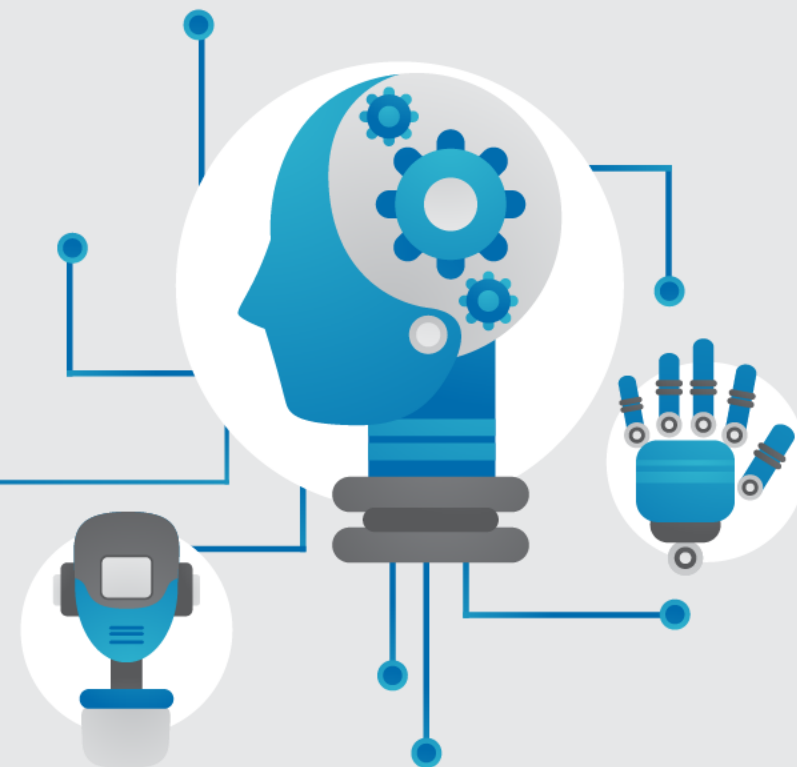


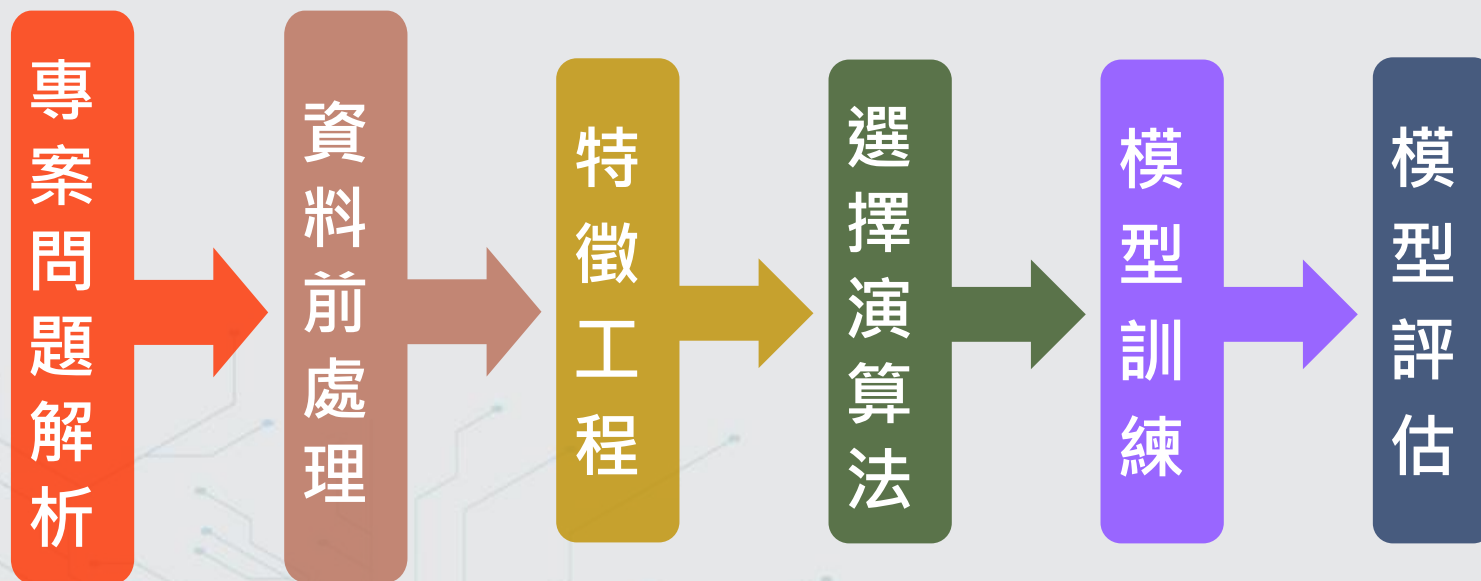
# 機器學習的建構準則





# 機器學習專案建構流程

機器學習實務





# 機器學習專案建構流程

機器學習實務



專案問題解析

資料前處理

特徵工程

選擇演算法

模型訓練

模型評估



# 專案問題解析



## 確定問題目標



例如：利用保險公司的客戶交易資料，開發機器學習系統達到「提升業績」的目標。



## 可行性評估

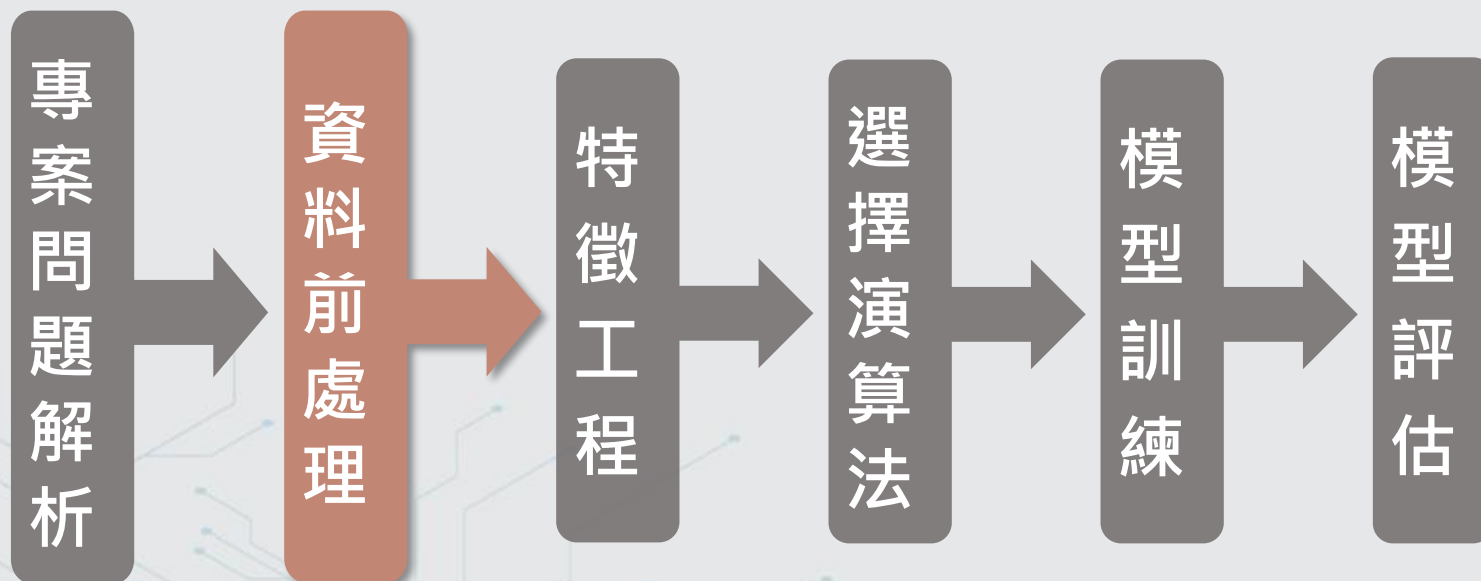
考量資料量多寡、軟硬體設備、投入人力資源等，評估專案開發的可行性。





# 機器學習專案建構流程

機器學習實務



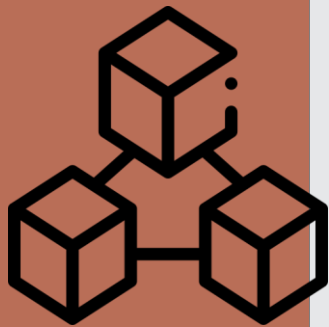


# 資料前處理



## 資料結構化

將半結構資料和非結構化資料都轉換成結構化資料。



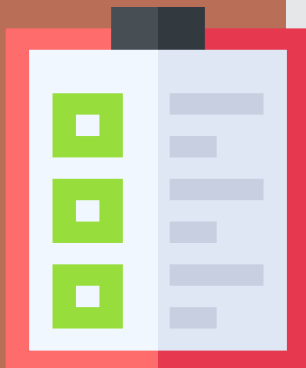
## 資料過濾

將資料中跟結果不相干的特徵資料過濾掉。



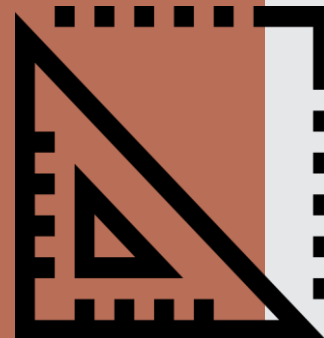
## 資料補齊

將資料表中空缺的欄位進行填補。



## 資料正規化

將資料處理限定於一定範圍內，用於加速演算法的收斂





# 資料結構



## 結構化資料

以矩陣結構儲存在資料庫的資料，可以透過二維表格結構來顯示。



## 半結構化資料

按照一定的結構儲存，但不是二維的表格結構。



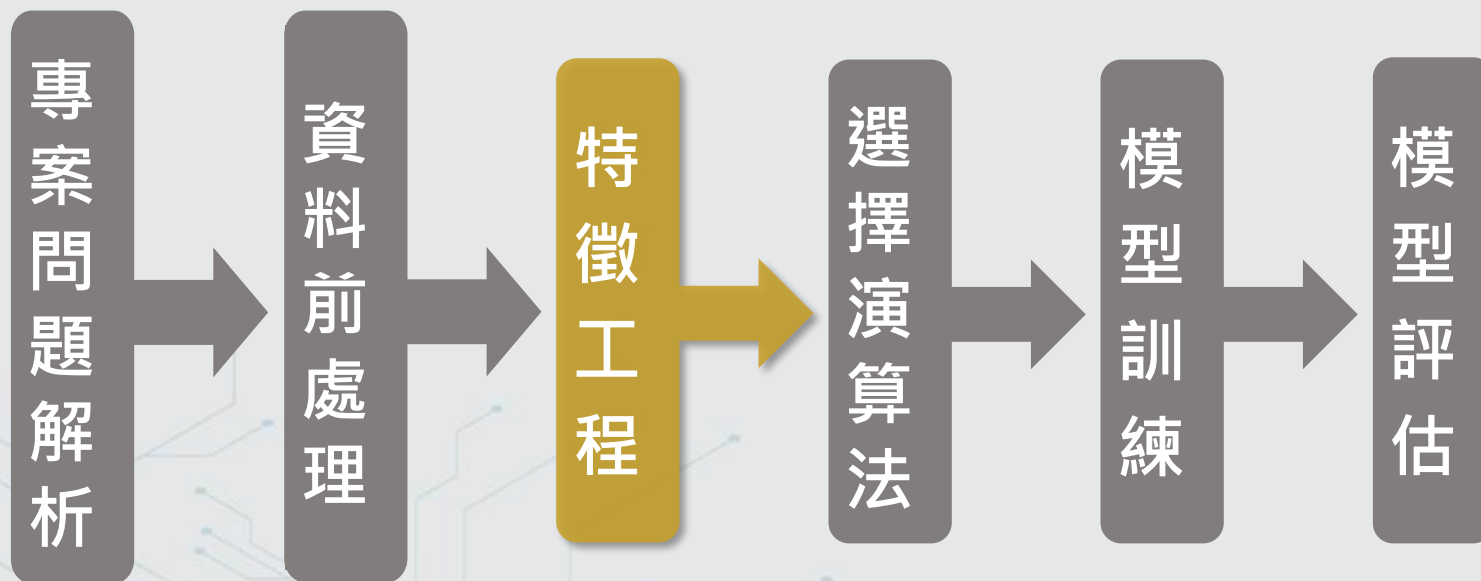
## 非結構化資料

未經整理過的資料。例如：圖片、文字、語音等。



# 機器學習專案建構流程

機器學習實務







# 特徵工程



## 特徵向量化

將特徵**量化**，例如ID、性別、區域、職業等。

## 特徵重要性評估

每個特徵對於目標列的**影響程度**。例如可透過邏輯迴歸的模型係數進行判斷。

## 特徵衍生

利用現有的特徵進行某種組合，**產生具有新含意的特徵**。例如利用產品的銷售記錄和銷售時間，可以組合算出產品熱度。

## 特徵維度縮減

處理高維度數據時，**挖掘關鍵欄位**，用以**減少輸入維度**。

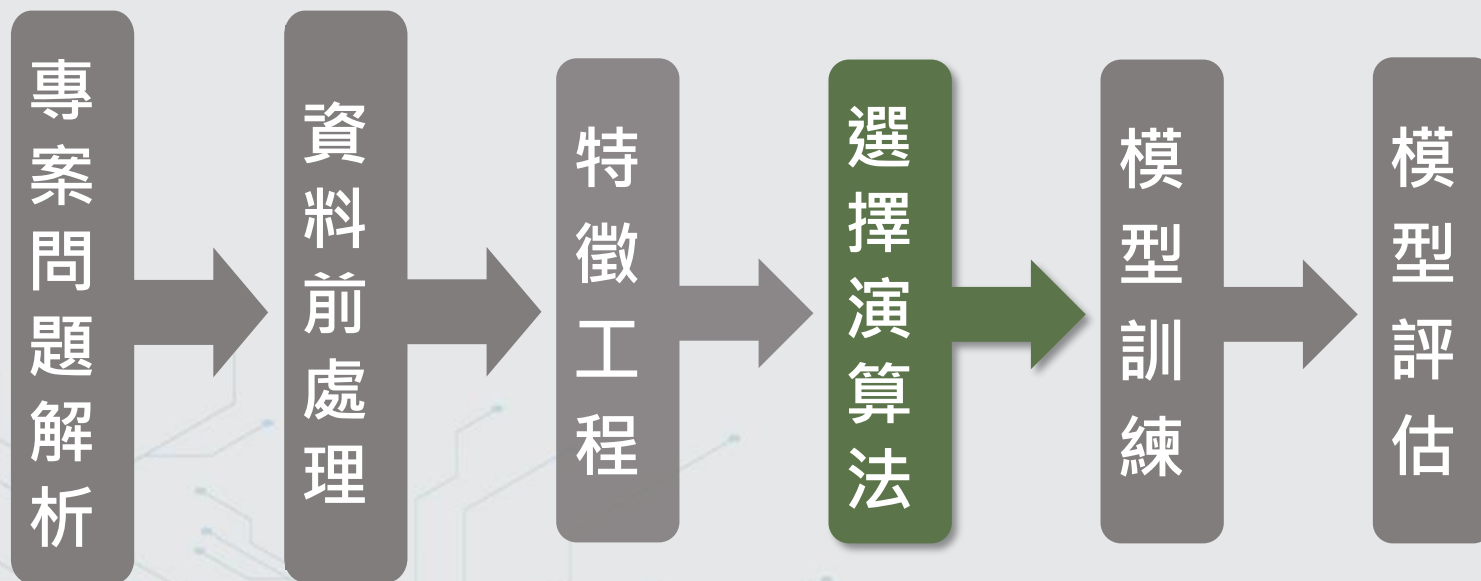
目的在：

- ( 1 ) 確保特徵間的相互獨立
- ( 2 ) 減少計算量
- ( 3 ) 去除雜訊



# 機器學習專案建構流程

機器學習實務





# 演算法選擇

機器學習實務



## 常規演算法

例如：  
分類演算法  
分群演算法  
迴歸演算法  
文字處理演算法

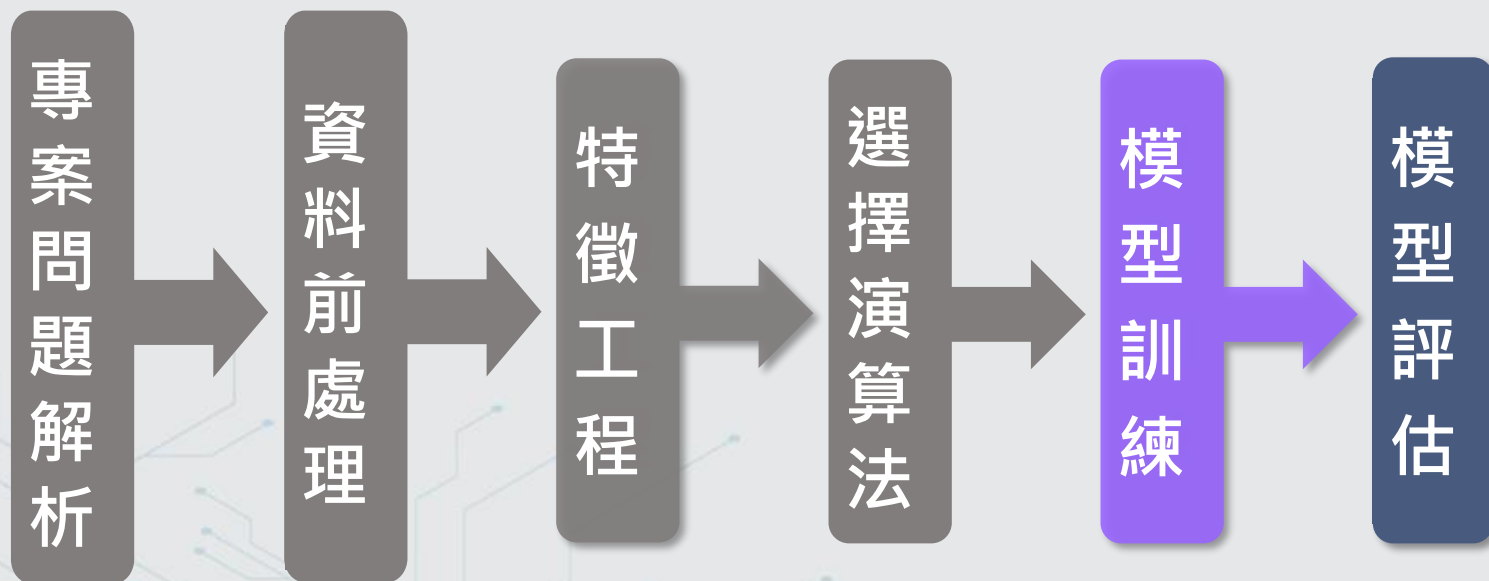
## 深度學習演算法

例如：  
深度神經網路  
卷積神經網路  
遞歸神經網路



# 機器學習專案建構流程

機器學習實務





# 模型訓練與評估



## 模型參數設定

輸入大小、模型結構、學習率和迭代次數等

## 模型參數設定

尋求效率更好的模型

## 模型參數設定

評估指標包含：  
正確率（Accuracy）  
精確率（Precision）  
召回率（Recall）  
F值（F-measure）等。