

AI分析嬰兒動作年齡

作者姓名 陳柏峯 孫浩文 王淳毅

資料庫來源：SRDA 幼兒發展資料庫建置計畫(家長面談問卷)

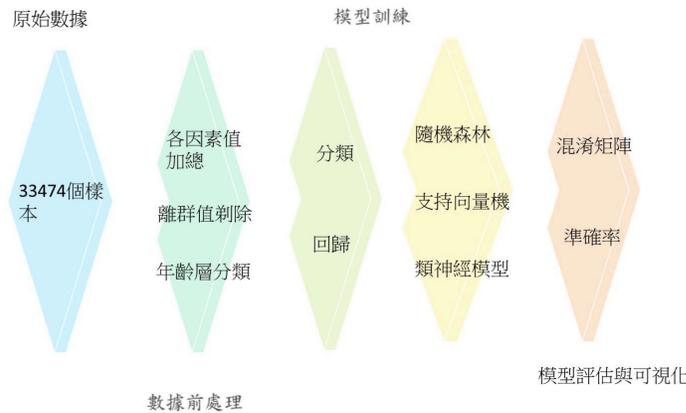
簡介

為了提供家長有辦法透過，協助評估孩子的早期發展潛力。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。

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。



成果展現

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，圖一為關於此問卷中與身體動作相關題目的型式。

題號	題目	圖示	完全不	不太	普通	非常
1	躺著時，會抓自己的腳趾玩。		1	2	3	4
2	趴著時，會用前臂撐起，頭能抬起。		1	2	3	4

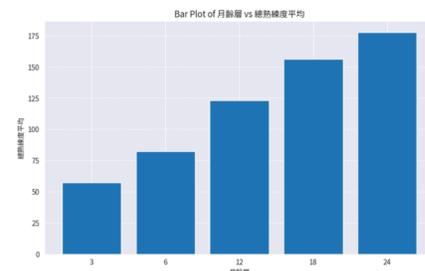
圖一、題目型式

關於身體動作發展問卷共有44題，圖二使用pandas匯入我們要進行資料分析的表格。

```
5 bodyag06 bodyag07 bodyag08 bodyag12 ... bodycf05 bodycf06 bodycf15 bodycf07 bodycf16 bodycf08 bodycf01 bodycf02 Age total
0 1.0 1.0 1.0 1.0 ... 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 2.896890 59.695890
0 1.0 1.0 1.0 1.0 ... 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 2.496030 50.496030
0 1.0 1.0 1.0 1.0 ... 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 2.496030 51.496030
0 1.0 1.0 1.0 1.0 ... 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 3.254795 62.254795
0 1.0 1.0 1.0 1.0 ... 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 3.057534 64.057534
... ..
0 4.0 4.0 4.0 4.0 ... 4.0 4.0 4.0 3.0 4.0 2.0 1.0 1.0 24.197260 182.197260
0 4.0 4.0 4.0 4.0 ... 4.0 4.0 4.0 1.0 4.0 1.0 2.0 1.0 25.545205 184.545205
0 4.0 4.0 4.0 4.0 ... 4.0 4.0 4.0 2.0 4.0 2.0 2.0 1.0 24.821918 187.821918
0 4.0 4.0 4.0 4.0 ... 4.0 4.0 4.0 3.0 3.0 2.0 1.0 1.0 24.789041 187.789041
0 4.0 4.0 4.0 4.0 ... 4.0 3.0 4.0 4.0 4.0 2.0 2.0 2.0 23.636356 187.636356
```

圖二、dataframe表格

圖三為各月齡層對於熟練度的平均值，方便我們簡單看出兩者的關聯性。



圖三、月齡層與熟練度關聯

分類模型

畫分為3,6,12,18,24月齡組，共五種分類進行分析，結果如圖四，每個月齡組，正負差不超過一月齡。

迴歸模型

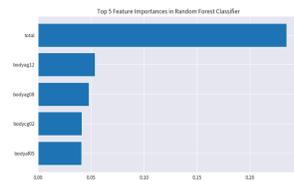
圖五為在各模型中實際值與預測值的比較，若藍點靠近紅色虛線表示此模型在預測上準確率較高。

分類與迴歸模型整體比較

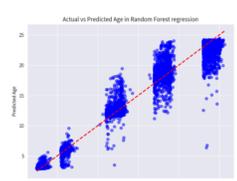
圖六為各模型下的訓練準確率比較，我們發現在分類問題上，每個模型皆達到超過九成的準確率。



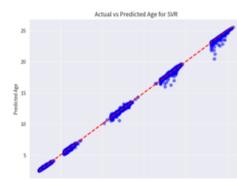
隨機森林混淆矩陣



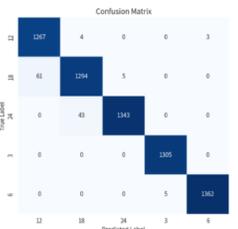
五大重要指標



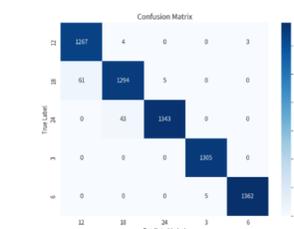
隨機森林



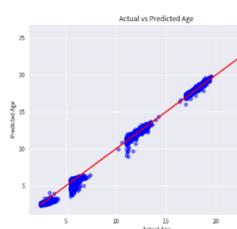
SVM



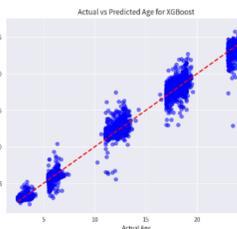
SVM 混淆矩陣



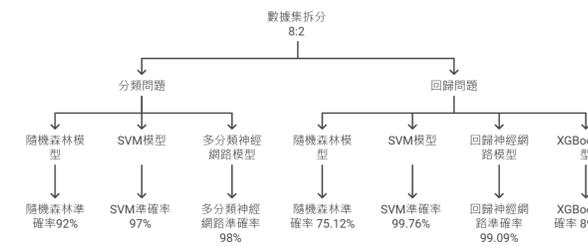
類神經網路混淆矩陣



神經網路



XGBoost



圖六、各模型準確率比較

圖四、三種模型混淆矩陣預測效果比較

圖五、四種迴歸模型的預測效果

結論

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。

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。