

團隊程式說明

報名序號：111011 (報名序號(格式:111XXX)已寄至隊長email)

團隊名稱：_____Urban2.0_____

註1：請用本PowerPoint 文件撰寫團隊程式說明，請轉成PDF檔案繳交。

註2：本文件格式及安裝內容僅供參考，請依實際需求安裝並撰寫即可。

註3：依據競賽須知第七條，第4項規定：

測試報告之簡報資料不得出現企業、學校系所標誌、提及企業名稱、學校系所、教授姓名及任何可供辨識參賽團隊組織或個人身分的資料或資訊，違者取消參賽資格或由評審會議決議處理方式。

一、程式執行的安裝環境說明

- 作業系統：Ubuntu 18.04.6 LTS
- 程式語言：Python 3.8
- 工具軟體：
 - jupyter==1.0.0
 - lightgbm==3.2.1
 - matplotlib==3.5.3
 - numpy==1.23.2
 - pandas==1.4.3
 - plotly==5.10.0
 - pyyaml==6.0
 - scikit-learn==1.1.2
 - seaborn==0.11.2

一、程式執行的安裝環境說明 (cont.)

1. 在Ubuntu 18.04.6下安裝Anaconda
2. 創建本專案的conda environment並安裝需使用的套件
 - a. `conda create --name verify python=3.8 -y`
 - b. `conda activate verify`
 - c. `pip install -r requirements.txt`
3. 創建程式執行所需之kernel
 - a. `ipython kernel install --name verify --user`

二、程式執行步驟說明

- 我們認為本次競賽需要花更多心力在探索式資料分析 (EDA) 上，而非採取複雜的建模手法。因此，我們提供兩個主要的檔案來讓主辦單位驗證最終預測結果是如何產生。

1. `eda_preliminary_round.ipynb`

- 此notebook包含詳細的資料分析以及對於資料的解讀。

2. `gen_submission_preliminary_round.ipynb`

- 經過反覆的資料分析及揣摩，我們發想出獨特的建模預測手法。
- 此notebook將一步步帶領主辦單位得到最後的預測結果。

二、程式執行步驟說明 (cont.)

1. 在剛剛創建的conda environment內開啟jupyterlab server
 - a. `jupyter-lab --port <port_number>`
 - b. 打開瀏覽器並至 `http://localhost:<port_number>/` 開啟jupyterlab
2. 開始執行程式並驗證資料分析過程
 - a. 打開notebooks/eda_preliminary_round.ipynb
 - b. 在右上角選取剛才創建的kernel – verify
 - c. 依次執行各個cell之程式碼
 - d. 執行完成後，data/processed/中會有一個由我們重新建構的dataset complete.csv

二、程式執行步驟說明 (cont.)

3. 開始執行程式並獲取最終預測結果

- a. 打開notebooks/gen_submission_preliminary_round.ipynb
- b. 在右上角選取剛才創建的kernel – verify
- c. 依次執行各個cell之程式碼
- d. 執行完成後，預測結果就會被寫入111011_TestResult.xlsx中
- e. 完成!

二、程式執行步驟說明 (cont.)

Generate Final Submission File

最後，我們將預測結果生成並放到 `111011_TestResult.xlsx` 中。

```
[18]: assert not (df_complete[TARGET] == 0).any().any()

[19]: df_sub = df_complete[df_complete["test"]][TARGET+["test_index"]]
      assert len(df_sub) == 100

[20]: df_final_sub = pd.read_excel("../data/raw/2111999_TestResult.xlsx")
      for _, r in df_sub.iterrows():
          df_final_sub.loc[r["test_index"], TARGET] = r[TARGET]

      assert not df_final_sub.isna().any().any()

[42]: df_final_sub.to_excel("../111011_TestResult.xlsx", index=False)|
```

圖1. 將預測結果寫入111011_TestResult.xlsx