

PROBLEM STATEMENT - 2

Car Park Availability Analysis and Predictive Modelling

The challenge is aimed at making use of machine learning to predict the availability of parking spaces in specific Hong Kong car parks through different periods of time. The dataset made available to participants is sourced from a Hong Kong government API, and can be found on data.gov.hk.

In this project, we aim to create a machine learning model to predict the amount of vacancies at a specific car park or district at a given time of day, which will be inputted by the user. The deployed model will take in the user's requests of time and return the expected amount of vacancies by carpark.

The datasets are available at:

Basic Info: https://resource.data.one.gov.hk/td/carpark/basic_info_all.json

CarPark Availability: https://resource.data.one.gov.hk/td/carpark/vacancy_all.json

PROBLEM DESCRIPTION

1. [Basic] Probable Use case to implement - Predicting the availability of given car parks

- User will provide the full date and time rounded off to the closest 15 minute interval as an input
- The output should be a prediction of the amount of car park spaces available at the given car park.

Description:

The basic version of the problem will only require participants to predict the car park availability at the user's input time in Lee Garden One Car Park.

2. [Advanced] Probable Use case to implement - Predicting the availability of car parks across all HK districts.

- User will provide as an input:
 - full date
 - time rounded off to the closest 15 minute interval
 - The district they are searching
- The output should be a prediction of the amount of car park spaces available in each car park of the district provided by the user.

Description:

The advanced version of the problem will require participants to predict the availability of all car parks in the district given by the user. Preferably, the results should show the car parks ranked in order from the highest amount of parking spaces available to the lowest amount of parking spaces available.

ABOUT DATASET

The datasets are two JSON APIs provided by the HK government on data.gov.hk. One of the APIs, named `basic_info_all` consists of basic information regarding all the car parks in Hong Kong, including `name`, `park_id` and `district` of car park. The other API, called `vacancy_all` contains the `park_id` alongside specific information on car park vacancies.

For students who will be undergoing the advanced challenge, please make sure you only collect data from car parks with “`type`” value “`P`”, signifying private car spaces, and “`vacancy_type`” value “`A`”, which gives a continuous value regarding vacancy.

Please see the below image for reference of acceptable car park data for collection.

```
{
  "park_id": "tdc1p1",
  "vehicle_type": [
    {
      "type": "P",
      "service_category": [
        {
          "category": "HOURLY",
          "vacancy_type": "A",
          "vacancy": 124,
          "lastupdate": "2021-06-13 00:40:06"
        }
      ]
    }
  ]
},
```

SUBMISSION GUIDELINES

What to submit?

Each group's submission should include, but not be limited to:

1) Codes (Score: 50%)

A Python notebook (**submitted in Google Colab .ipynb format**) of the model including:

- Data collection from GovHK API and database set up
- Exploratory Data Analysis (EDA)
- Model training
- Model prediction and evaluation
- Function for adding single instance of input of date and time, and returning prediction

2) Documentations (Score: 50%)

Presentation slides (**submitted in PDF format**) containing:

- Introduction of problem and approach
- Findings from EDA, data preprocessing and visualization of data
- Workflow documentation of working process, including:
 - What features were created and selected
 - Evaluation of results of model deployed
 - Problems in process of developing model and solutions, including:
 - Explanation of model utilized for deployment and the background of why it was chosen
 - Explanation should also include models considered previously, and reasons to why they weren't chosen for deployment
 - Limitations of the models as well as problems that occurred in the process should be documented as well
- Possibilities after deployment, and next steps of development
 - The possibilities after deployment, and propose the next steps of development for future applications.

Remarks: The slides prepared should be suitable for a 10-minute presentation.

How to submit?

The submission deadline is June 20, 23:59 HKT. Here are the points to note:

Email Subject: DTT Mini-project Submission - (*Your Group Number*)

Attachment: A compressed file (zip-file) renamed as (*Your Group Number*).zip

*Your zip file should include an *.ipynb* file and a PDF.

EXPECTATION

- Participants are expected to choose one of the problems (Basic or Advanced) to complete. Those who have chosen to attempt the Basic problem might get a maximum score of 80; while those who attempt the Advanced problem could get a full mark of 100.
- Participating teams with individuals who are caught cheating will be disqualified from the competition immediately.

IMPORTANT DATES

Submission deadline: June 20 (23:59)

*Late submission will not be accepted.

Announcement of result: June 25

*The results of the Hackathon will be announced through email. The selected teams will be notified and invited to participate in the presentation day.

Presentation day: July 2