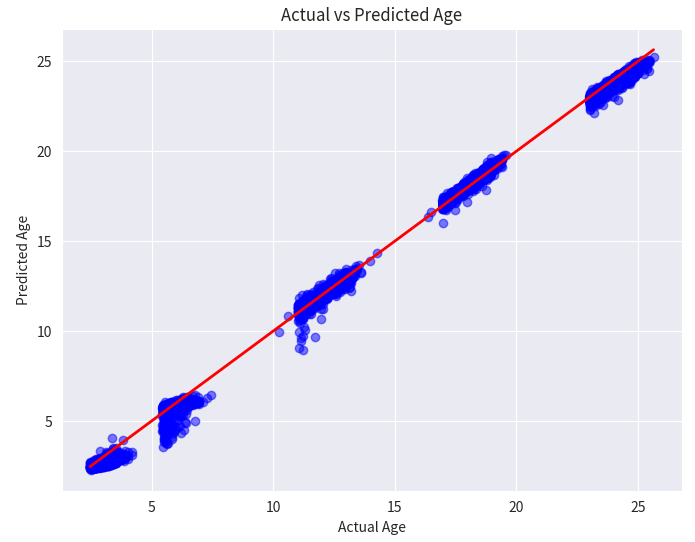
2.此Ranodm Forest模型的五大重要性指標，提供給我們得知哪些因素造成年齡的影響較大。

(2)迴歸問題

下圖為在各模型中實際值與預測值的比較，若藍點靠近紅色虛線表示此模型在預測上準確率較高。

Random Forest模型(準確率75.12%) SVM模型(準確率99.76%)



回歸神經網路模型(準確率99.09) XGBoost模型(準確率89.67%)

Random Forest 依舊在回歸模型上表現較其他模型差，有許多離群值的產生。

**肆、推論與討論**

我們的研究結果顯示，年齡與動作表現之間存在高度相關性。無論採用回歸分析或將問題轉化為分類模型，我們的模型均展現出卓越的效能。其中，SVM 模型的準確率高達 99.76%，顯示出極強的預測能力。因此，若以預測嬰兒年齡為主要目標，我們的模型無疑是十分成功的，可為家長提供量身訂製的兒童肢體發展建議。

然而，我們也觀察到，目前模型僅使用了身體動作相關的變數，這可能是模型達到高準確率的原因之一。因此，在未來的研究中，我們計畫引入更多多樣化的變數，以提升模型的泛用性並拓展其應用範圍，並能知曉在各個領域下，孩子們的天賦所在。我們期待，隨著變數範疇的增加，模型依然能維持優異的準確率，進一步證明其穩健性與實用價值。

**伍、創新性與延展性**

目前，我們的模型主要基於台灣地區的數據，因此其應用範圍暫時限於台灣人群。然而，隨著未來能夠獲取更多來自不同地區和多元人群的資料，我們有信心此模型將具備跨區域的適用性，從而服務於更廣泛的國際市場。

此外，目前模型僅針對幼兒的身體動作進行分析。在未來，我們計畫擴展數據範疇，納入認知發展、語言能力等多維度的發展指標，以使預測結果更為全面、多元。同時，這將顯著提升模型在不同發育階段與文化背景中的準確性和可靠性。我們相信，隨著這些拓展，該模型將成為更強大、靈活的工具，為幼兒發展的科學研究和實務應用提供更豐富的價值。

**陸、參考文獻**

1.張鑑如（2024）。幼兒發展調查資料庫建置計畫：3 月齡組第一波 3 月齡（D00180）【原始數據】。取自中央研究院人文社會科學研究中心調查研究專題中心學術調查研究資料庫。https://doi.org/10.6141/TW-SRDA-D00180-2

2.張鑑如（2024）。幼兒發展調查資料庫建置計畫：3 月齡組第二波 6 月齡 （D00181）【原始數據】。取自中央研究院人文社會科學研究中心調查研究專題中 心學術調查研究資料庫。<https://doi.org/10.6141/TW-SRDA-D00181-2>

3.張鑑如（2024）。幼兒發展調查資料庫建置計畫：3 月齡組第三波 12 月齡 （D00214）【原始數據】。取自中央研究院人文社會科學研究中心調查研究專題中 心學術調查研究資料庫。https://doi.org/10.6141/TW-SRDA-D00214-3

4.張鑑如（2023）。幼兒發展調查資料庫建置計畫：3 月齡組第四波 18 月齡 （D00215）【原始數據】。取自中央研究院人文社會科學研究中心調查研究專題中 心學術調查研究資料庫。 https://doi.org/10.6141/TW-SRDA-D00215-2

5.張鑑如（2024）。幼兒發展調查資料庫建置計畫：3 月齡組第五波 24 月齡 （D00228）【原始數據】。取自中央研究院人文社會科學研究中心調查研究專題中 心學術調查研究資料庫。https://doi.org/10.6141/TW-SRDA-D00228-2