

Day 16 探索式數據分析

模型初體驗 Logistic Regression



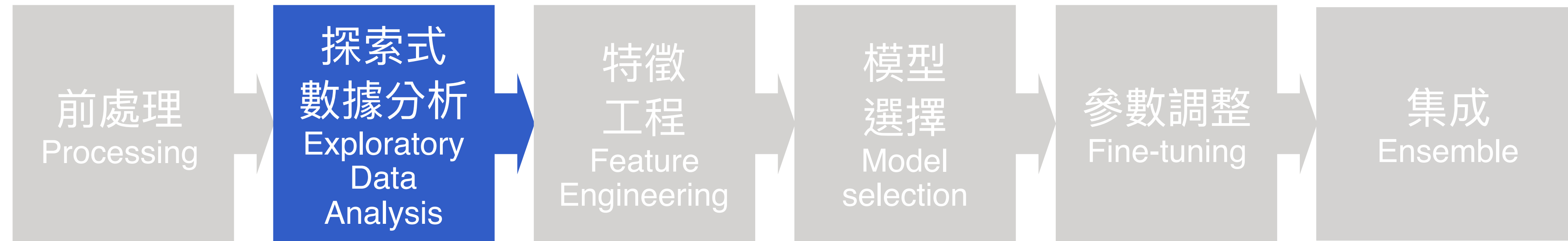
出題教練

游為翔 / 杜靖愷

知識地圖 探索式數據分析 模型初體驗

機器學習概論 Introduction of Machine Learning

監督式學習 Supervised Learning

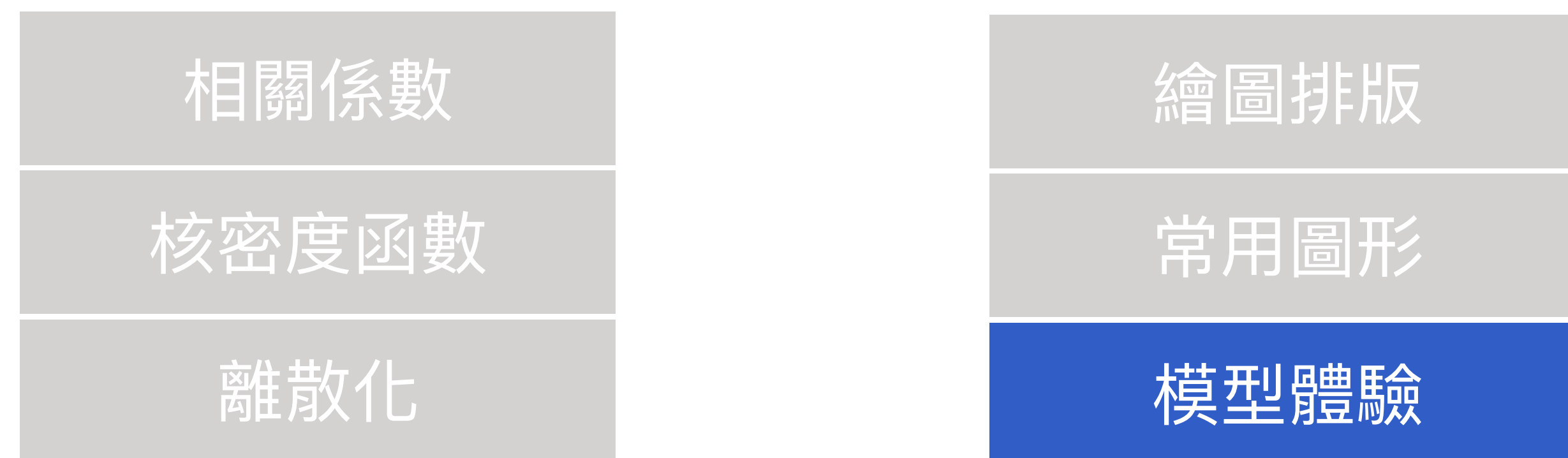


非監督式學習 Unsupervised Learning



探索式數據分析 Exploratory Data Analysis (EDA)

統計值的視覺化



本日知識點目標

以最簡單的模型體驗一次完整上傳 kaggle 結果的過程

A baseline: Logistic Regression

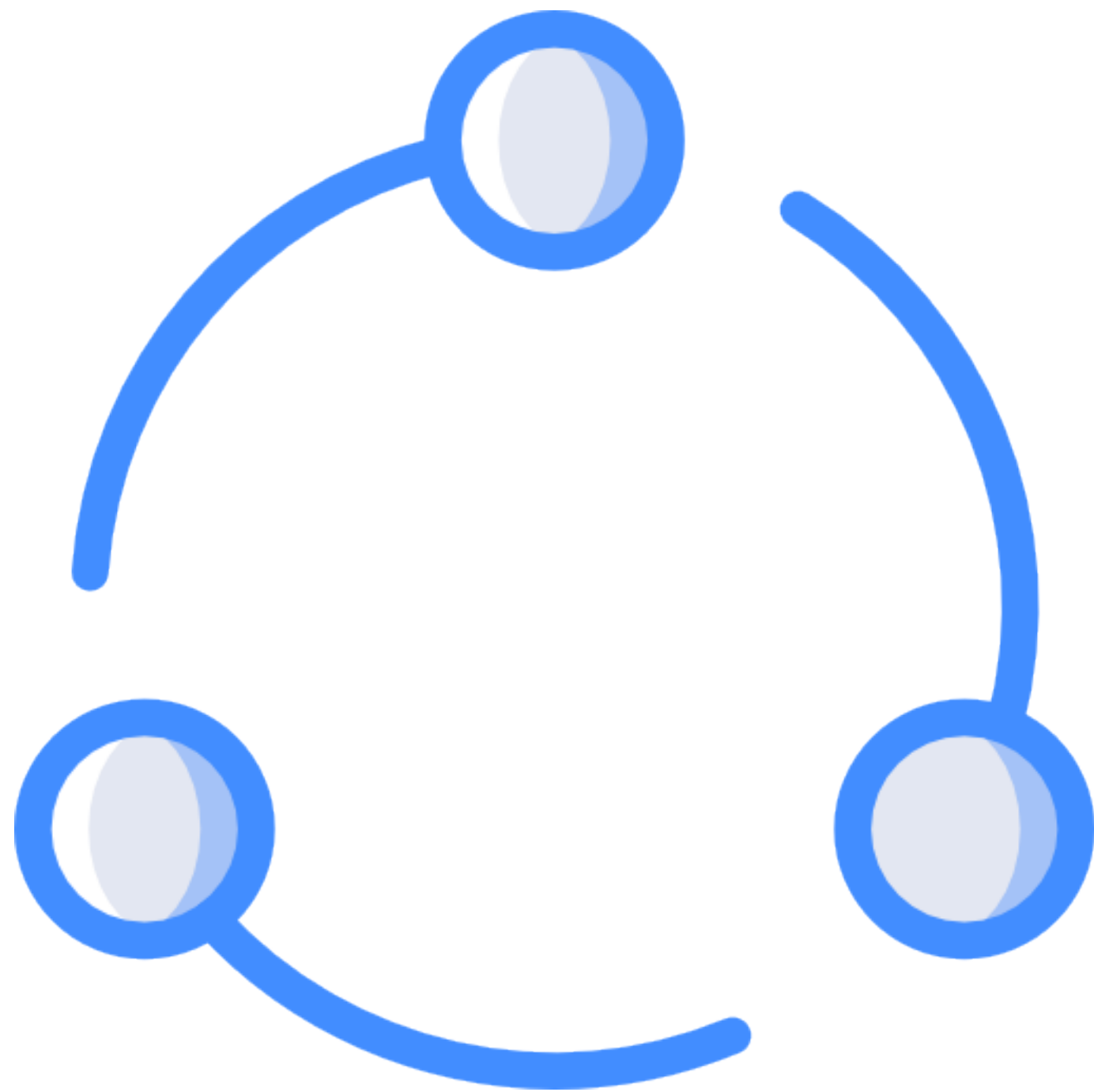
我們最終的目的是要預測客戶是否會違約遲繳貸款的機率，在我們開始使用任何複雜的模型之前，有一個最簡單的模型當作 baseline 是一個好習慣。

我們使用 Logistic Regression 作為 baseline，理論的部分可以參考

- Andrew Ng's videos about logistic regression

<https://www.youtube.com/playlist?list=PLNeKWBMsAzboR8vvhnlanxCNr2V7ITuxy>

- 大部分的情況，如何通過資料得到模型的具體計算過程的程式碼是不需要我們自己的，python 有很豐富的資源，我們只要把相關的套件 import 進來，看官方文檔或者 google 怎麼使用就好。（通常只是一個 .fit 的過程）
- 相反地，時間會花更多在讀取資料、處理資料上，甚至，第一次把模型結果儲存為對的格式可能比訓練模型還花時間。



解題時間 It's Your Turn

請跳出PDF至官網Sample Code & 作業
開始解題

