Day 4 資料清理數據前處理 EDA資料類型介紹

知識地圖 機器學習前處理 欄位的資料類型介紹及處理

機器學習概論 Introduction of Machine Learning

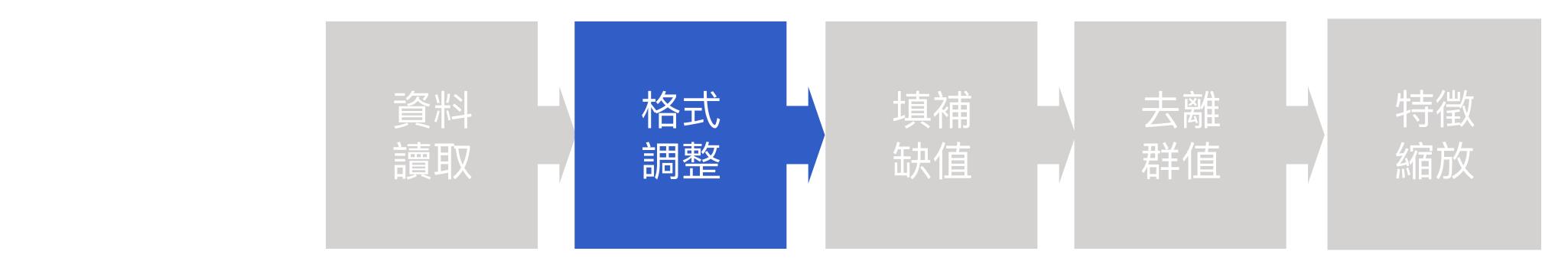
監督式學習 Supervised Learning

探索式 特徵 模型 數據分析 集成 參數調整 前處理 選擇 工程 Exploratory Ensemble Processing Fine-tuning Model Feature Data selection Engineering Analysis

非監督式學習 Unsupervised Learning

> 分群 Clustering 降維 Dimension Reduction





資料類型

資料的欄位變數一般可分

- 離散變數: 只能用整數單位計算的變數
 - 房子的房間數量、性別、國家
- 連續變數: 在一定區間內可以任意取值的變數
 - Ex: 測量的身高、飛機起飛到降落所花費的時間、車速

當然還有日期、boolean 等等不同的格式,實務中遇到再 google 就好



01

Pandas DataFrame中 最常見的欄位資料類型有三種

02 float64: 浮點數,可表示離散或連續變數

● int64 : 整數,可表示離散或連續變數

object : 包含字串,用於表示類別型變數



請跳出PDF至官網Sample Code&作業開始解題

