



圖靈觀眾

預測高中學生之 文理學業表現

成員：統計108 劉臣洋 | 物理112 李國成 | 物理112 陳昱丞

報告總覽

目錄

1. 動機
2. 資料簡介
3. 資料來源
4. 面臨的挑戰
5. 解決方案
6. 預測結果的評估
7. 時程規劃

1. 預測學生成績

透過收集學生的相關數據來預測該學生的文理考試成績

2. 嘗試瞭解環境對學生成績的影響

在所有收集的數據中，到底有哪些因素對學生的成績會有顯著的影響？



動機

資料簡介

- **Input (X)**

葡萄牙語成績及數學成績的兩個資料集

每個資料及有30個變數

資料集筆數

葡萄牙語：649 筆資料

數學：395 筆資料

- **Output (Y)**

成績的總和 (三次考試各20分)、

以60%成績為基準並判斷是否會過

資料簡介

- 資料形態

類別資料：是否有參加課外活動、是否談過戀愛、父母是否離婚等

數值資料：年齡、課後自由時間等

- 資料內容

個人相關：年齡、性別、課後自由時間等

家庭相關：父母教育程度、職業，家人數量等

KAGGLE

Student Performance Data Set

Predict student performance in
secondary education (high school)

網址：

<https://www.kaggle.com/impapan/student-performance-data-set>

數據來源

兩間葡萄牙高中
收集來的**成績和問卷**

資料來源

1. 變數很多，需要處理高維度資料

2. 需要處理類別形態資料

面臨的挑戰

3. 資料筆數不多，總和只有1000多筆資料

4. 有些特殊的資料
(像：先天環境不好但成績優秀的學生)

解題方案

1. 資料探勘與前處理
2. 透過PCA來降低資料的維度
3. 透過比較不同的模型，找到最合適的分析模型

評估預測結果

Accuracy

MSE

交叉驗證

時間規劃

1週

資料探勘與前處理

0.5週

處理及調整模型

0.5週

整合所有分析結果，
並導出結論