

Day 17 特徵工程

特徵工程簡介



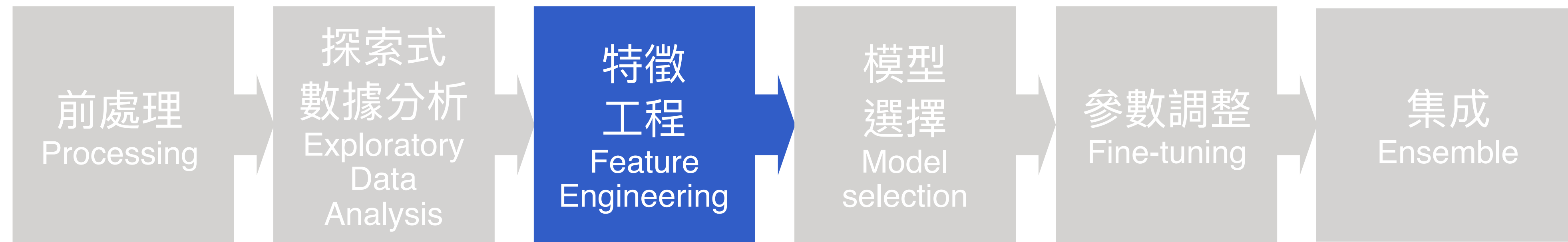
出題教練

陳明佑

知識地圖 特徵工程 特徵工程簡介

機器學習概論 Introduction of Machine Learning

監督式學習 Supervised Learning



非監督式學習 Unsupervised Learning



特徵工程 Feature Engineering

概論

數值型特徵

類別型特徵

時間型特徵

填補缺值

去離群值

去偏態

特徵縮放

類別型特徵處理

時間型特徵處理

特徵組合

特徵篩選

特徵評估



本日知識點目標

- 初步理解特徵工程的概念
- 能從程式中辨識特徵工程的區塊與意義
- 知道特徵工程至少需要那些部分

什麼是特徵工程? (1 / 2)

通常我們會這樣考慮事情...

	重要性	緊急度	所需時間	
Question1	5 ★	1 ★	5 ★	→ 11 ★
Question2	2 ★	5 ★	1 ★	→ 8 ★
Question3	3 ★	4 ★	3 ★	→ 10 ★

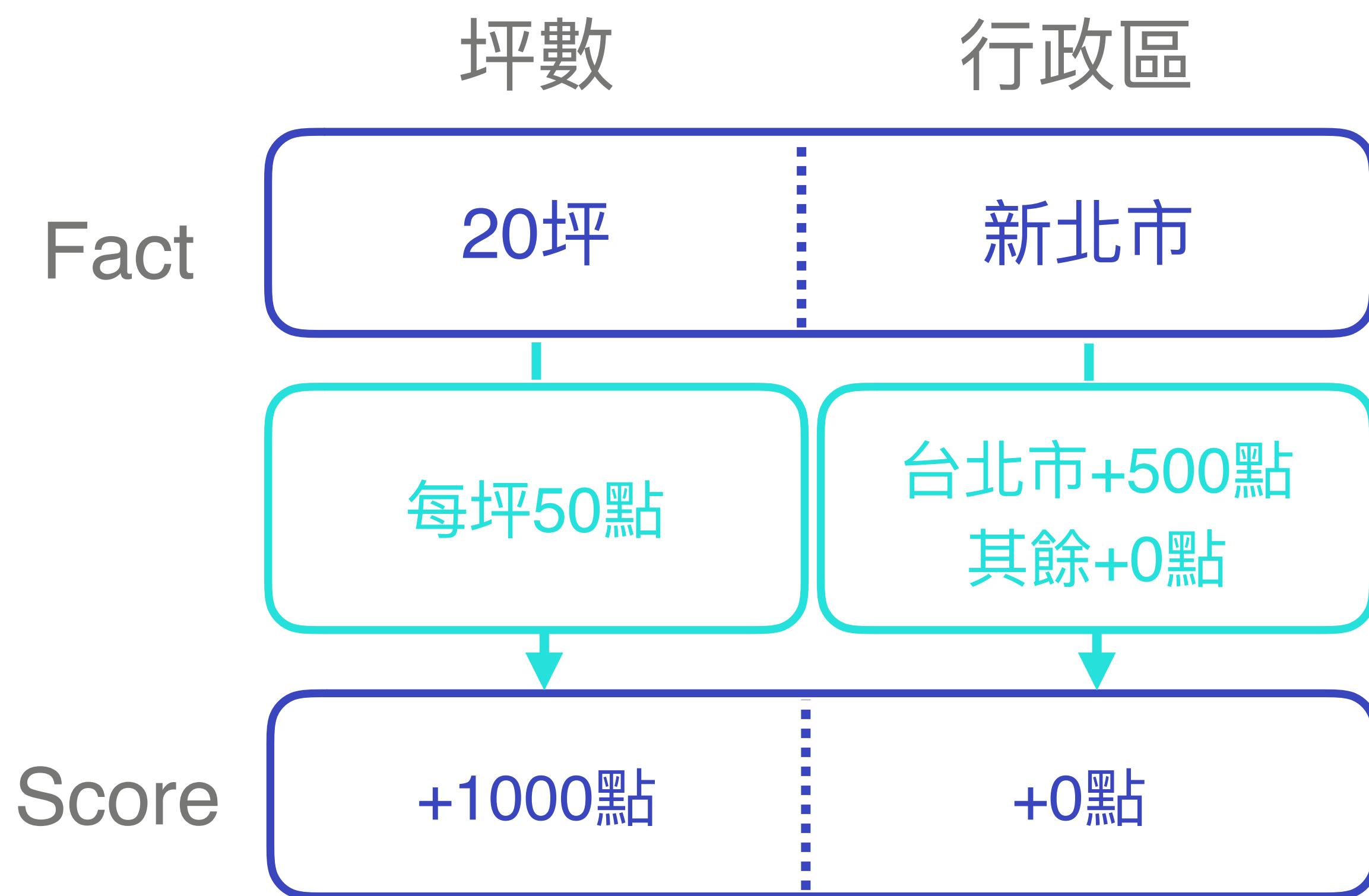
什麼是特徵工程? (2 / 2)

但其實考慮的關鍵是...

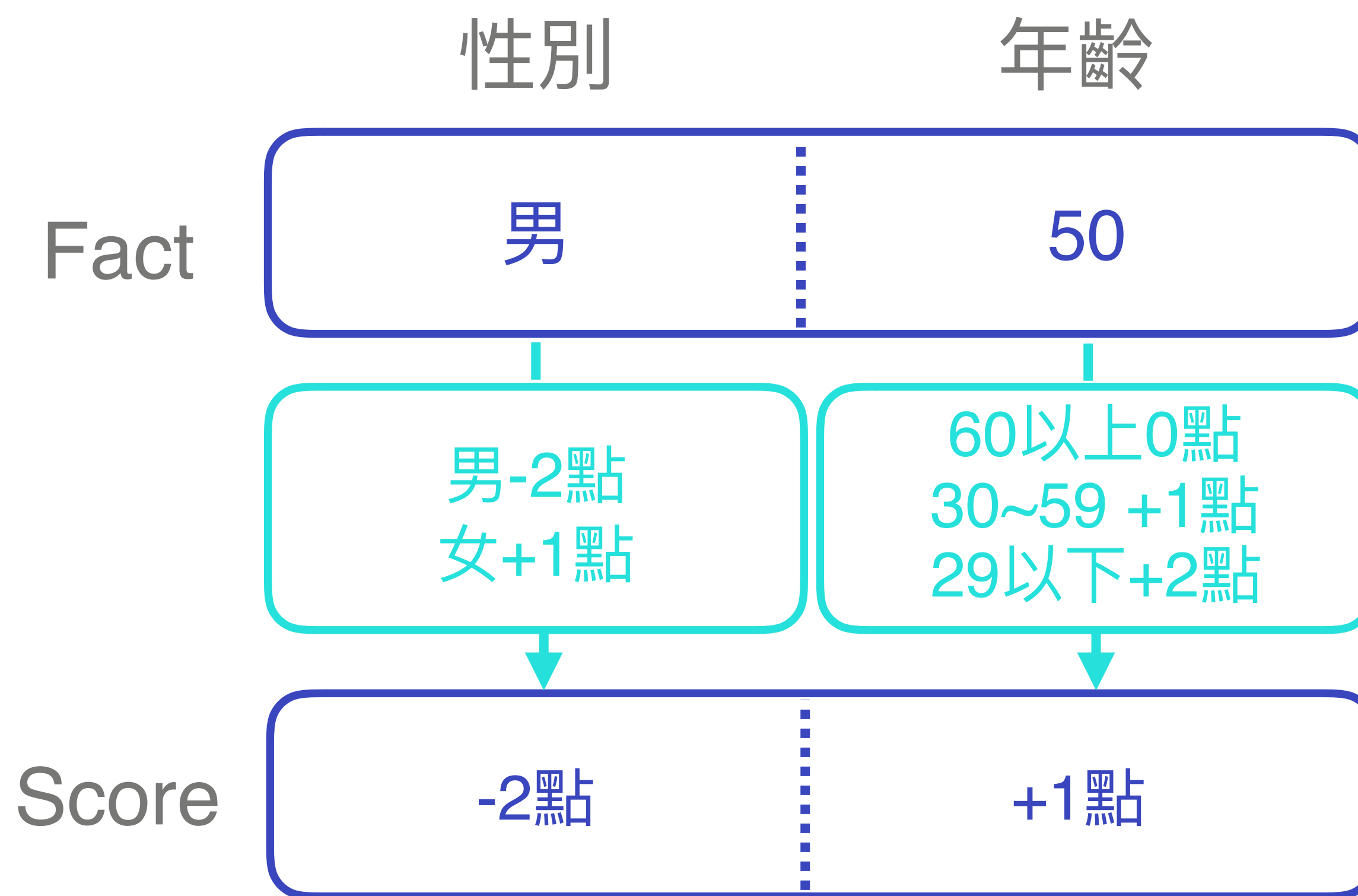


從事實到對應分數的轉換，我們稱為特徵工程

房價預測



鐵達尼號生存預測

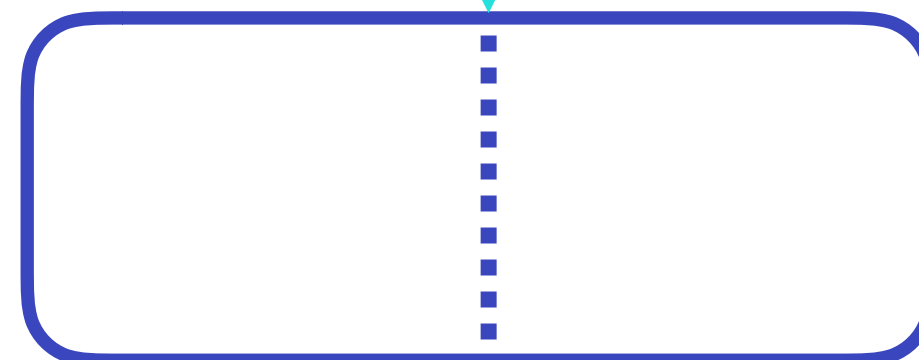
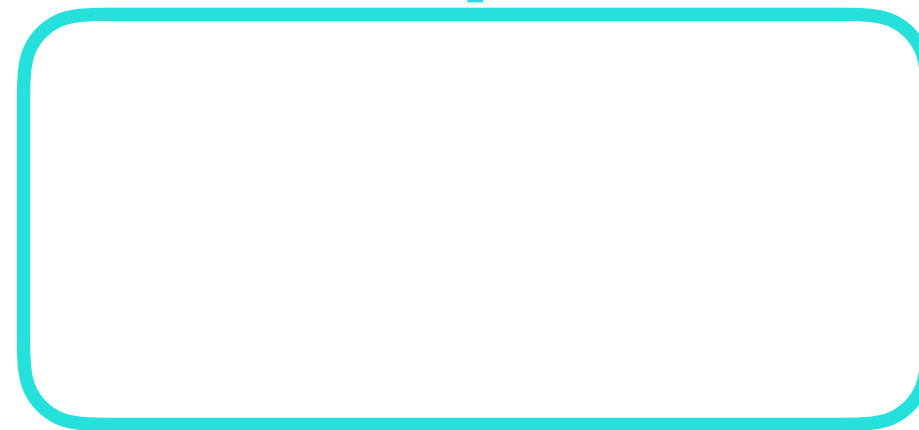
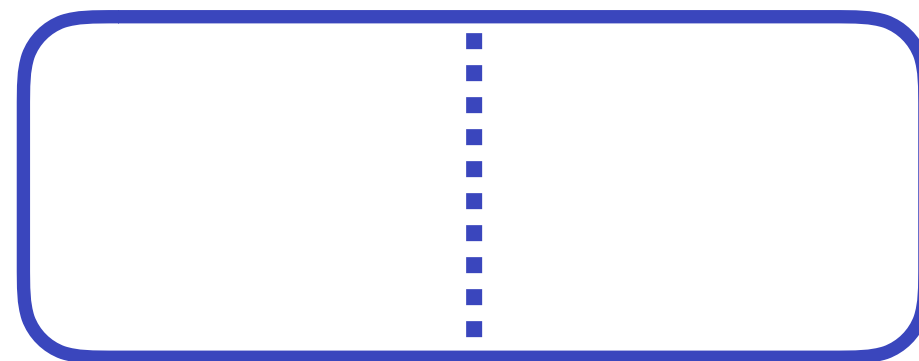


點數未必直接對應到總價或機率，但會是後續評估的依據

前述流程 / python程式 對照(範例：房價預測)

前述流程

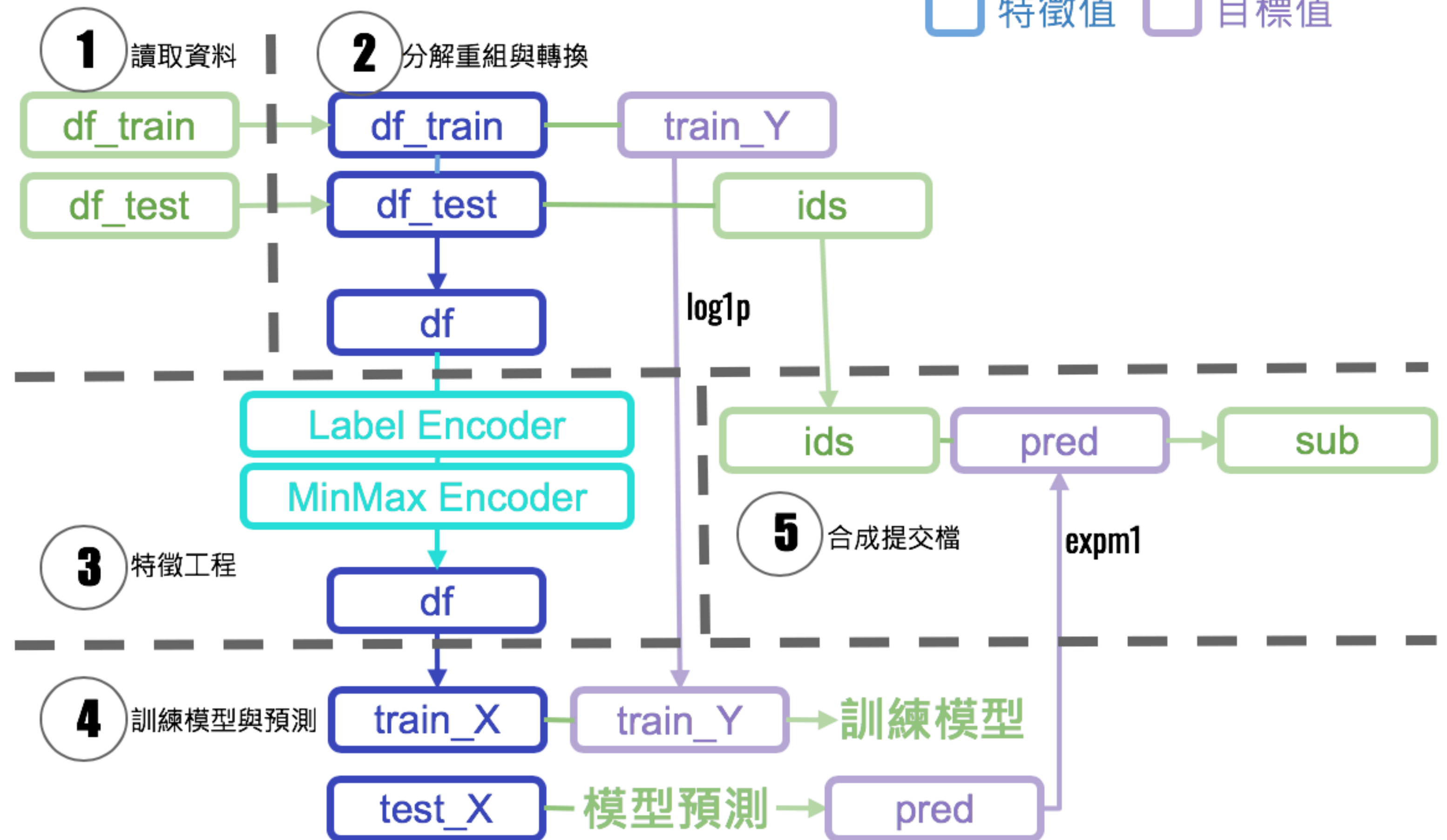
Fact

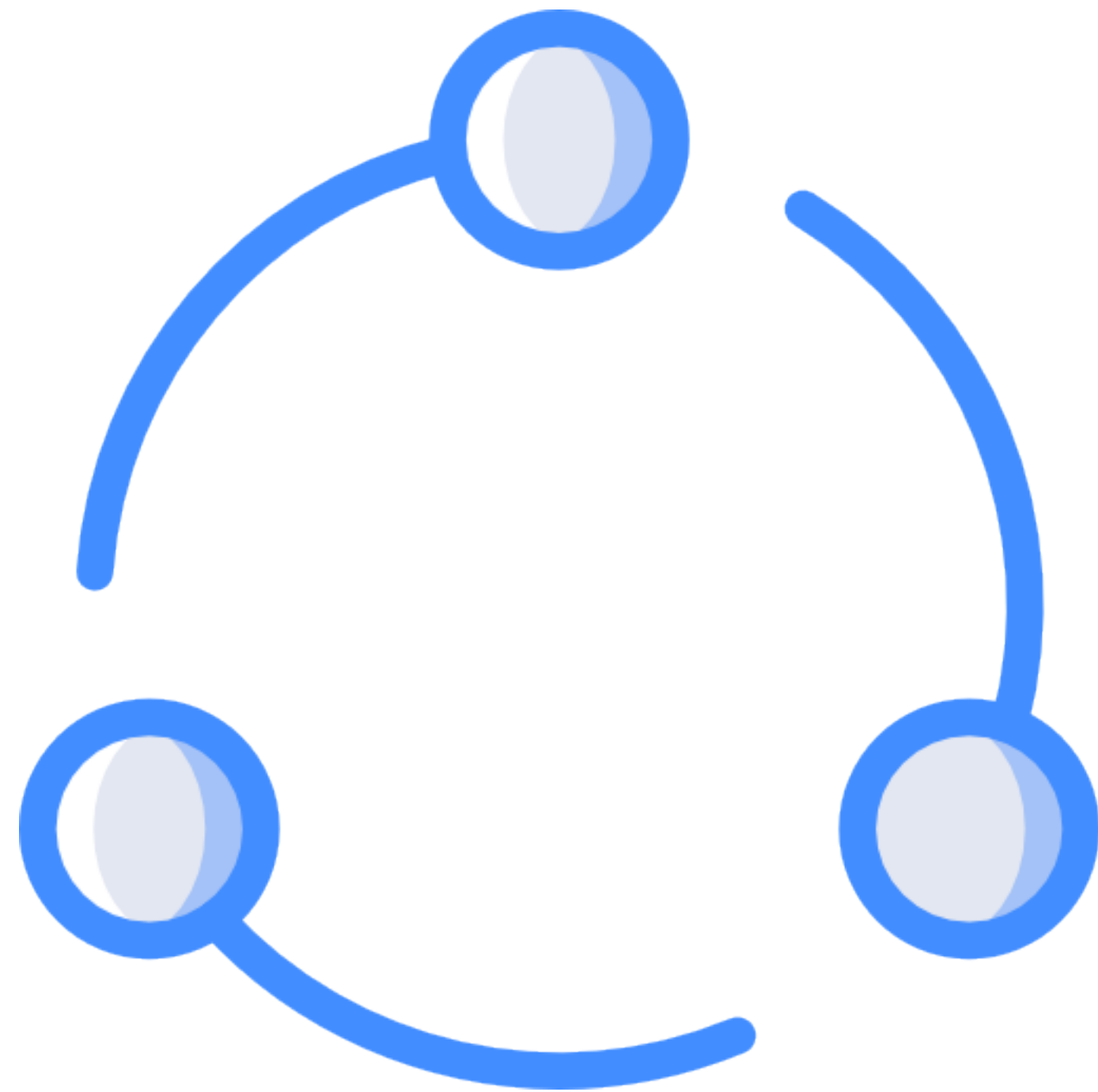


Score

python程式 (請參閱今日範例)

 特徵值 目標值





- 特徵工程是**事實對應**到後續評估分數的**轉換**
- 在程式語法中，特徵工程位於**資料彙整之後**，以及**擬合模型之前**
- 由於資料包含類別型(文字型)特徵以及數值型特徵，所以最小的特徵工程至少要包含一種**類別編碼**(範例使用標籤編碼)，以及一種**特徵縮放**方法(範例使用最小最大化)

解題時間 It's Your Turn

請跳出PDF至官網Sample Code & 作業
開始解題

