

永豐 AI GO 競賽-攻房戰

下載格式

提供參賽者下載之檔案共分三大類：

1. 訓練資料集，包含【房屋估價訓練資料集】，【外部輔助資料集】。
(於測試賽期間 2023/09/18-2023/11/13 開放下載)
2. Public 測試資料集。(於測試賽期間 2023/09/18-2023/11/13 開放下載)
3. Private 測試資料集。(於正式賽期間 2023/11/06-2023/11/13 開放下載)

所有資料集皆為 UTF-8(無 BOM 檔首)編碼格式之 csv 檔，說明如下：

1. 房屋估價訓練資料集

- 第一列(row)為欄位名稱

欄位名稱	欄位說明
ID	流水號
縣市	該筆不動產所在之縣市
鄉鎮市區	該筆不動產所在之行政區
路名	該筆不動產所在路名
土地面積	該筆不動產的土地面積 (已去識別化)
使用分區	有關都市計劃使用分區
移轉層次	該筆不動產所在樓層
總樓層數	該筆不動產總樓層
主要用途	該筆不動產的用途
主要建材	該筆不動產使用的建材
建物型態	該筆不動產種類
屋齡	該筆不動產屋齡
建物面積	該筆不動產建物面積 (已去識別化)
車位面積	該筆不動產車位面積 (已去識別化)
車位個數	該筆不動產車位個數
橫坐標	該筆不動產橫坐標 (TWD97 二度分帶座標值，已去識別化)
縱坐標	該筆不動產縱坐標 (TWD97 二度分帶座標值，已去識別化)

欄位名稱	欄位說明
備註	該筆不動產之備註內容
主建物面積	該筆不動產主建物面積 (已去識別化)
陽台面積	該筆不動產陽台面積 (已去識別化)
附屬建物面積	該筆不動產附屬建物面積 (已去識別化)
單價	該筆不動產單價。 預測目標 (新台幣/坪，已去識別化)

- 第二列(row)開始每一列(row)為一筆資料。有一部分欄位資料已使用 z-score 做去識別化。另，每筆不動產之橫、縱坐標有做小幅度移動做去識別化。

2. 外部輔助資料集

- 外部輔助資料集為透過公開資料整理而成的資料集，參賽者可自行運用於增加訓練資料集特徵，輔助建模。
- 第一列(row)為欄位名稱，每份資料集的欄位不盡相同，但每份資料集的最後兩個欄位為 lat 和 lng，說明如下：

欄位名稱	欄位說明
lat	該筆資料十進位緯度座標
lng	該筆資料十進位經度座標

- 第二列(row)開始每一列(row)為一筆資料。

上傳格式

提交檔案請使用.csv 檔，上傳檔案內容請使用 UTF-8(無 BOM 檔首)編碼，並使用 Unix 系統換行字符。請勿使用其他 Non-Printable Characters，以避免評分失敗的可能。檔案中第一列(row)需為欄位名稱，欄位名稱為“ID”，“predicted_price”。第二列(row)開始的每一列(row)必須含有與原資料集相同 ID，以及參賽者預測之單價，如下範例所示：

```
ID,predicted_price
PU-1,-1.348798769814341
PU-2,1.356476909089165
...
```

評分方式

本次競賽以 MAPE (Mean Absolute Percentage Error) 作為評估模型優劣的方法。MAPE 公式如下：

$$MAPE = \frac{1}{N} \sum_1^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%$$

其中 N 為資料集總數， y_i 為真實價格， \hat{y}_i 為參賽者模型預測價格。上傳的 Public 與 Private 測試資料集預測結果之 MAPE 會分別計算。