

使用機器學習演算法於賽馬資料分析之研究

系所／資訊工程學系

指導老師／黃世育

組員／黃博鴻、徐兆威、蔣宗儒、徐仲寬

世界上賽馬已經慢慢的變成我們平常的娛樂消遣，而一場賽馬比賽的結果，牽扯到許多因素，如馬匹、練馬師以及騎師都會深深影響結果，然而隨著科技的進步，賽馬比賽所產生的可預測性大幅的上升，進而機器學習(Machine Learning)的出現，但鮮少有將兩者結合的案例。

本專案的題目即是將機器學習與賽馬產生做結合，經由大量資料訓練模組，利用機器學習幫助我們自動產生出我們所需求的賽馬之排名預測。

公式 1 顯示 Python 中的機器學習(Scikit Learn)中的 R2 模型 (R Square_score Model) 公式。本專案透過將賽馬歷屆結果帶入此 R2 模型，進而分析出數據的成長方向而達到預測的目的。輸出的兩個數據分別稱為「R2」與「P value」。通過使用 R2 模型 (R Square_score Model) 公式，得到「R2」與「P value」並將數據匯成表格。圖 1、圖 2 為「R2」與「P value」所形成的表

格，經由將誤差極大狀況排除後呈現出模組訓練過後的預測精準度。實驗過程中常有數據不準或是空數據出現；難用於預測的狀況，但經由多次調整及修改，並藉由大量的預測數據使模組的輸出能更加理想。

一般來說，R2 模型分成兩種數據，也就是「R2」與「P value」。我們的研究與實驗將會圍繞在這兩個數據維持穩定上，針對賽馬排名預測的生成，並混合一些其他的演算法，以期待能在賽馬的機器學習領域得到不同的啟發與經驗。

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2}{\sum_{i=0}^{n-1} (y_i - \bar{y})^2}$$

公式 1：R2 模型公式

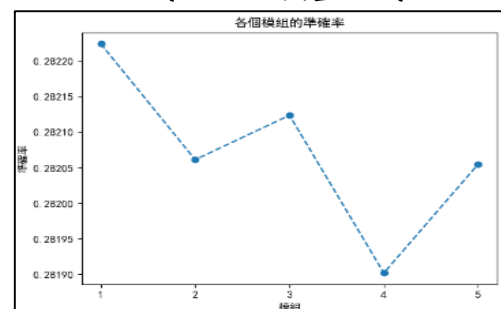


圖 1：R2 的實驗結果

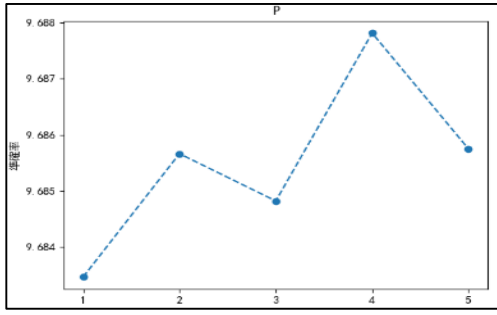


圖 2：P value 的實驗結果