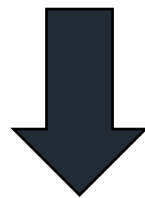


# 波士頓房價預測流程圖

1. 安裝套件，抓取資料，了解資料統計量有哪些
2. 整理 dataframe 數據集，透過相關係數探討房價與變數的相關性( $r > 0.5$ )
3. 分出訓練與驗證數據，檢查遺失值

探索數據



1. 列出房價統計量，分類連續跟非連續
2. 畫出散佈圖或盒鬚圖
3. 透過圖了解房價與各房子特徵的關係

可視化



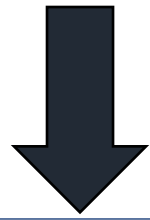
1. 將特徵進行標準化，使執行結果更準確
2. 建立回歸模型，找出閾值(平均)
3. 作特徵篩選，將係數與閾值(True: 大於閾值影響大)
4. 用圖表呈現出特徵重要性的排序

標準化



1. 前幾項重要性高的特徵，建立新的分析數據集(7 項)
2. 使用決策樹進行拆分工程

拆分數據



## 建立模型

1. 利用新數據計算  $r^2_{\text{train}}$ ,  $r^2_{\text{test}}$  及  $r^2_{\text{adj}}$
2. 建立七項變數的迴歸模型
3. 建立 XGB 模型



## 交互驗證

1. 用 k-fold 和 XGboost 的方式，將 train 數據拆分進行驗證



## 評估修正模型

1. 透過新數據計算判定係數  $R^2$ 、 $R^2_{\text{adj}}$ 、MSE 來評估修正模型
2. 最後得出用來預測的最佳化回歸模型



## 殘差圖表

1. 使用標準化殘插圖讓數據看得更準確
- 2.

