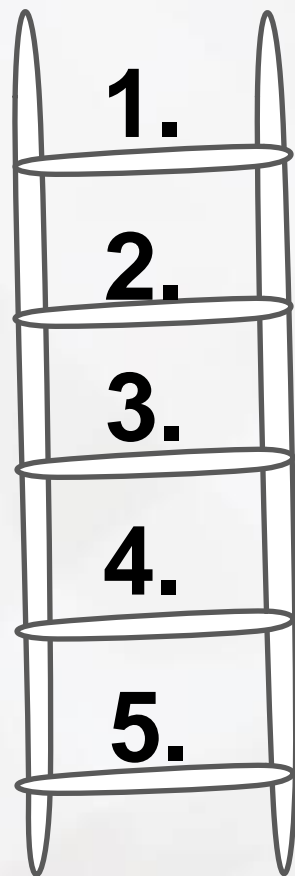
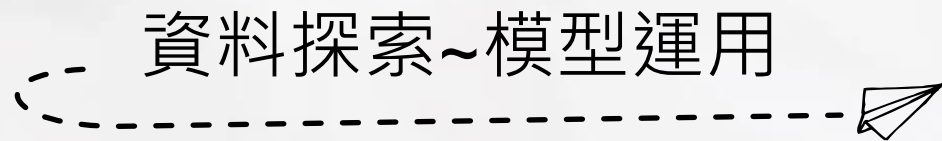


房產特徵價格評估模型

基於實價登入



資料探索~模型運用



資料獲取

使用爬蟲，爬取實價登入網站歷年數據。

資料處理

探索、清洗、特徵篩選、特徵工程。

選定模型

線性回歸、決策樹、XGBoost...

模型訓練

區分為訓練集與測試集，進入模型中訓練。

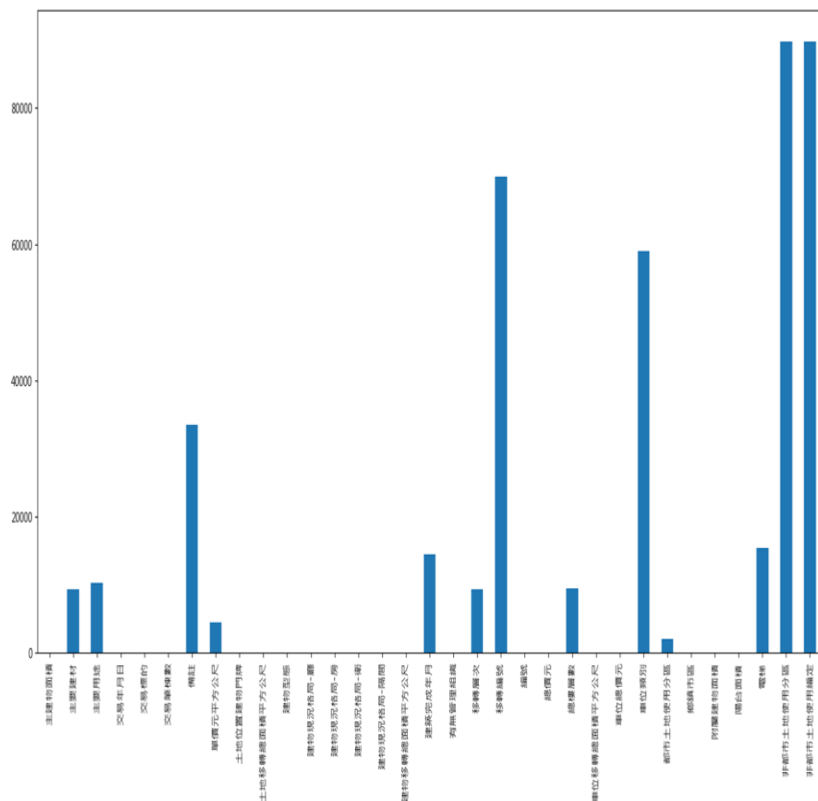
模型運用

建立一個簡單GUI介面，使用者輸入得到結果。

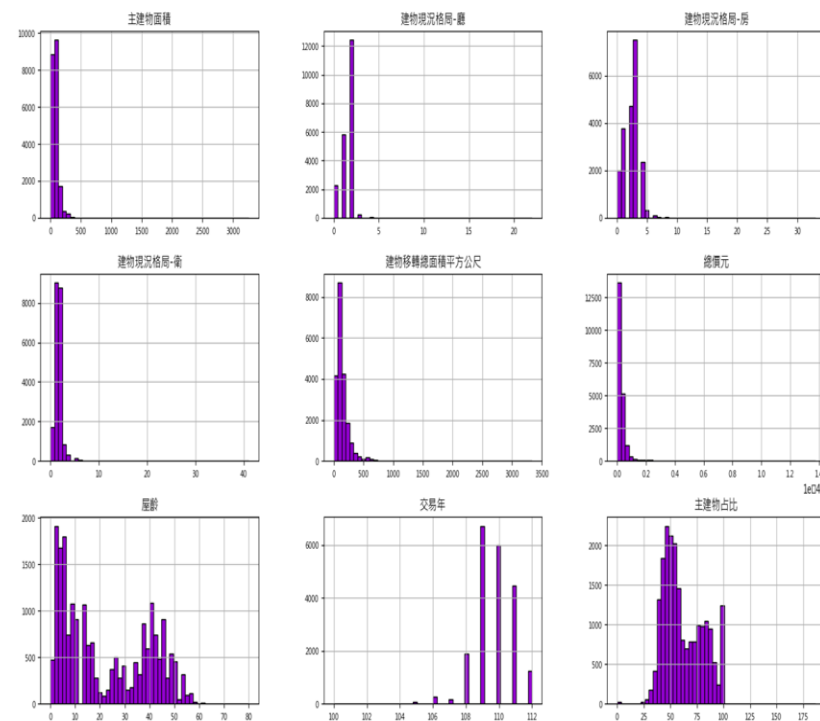
資料探索



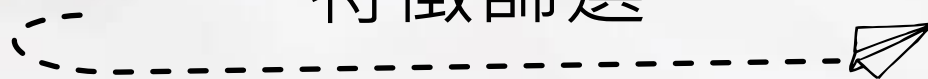
檢視缺失值



查看資料散布情況



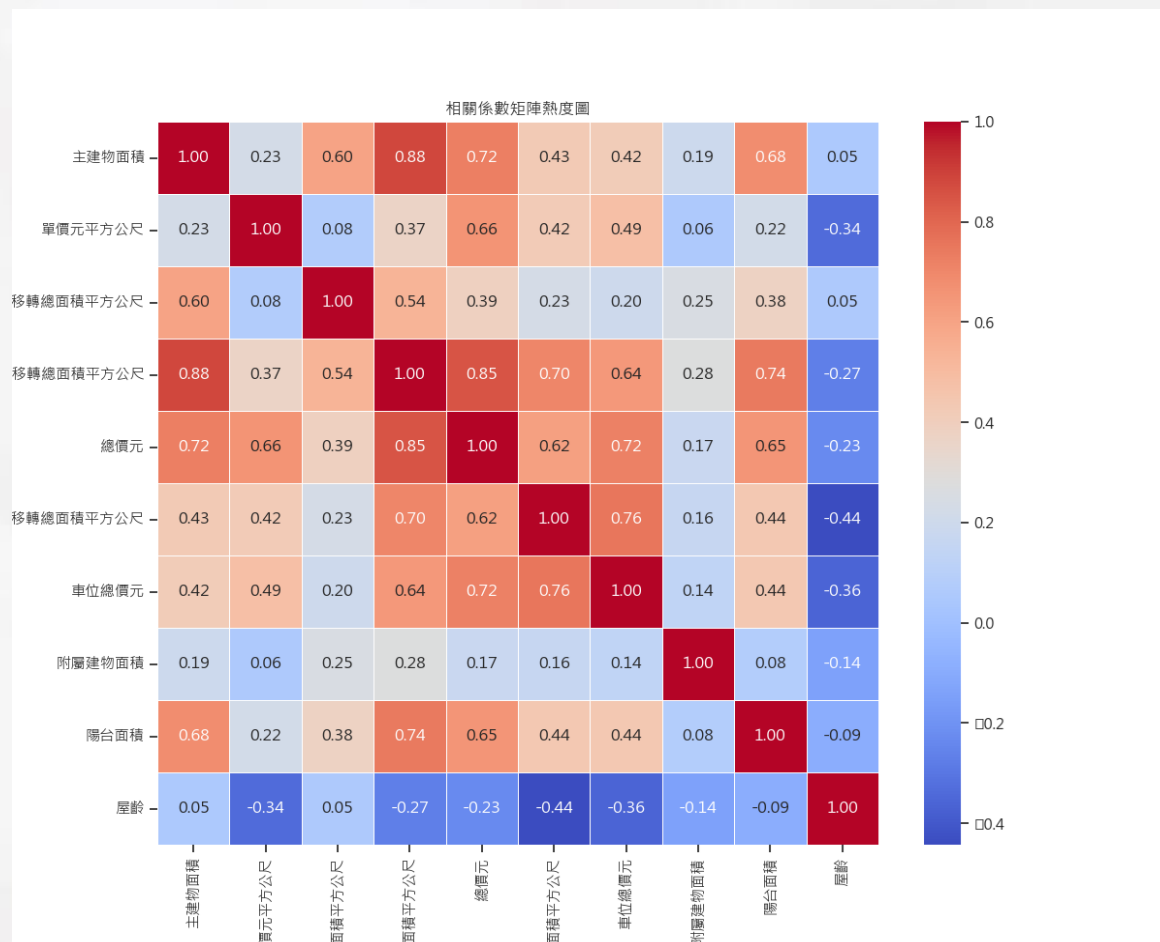
特徵篩選



★連續型特徵

連續型特徵用相關係數，篩選與房價最具關係的特徵；為了避免共線性在篩選特徵時排除兩兩相對中的其中一個。

共線性篩選



```
var2]:..2f}")
```

高度相關的變數對：

車位移轉總面積平方公尺 和 建物移轉總面積平方公尺：相關係數 = 0.70

車位總價元 和 總價元：相關係數 = 0.72

陽台面積 和 建物移轉總面積平方公尺：相關係數 = 0.74

建物移轉總面積平方公尺 和 主建物面積：相關係數 = 0.85

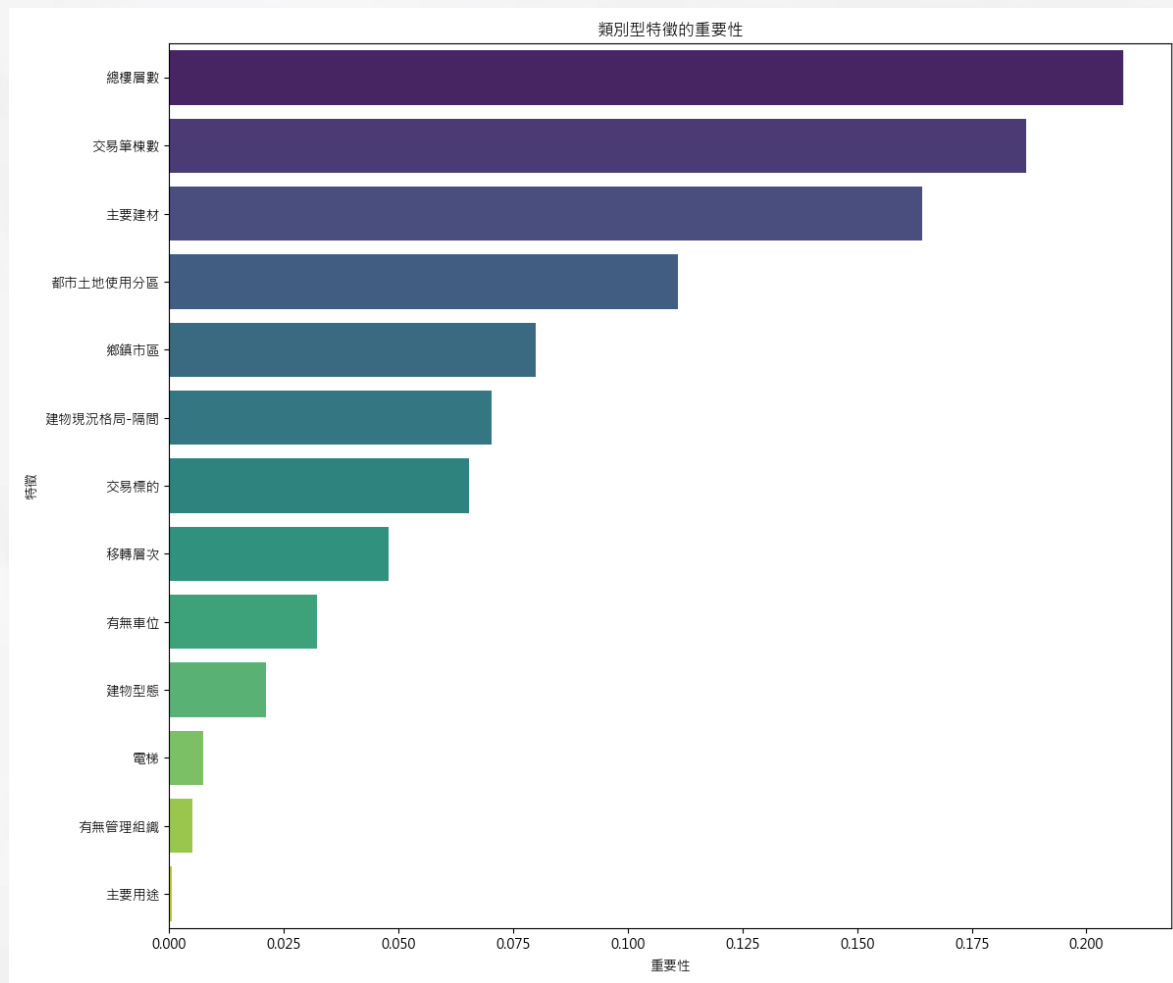
總價元 和 建物移轉總面積平方公尺：相關係數 = 0.85

車位總價元 和 車位移轉總面積平方公尺：相關係數 = 0.76

總價元 和 主建物面積：相關係數 = 0.72

```
In [17]:
```

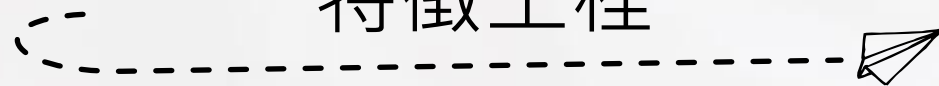
特徵篩選



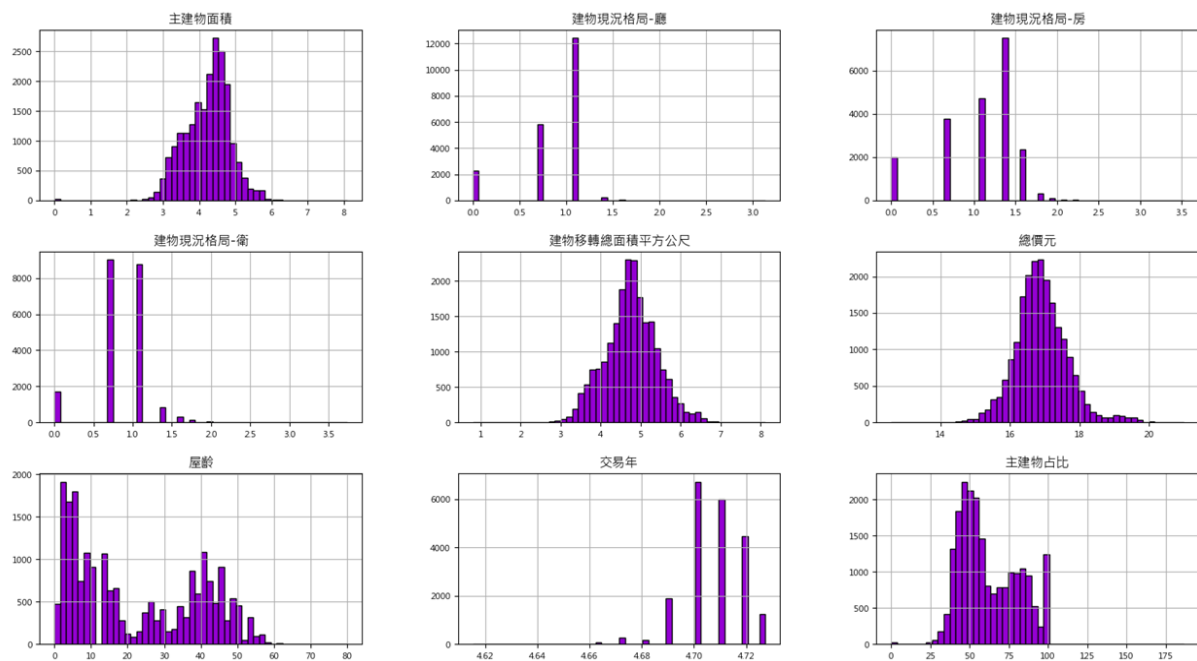
★ 類別型特徵

類別特徵在做完one-hot編碼後代入到隨機森林篩選特徵重要性。

特徵工程



對偏態的特徵作對數轉換



對於隨機森林，這對數轉換有助於提高預測的穩定性和準確性。

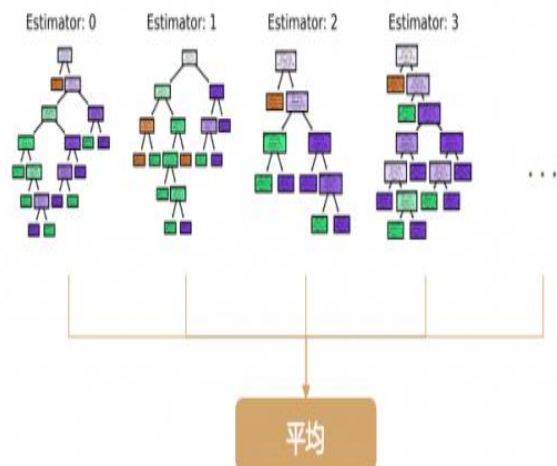
模型訓練

使用sklearn套件進行模型訓練

這裡使用隨機森林，進行房價預測。

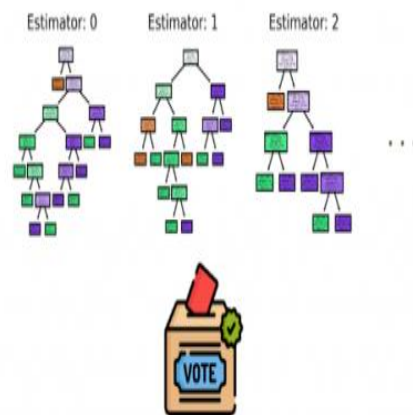
Mean Squared Error (MSE): 312086130970037.7
Root Mean Squared Error (RMSE): 17665959.667395305
Mean Absolute Error (MAE): 4712637.904179566
R2 Score: 0.8322855174606917

Regression



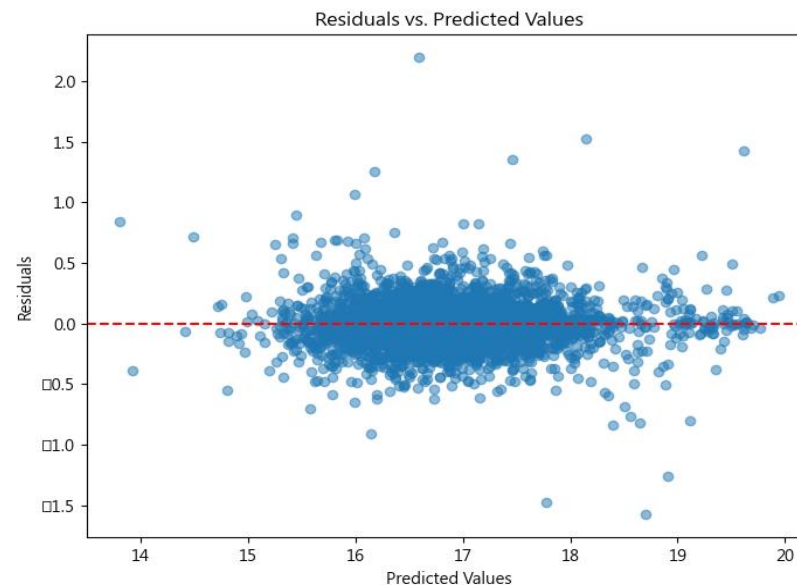
第13屆 IT 邦幫忙 鐵人賽 AI & Data 組

Classification

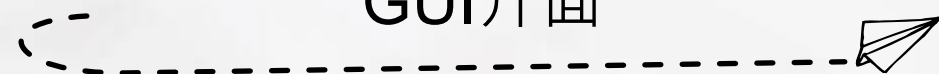


10程式中

殘差圖



GUI介面



使用tkinter套件架設GUI介面

成品展示

房價預測與附近設施查詢

主建物面積(平方公尺)

建物型態

建物現況格局-廳

建物現況格局-房

建物現況格局-衛

建物移轉總面積(平方公尺)

有無組織管理

移轉層次

總樓層數

鄉鎮市區

電梯

屋齡

有無車位

地址

學校: 0
銀行: 0
醫院: 0
公車站: 0
捷運站: 0
商店: 0

確定