

# Data Mining

## Final Project

### 個人報告

題目範圍:

從以下三個 Kaggle 線上競賽中挑選一個資料集

<https://www.kaggle.com/competitions/child-mind-institute-problematic-internet-use/overview>

<https://www.kaggle.com/competitions/nlp-getting-started/overview>

<https://www.kaggle.com/competitions/playground-series-s4e11/data>

結合資料前處理、探勘、模組程式，對該資料集提供一個完整探勘分析的程式。

了解該方法的原理，改變不同參數的設定或前處理方式，進行效能比較。

將自己實作程式預測結果上傳競賽系統並顯示Leaderboard結果。

### 小組報告

小組同學選用同一組資料集，了解小組其他同學進行的結果，提供一份小組報告，對個別結果提供一個整合報告，包括整體結果的統整觀察及結論。

若小組想採用其他資料集，請將欲分析資料集說明及分析目的於12/6前先寄給老師。

### Project demo & 繳交小組report (40%):

簡要說明系統實作目的及方式，demo實作程式及結果(每組約15 min)

各組除了在指定時間demo給老師看，每位同學相互觀摩其他至少4個組別的 project demo，並給定看過組別的優點及可改進建議，投給其中一組為推薦組別(只能投給其他組)。獲得前三高票的組別可獲得老師提供的獎品。

評分依據: 小組內容了解的熟悉度，合作程度

Source code (10%) 在報告中提供雲端下載連結

### Personal report (包含以下項目) (50%)

- 1) 專題實作資料集及探勘目的說明
- 2) 採用data mining方法
- 3) 程式/環境設定，執行方式說明
- 4) 改變控制參數/技術說明(須說明為何想改變控制的想法)
- 5) 評估方法(例如Accuracy, Error Rate, Precision, Recall, F-measure,執行時間等)
- 6) 結果及討論