



Global AI Challenge for Building E&M Facilities – AI Competition

Challenge Rules & Guidelines

Last updated: 6th Dec 2021

Contents:

1. Overview
2. Timeline
3. Data Description
4. Evaluation Criteria
5. Platform login
6. Overall workflow
7. Download dataset
8. Submission
9. Support
-  10. Structure of Python Scripts

1. Overview

Problem:

Turning the chillers on and off in a commercial building is more than just pressing a few buttons. When determining which chillers should be on or off and their operation parameters for large buildings, facility management needs to make assumptions such as the upcoming cooling load demand and external environment factors. However, these assumptions do not reflect real conditions, and may waste electricity energy when the cooling load is assumed too high or have bad occupants' ratings when assumed too low.

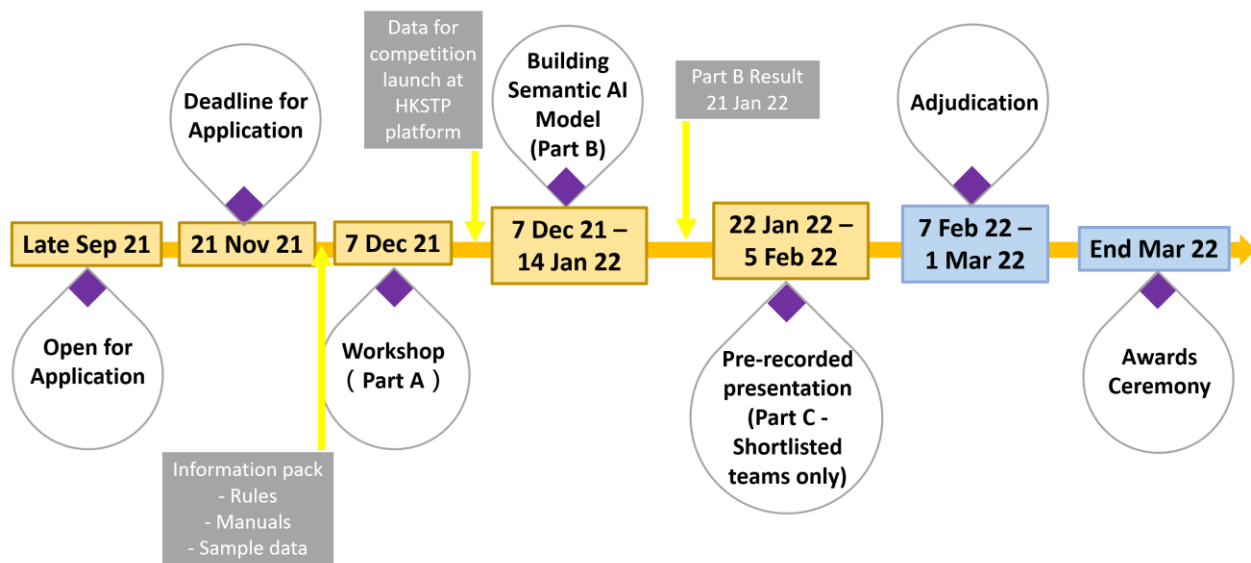
Proposed solutions:

If a more accurate building cooling load forecast can be predicted, facility managers will be able to determine the optimal chiller operation strategy to improve energy efficiency.

Details:

In this competition, you will develop an AI model to predict the **hourly cooling load** of a commercial complex. The data you will use for your prediction model training and testing comes from a multi-chiller system of a commercial complex in Hong Kong. The data will be provided to you in a CSV format, and you are required to use this data to develop a Cooling Load Prediction model that can accurately **predict the next 3 months of hourly cooling load** of this commercial complex.

2. Timeline



Summary of Deadlines (GMT+8):

[Optional] Interim submission: 28th Dec 2021, 23:59

Final submission: 14th Jan 2022, 23:59

Questionnaire submission: 14th Jan 2022, 23:59

[Only shortlisted teams] Pre-recorded presentation submission: 5th Feb 2022, 23:59

3. Data Description

Data to be released to contestants for developing model:

The training data set contains cooling load data related to a commercial building complex located in Hong Kong, from April 1, 2020, to September 30, 2021. ***Data pre-processing work is required.**

The first 15 months of the data (April 1, 2021, to June 30, 2021) is provided as a single CSV file which can be downloaded directly from HKSTP platform (xdp.bip.hkstp.org/user/login). The last three months of the training data (July 1, 2021, to September 30, 2021) will have to be extracted by contestants from the Semantic AI Platform.

Using the provided training data, contestants are required to develop a prediction model that can accurately predict the **Hourly Cooling Load** of this commercial building complex for **the next 3 months (October 1, 2021 – 31 December, 2021)**.

Description of the Training Data

The first 15 months (April 1, 2020 to June 30 2021) of the training data is provided in a single CSV. This data set has eight columns in it. The data is in a 15 minutes sampling interval.

The variables in this CSV file are Timestamp, Temperature, Humidity, UV Index, Rainfall, Cooling Load for South Tower, Cooling Load for North Tower and the Combined Cooling Load of these two towers (i.e., summation of values in the Cooling Load for North Tower and South Tower).

Name of CSV: CoolingLoad15months.csv (4MB)

Total number of records: 43776

Timestamp	Average_OAT	Humidity	UV_Index	Average_Rainfall	ST_Cooling Load	NT_Cooling Load	CoolingLoad
2020-02-03 00:00:00	18.07	81	0.1	0	264	152	416
2020-02-03 00:15:00	18.07	81	0.1	0	264	152	416
2020-02-03 00:30:00	18.07	81	0.1	0	264	152	416
2020-02-03 00:45:00	19.22	78	0.5	0	281	210	491
2020-02-03 01:00:00	19.09	78	0.5	0	211	199	410
2020-02-03 01:15:00	19.21	78	0.5	0	223	204	427
2020-02-03 01:30:00	20.78	78	0.5	0	290	251	541
2020-02-03 01:45:00	20.11	78	2	0	286	281	567
2020-02-03 02:00:00	20.14	78	2	0	308	345	653
2020-02-03 02:15:00	19.80	78	2	0	317	423	740
2020-02-03 02:30:00	19.65	78	2	0	307	448	755
2020-02-03 02:45:00	19.69	75	3	0	324	448	772

Description of the variables:

1. Timestamp: **Timestamp of the observation** (yyyy-mm-dd hh:mm:ss)
2. **Average_OAT**: Average outdoor **air** temperature (degree Celsius)
3. **Humidity**: Relative Humidity **of outdoor air**(%)
4. UV Index: UV Index
5. **Average Rainfall**: Average rainfall (mm)
6. ST_CoolingLoad: Cooling Load from South Tower (kW)
7. NT_CoolingLoad: Cooling Load from North Tower (kW)
8. CoolingLoad: Summation of values from ST_CoolingLoad & NT_CoolingLoad (kW)

The hourly cooling load can be computed as the average of the 15 minutes cooling load readings in that hour. For example, the hourly cooling load at 00:00 is:

$$(CoolingLoad\ at\ 00:00 + CoolingLoad\ at\ 00:15 + CoolingLoad\ at\ 00:30 + CoolingLoad\ at\ 00:45) / 4$$

The remaining three months of the training data (July 1, 2021, to September 30, 2021) can be retrieved from the Semantic AI platform.

<https://semantic1.globalaichallenge.com/competition>

<https://semantic2.globalaichallenge.com/competition>

(data launch at 18:00 (GMT+8) on 7th Dec 2021)

Contestants can navigate to the Semantic AI Platform (using either of the two links above) to download the respective data to construct variables (features and target) of the last three months training data. Refer to Section 7 Download dataset for the details of downloading the rest of the training data from the Semantic AI platform.

The mapping of the semantic data point to variables in the training dataset (CoolingLoad15months.csv) is depicted in the below table:

Point Number	Point Name	Description	Unit	Name of CSV	Variables
1	CP-ST-CHWS-EM:FLOWRATE	Chilled Water Supply Flow Rate from South Tower Header Pipe	L/s	3159.csv	Compute value for ST_CoolingLoad
2	CP.ST.CHWS.TEMP.1	Temperature of chilled water that is supplied to south tower	°C	1411.csv	
3	CP.ST.CHWR.TEMP.1	Temperature of chilled water that is returned from south tower	°C	1409.csv	
4	CP-NT-CHWS-EM:FLOWRATE	Chilled Water Supply Flow Rate from South Tower Header Pipe	L/s	3133.csv	Compute value for NT_CoolingLoad
5	CP..NT.CHWS.TEMP.1	Temperature of chilled water that is supplied to north tower	°C	1361.csv	
6	CP.NT.CHWR.TEMP.1	Temperature of chilled water that is returned from north tower	°C	1401.csv	
7	ST.RF.OUTDOOR.TEMP.1	Temperature of outside air	°C	3007.csv	Average_OAT
8	ST.RF.OUTDOOR.HUMV	Relative humidity of outside air	%	3028.csv	Humidity
9	ST.RF.OUTDOOE.RAIN1	Amount of precipitation fallen	mm	9998.csv	Average_Rainfall
10	ST.RF.OUTDOOR.UV1	UV index		9999.csv	UV_Index
11	CP.CWR.TEMP.1	Temperature of water entering a condenser from cooling tower	°C	1392.csv	
12	CP.CWS.TEMP.1	Temperature of water leaving a condenser to cooling tower	°C	1399.csv	

Contestants can refer to the below formula to compute the values of Cooling Load for South Tower (ST_CoolingLoad) and North Tower (NT_CoolingLoad). The combined cooling load (CoolingLoad) is the sum of the cooling load for South Tower (ST_CoolingLoad) and North Tower (NT_CoolingLoad).

Cooling Load = mc ΔT

- m is the flow rate (L/s) x density of water (1kg/L)
- C is the heat capacity (4.19kJ/kg °C)
- ΔT is the difference between the inlet (return) and outlet (supply) chilled water temperature (°C)

$$\text{Cooling Load} = \text{Flow Rate} \times 1 \times 4.19 \times \Delta T$$

- Total Cooling Load (CoolingLoad) = South Tower Cooling Load + North Tower Cooling Load
- The weather-related data can be obtained from point number 7, 8, 9 and 10

Data to be released to contestants for testing their models:

A single seven-day testing data (in CSV format) from September 24, 2021, to September 30, 2021 is prepared for contestants. Using this seven-day testing data, contestants are required to generate a CSV dataset with predicted **hourly cooling load** (i.e., combined hourly cooling load for South Tower and North Tower) for that period using their to-be submitted Cooling Load Prediction models. The output dataset (CSV) must have the exact format as described in below. Contestants who fail to submit the prediction output CSV file with the name and format described below will be disqualified.

Description of the Testing Data

- Name of the CSV: CoolingLoad7days.csv (31kB)
- Period: September 24, 2021, to September 30, 2021 (data in 15 minutes interval)
- Total of 672 records

Timestamp	Average_OAT	Humidity	UV_Index	Average_Rainfall
2021-10-01 00:00:00	*****	*****	*****	*****
2021-10-02 00:15:00	*****	*****	*****	*****
2021-10-03 00:30:00	*****	*****	*****	*****
2021-10-04 00:45:00	*****	*****	*****	*****
2021-10-05 01:00:00	*****	*****	*****	*****
..	*****	*****	*****	*****
..	*****	*****	*****	*****
2021-12-31 22:15:00	*****	*****	*****	*****
2021-12-31 22:30:00	*****	*****	*****	*****
2021-12-31 22:45:00	*****	*****	*****	*****
2021-12-31 23:00:00	*****	*****	*****	*****
2021-12-31 23:15:00	*****	*****	*****	*****
2021-12-31 23:30:00	*****	*****	*****	*****
2021-12-31 23:45:00	*****	*****	*****	*****

Expected Output File Name and Data Format from the Prediction Model

The name and data format from the **Cooling Load Prediction Model** should be:

Name of CSV: *YourTeamName_output.csv*

Number of columns: 2

Name of Column 1: Timestamp

- Format: yyyy-mm-dd hh:mm:ss

Name of Column 2: CoolingLoad

Format: Numeric

Description of the Testing Output CSV (to be submitted):

Contestants are required to submit this 7-day hourly prediction output CSV file with the following convention for filename, column names and format in it:

Name of CSV: *YourTeamName_output.csv*

Number of records in the csv: 168

Number of columns in the csv: 2

Name of Column 1: Timestamp

- Format: yyyy-mm-dd hh:mm:ss

Name of Column 2: CoolingLoad

- Format: Total Cooling Load of the Chiller System (i.e., South Tower Cooling Load + North Tower Cooling Load) – Numeric

Timestamp	CoolingLoad
2021-09-24 00:00:00	787
2021-09-24 01:00:00	733
2021-09-24 02:00:00	617
2021-09-24 03:00:00	588
2021-09-24 04:00:00	598
2021-09-24 05:00:00	643
...	..
...	..
...	..
...	..
2021-09-30 20:00:00	1897
2021-09-30 21:00:00	1589
2021-09-30 22:00:00	1344
2021-09-30 23:00:00	1105

Data that will be used for model evaluation

Contestants' submitted models will be evaluated based on the accuracy of the **hourly cooling load** from your Cooling Load Prediction model for the next three months (**October 1, 2021, to December 31, 2021**).

The format of the evaluation data set is same as the format of the testing data set (CoolingLoad7days.csv).

Timestamp	Average_OAT	Humidity	UV_Index	Average_Rainfall
2021-09-24 00:00:00	20.89	82	0.1	0
2021-09-24 00:15:00	20.52	80	0	0
2021-09-24 00:30:00	20.94	80	0	0
2021-09-24 00:45:00	18.68	80	0	0
2021-09-24 01:00:00	19.02	80	0	0
..
..
2021-09-30 22:15:00	19.14	81	0	0
2021-09-30 22:30:00	19.02	81	0	0
2021-09-30 22:45:00	19.72	78	0	0
2021-09-30 23:00:00	19.84	78	0	0
2021-09-30 23:15:00	19.63	78	0	0
2021-09-30 23:30:00	19.42	79	0	0
2021-09-30 23:45:00	19.20	80	0	0

4. Evaluation Criteria

The evaluation metric for this part of competition is based on the following:

- ✧ The accuracy of your Hourly Cooling Load Prediction model is measured by the Root Mean Square Error (RMSE):

$$\epsilon = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - a_i)^2}$$

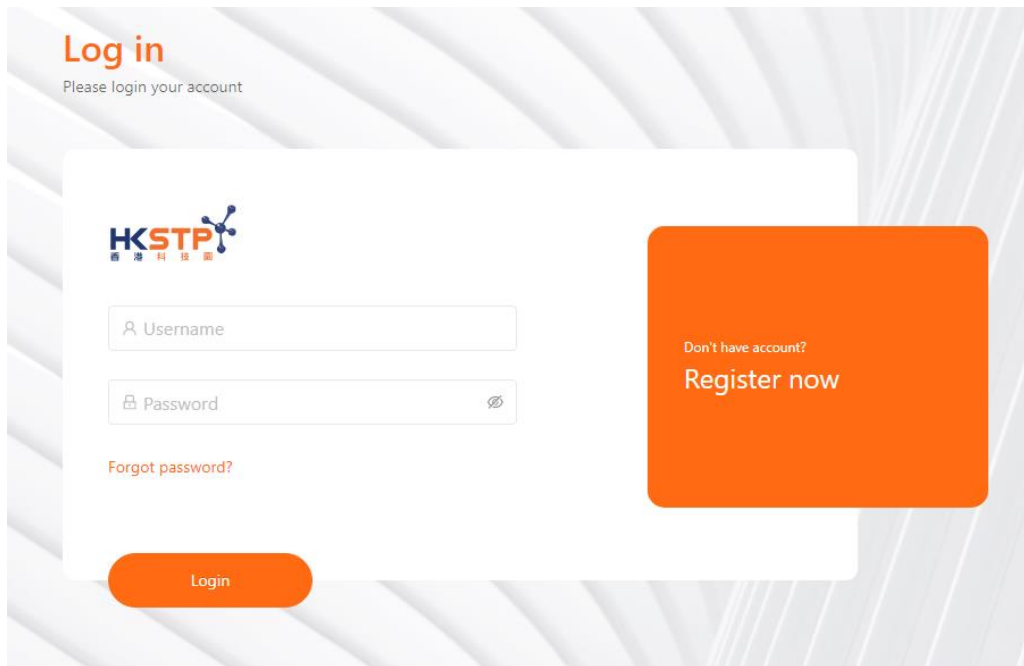
Where

- ϵ is the RMSE value (score)
- n is the total number of forecasts, which is the next 3 months of the data in the dataset
- p_i is your prediction value of the cooling load demand in kW, and a_i is the actual cooling load in kW

5. Platform login

Contestants will receive a **“first time login password”** e-mail. Login will be available since Day 1 (December 7, 2021) of the contest. Please follow the steps below to reset a new password after first time login.

- Go to <https://xdp.bip.hkstp.org/user/login>



Log in
Please login your account

HKSTP
香港科技園

Username

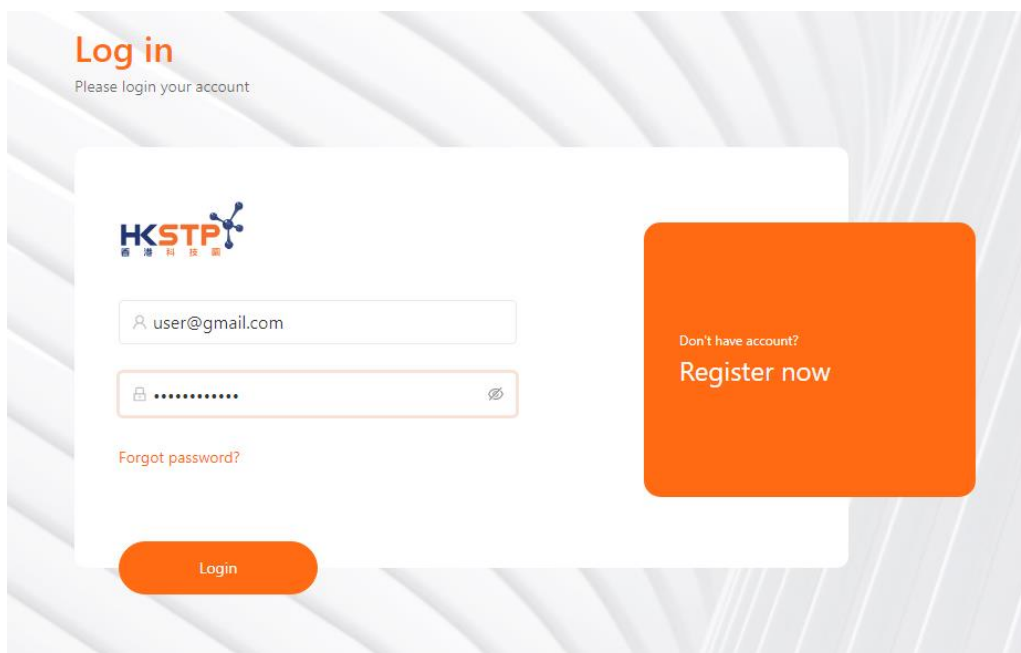
Password

[Forgot password?](#)

Login

Don't have account?
Register now

- Login with the password provided in the e-mail



Log in
Please login your account

HKSTP
香港科技園

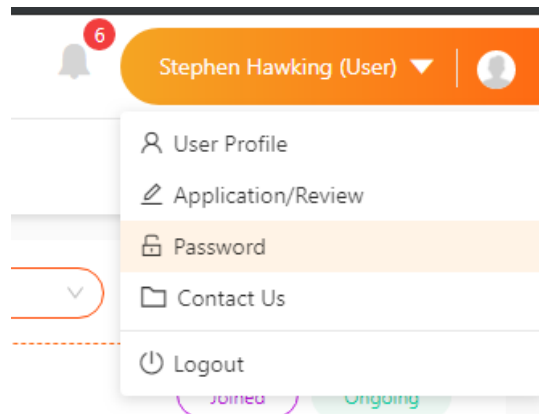
user@gmail.com

[Forgot password?](#)

Login

Don't have account?
Register now

- Navigate to the setting drop-down list and select “Password”



- Change your password and “Save”

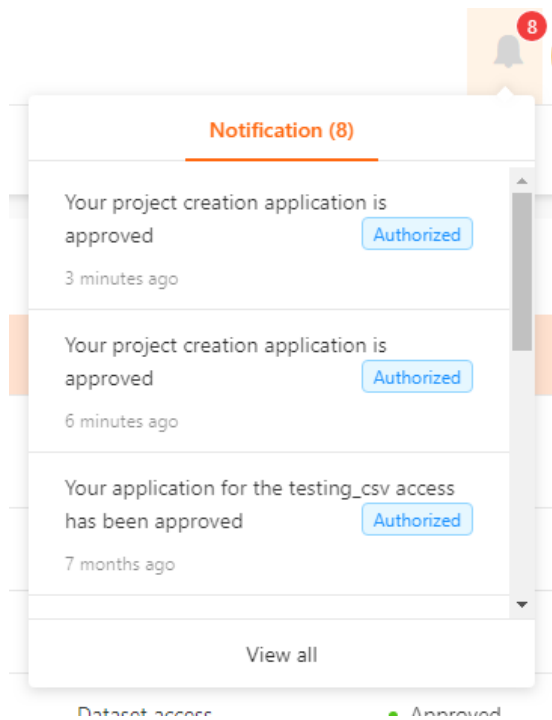
Current password:

New password:

Confirm password:

Login & join project

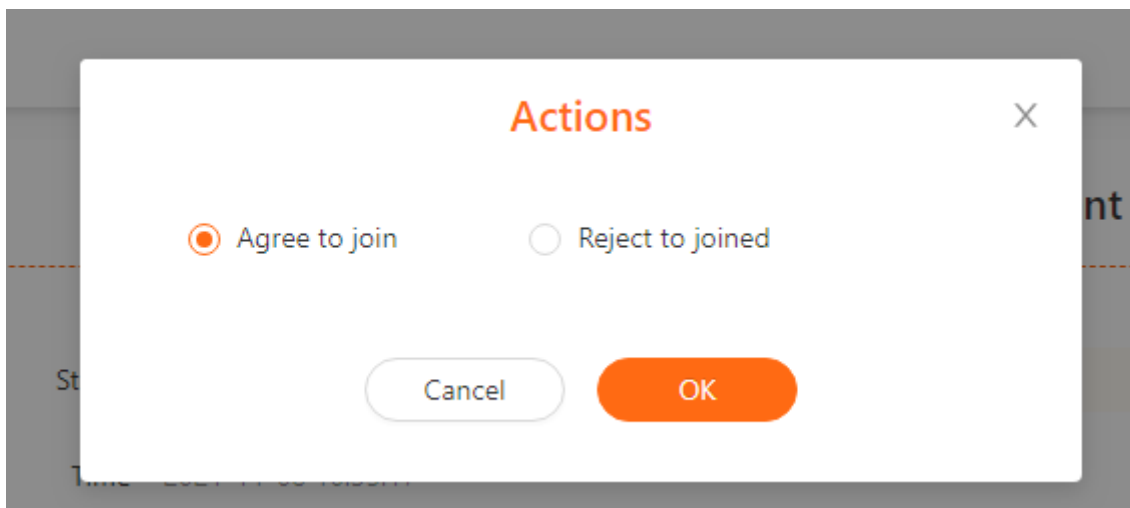
1. Once logged into the Platform, click on the “notification icon” → “view all”



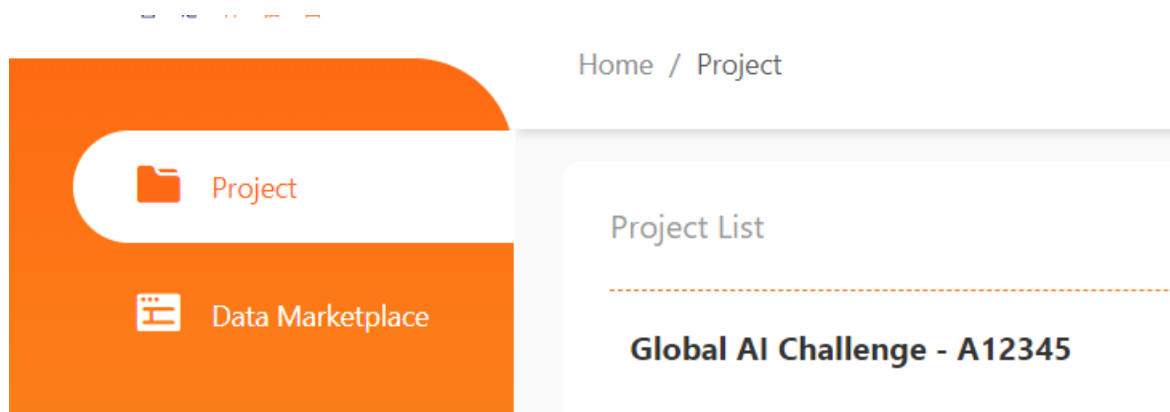
2. Find the project invitation and click “view” on the right

Time	Notification detail	Review type	Review status	Actions
2021-11-08 10:35:30	Darwin (Group Leader) Charles(cd287588@gmail.com) invited you to join this project: emsd - A	Project	• Pending	<div>View</div> <div></div>

3. Click “Actions” and select “Agree to join”



4. Contestant should be able to see the project in “Project” session



(where A12345 refers to your own team name)

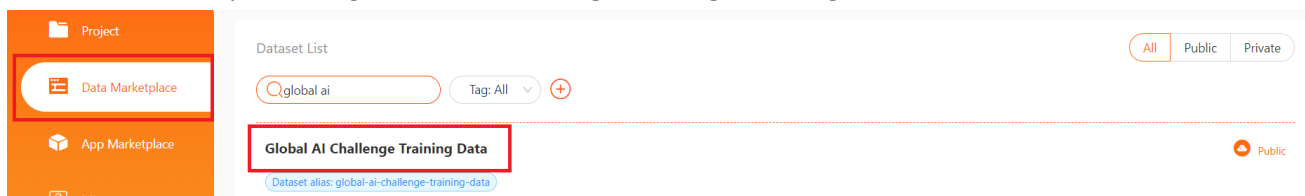
Overall workflow

1. Download training data for model training
2. Build and train the model with your own resources
3. **[Optional]** Go to STP Platform and submit interim submission files
4. Go to STP Platform and submit all the required deliverables
5. Evaluate your model with the evaluation dataset

6. Download dataset

Training Dataset (first 15 months (April 1, 2020 – 30 June, 2021))

1. Contestant should download the training data from the Data Marketplace after logging in to the Platform by searching “Global AI Challenge Training & Testing Data”



2. Go to “Global AI Challenge Training Data” → “Overview” → Click on the link and download the datasets for your model training purposes

Training Dataset for **competition** (15 months of data)

https://global-ai-challenge-training-data.s3.ap-east-1.amazonaws.com/Training+Dataset_CoolingLoad15months.rar

Testing Dataset for **competition** (7 days of data)

https://global-ai-challenge-training-data.s3.ap-east-1.amazonaws.com/Testing+Dataset_CoolingLoad7days.rar

(data launch at 18:00 (GMT+8) on 7th Dec 2021)

Training Dataset (the subsequent 3 months (April 1, 2020 – 30 June, 2021))

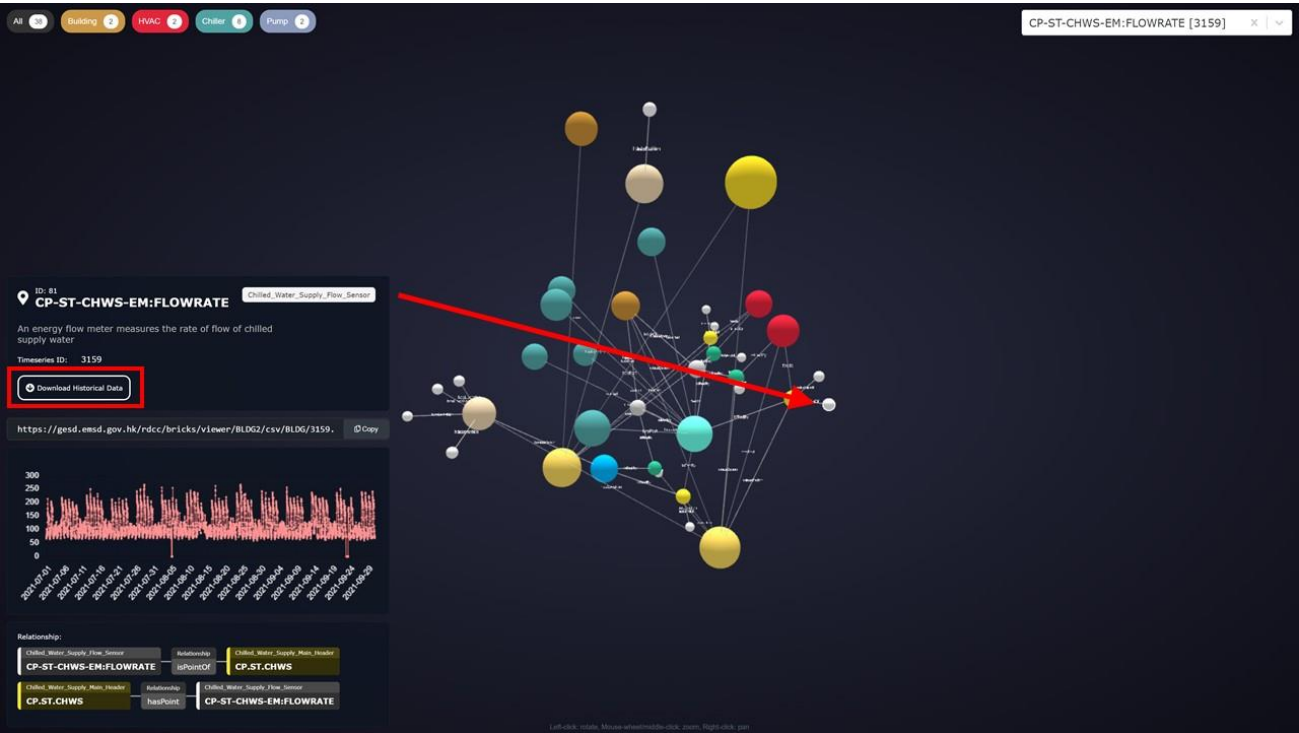
The subsequent 3 months training data can be extracted from Semantic AI Platform.

<https://semantic1.globalaichallenge.com/competition>

<https://semantic2.globalaichallenge.com/competition>

(same content in both links, data launch at 18:00 (GMT+8) on 7th Dec 2021)

Example: to download the flow rate value of chilled water goes to the south tower, click on the download link in the node (CP-ST-CHWS-EM:FLOWRATE) . A csv with name 3159.csv will be downloaded automatically.



Timestamp	Data
2021-07-01 00:00:00	70.31
2021-07-01 00:15:00	65.42
2021-07-01 00:30:00	68.61
2021-07-01 00:45:00	68.92
2021-07-01 01:00:00	75.71
2021-07-01 01:15:00	74.89
2021-07-01 01:30:00	72.99
2021-07-01 01:45:00	82.26
2021-07-01 02:00:00	102.47
2021-07-01 02:15:00	101.20
....
....
2021-09-30 22:15:00	71.98
2021-09-30 22:30:00	70.83
2021-09-30 22:45:00	76.02
2021-09-30 23:00:00	72.19
2021-09-30 23:15:00	72.96
2021-09-30 23:30:00	73.47
2021-09-30 23:45:00	67.42

Figure 1 3159.csv

7. Submission

The contestants are required to submit:

[Optional] Interim submission [by 28th Dec 2021 23:59 (GMT+8)]:

1. A CSV file of predicted model output of the testing data (7 days)
(file name: **YourTeamName_interimoutput.csv**)
2. Source codes of the testing model that can be used to generate the predictive hourly cooling load from the testing dataset
(file name: **YourTeamName_interimtesting.py**)
3. Zip file of the pickle file of Interim model (.zip)
(file name: **YourTeamName_interimmodel.zip**)
4. A text file with listing out the libraries or modules versions being used
(file name: **YourTeamName_interimrequirements.txt**)
5. A text file with listing out the language and version being used
(file name: **YourTeamName_interimversion.txt**)

Judges will run your testing model and confirm if the submitted interim model can be run successfully on or before 4th Jan 2022.

Final submission [by 14th Jan 2022 23:59 (GMT+8)]:

1. Part A: PDF file of the completed online questionnaire
(file name: **YourTeamName_questionnaire.pdf**)
2. Source codes that can be used to generate the cooling load prediction model (e.g., pickle) from the training dataset.
(file name: **YourTeamName_finaltraining.py**)
3. Zip file of the model file of Final model (.zip)
(file name: **YourTeamName_finalmodel.zip**)
4. Source codes of the testing model that can be used to generate the predictive hourly cooling load from the testing dataset
(file name: **YourTeamName_finaltesting.py**)
5. A CSV file of predicted model output of the testing data (7 days)
(file name: **YourTeamName_output.csv**)
6. A text file with listing out the libraries or modules versions being used
(file name: **YourTeamName_requirements.txt**)
7. A text file with listing out the language and version being used
(file name: **YourTeamName_version.txt**)

8. A shell script with commands to execute your final training and testing programs if they testing cannot be executed directly by entering `python – yourpythonprogram.py`
(file name: **YourTeamName_command.sh**)
9. A two-page report containing the introduction and detailed description of the AI model
Format: Single-spaced; Font Size: 12-point; Font: Times New Roman; Maximum: 2 pages
(file name: **YourTeamName_finalreport.doc**)
10. **[For Tencent cloud service users]** Screen captures of development works on the provided cloud platform
(file name: **YourTeamName_cloudplatform.jpg**)

***Remark:**

- Only Final Submission will be evaluated.

7.1 Questionnaire Submission

Contestants are required to complete an online questionnaire at below link,

https://corexms8hdymxr9zf24p.qualtrics.com/jfe/form/SV_9S8hOI59HzB3m6O

- To express their views on the benefits of semantic AI in building management;
- To provide any suggestions on areas of improvement.

You may want to try Energon before the questionnaire:

https://github.com/fangger4396/energon_example/blob/main/emsd_example_2.md

After you complete the questionnaire, please download the completed questionnaire as **PDF** and submit the PDF at HKSTP Platform.

Survey Completion
0%  100%

We thank you for your time spent taking this survey.
Your response has been recorded.

Below is a summary of your responses

[Download PDF](#)

7.2 Model Preparation

Contestant's model submitted will be tested by the AI competition judges after the submission period.

Language and Version Requirements:

Python 3 or above

Below the VM specifications that the judges would run the submitted mode:

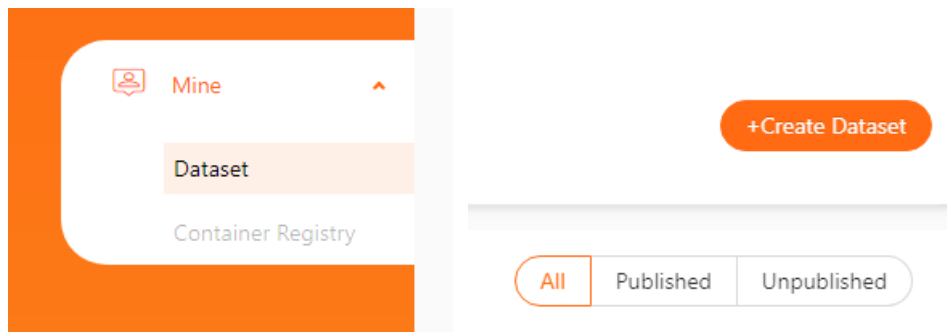
4 CPU core

16GB RAM

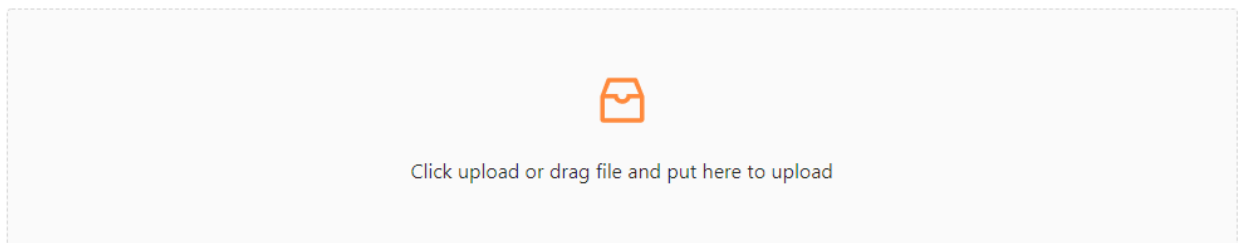
All the pre-processing steps that used to produce the final training dataset shall be included in the training.py submission code.

8.3 Deliverables submission

1. Contestants are required to submit all deliverables to HKSTP Platform before the deadlines.
2. Navigate to Mine → Dataset. Click “Create Dataset” on the right top corner



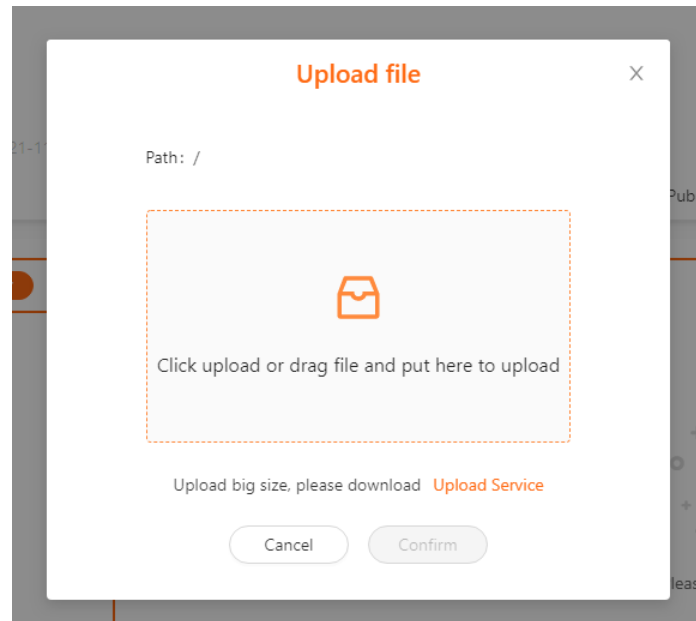
3. For interim submission **[by 28th Dec 2021 23:59 (GMT+8)]**, create a dataset with the following,
Dataset name: **Global AI Challenge O-P12345 Interim Submission**
Dataset alias: **global-ai-challenge-o-p12345-interim-submission**
***O-P12345 refers to your own team name**
4. For final submission **[by 14th Jan 2022 23:59 (GMT+8)]**, create a dataset with the following,
Dataset name: **Global AI Challenge O-P12345 Final Submission**
Dataset alias: **global-ai-challenge-o-p12345-final-submission**
***O-P12345 refers to your own team name**
5. Upload you files as below



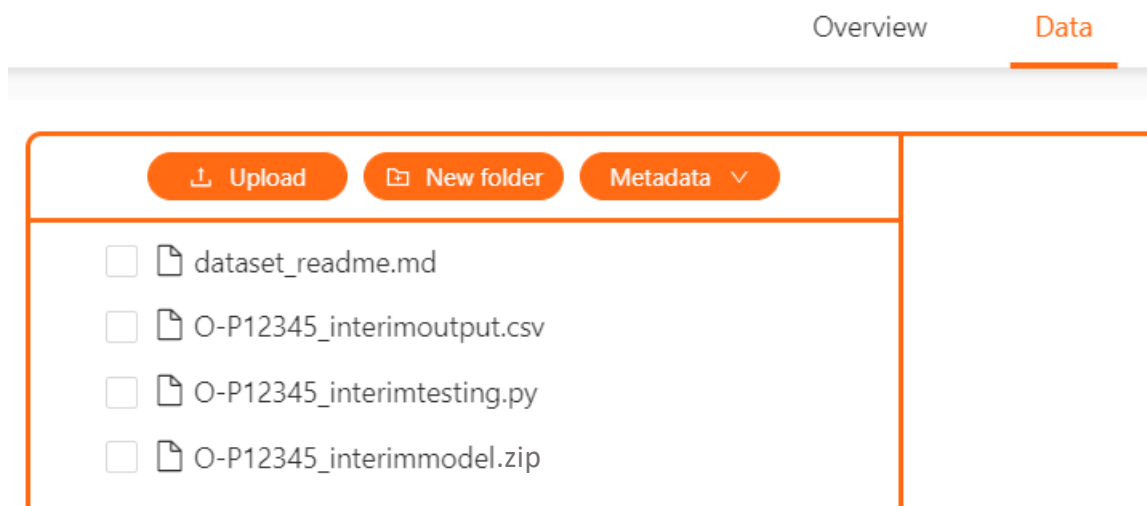
Upload big size, please download [Upload Service](#)

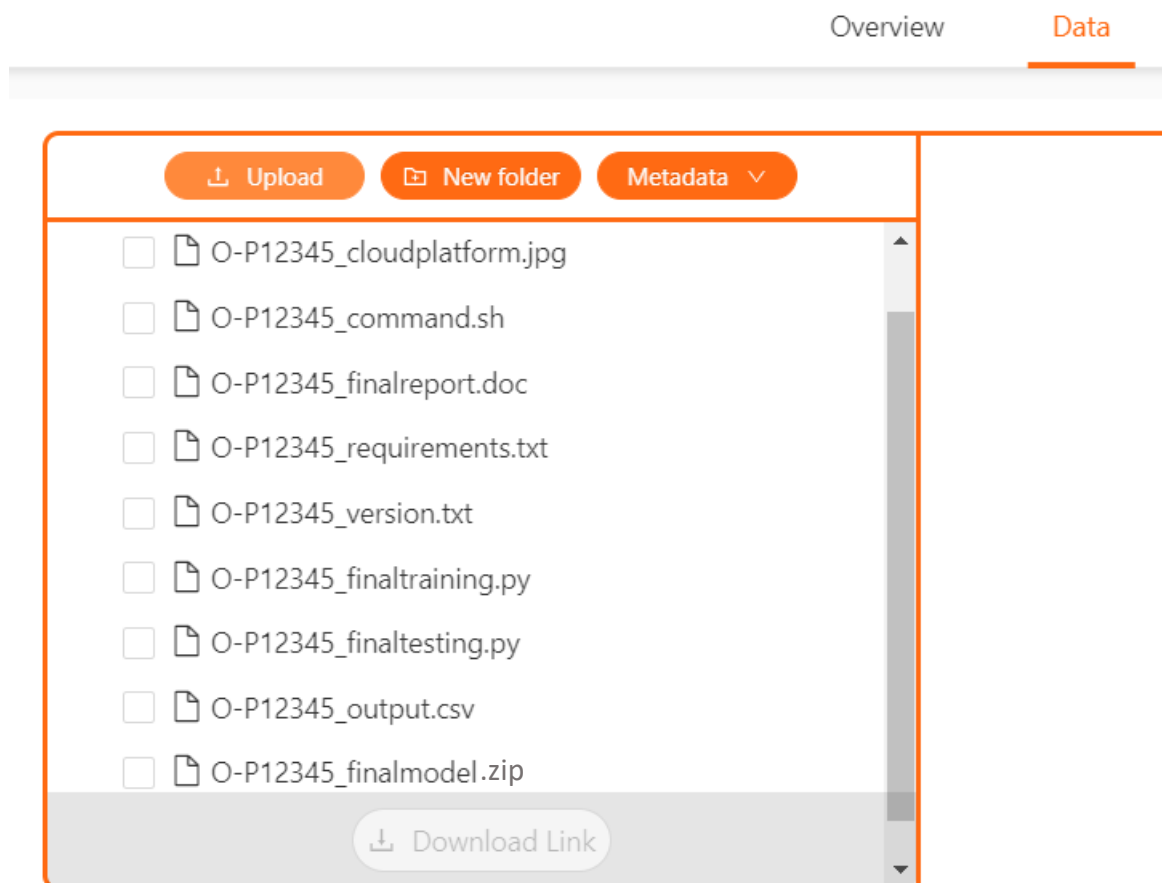
Next step

6. Alternatively, you can upload it inside the dataset session (tab “Data” → “Upload”)

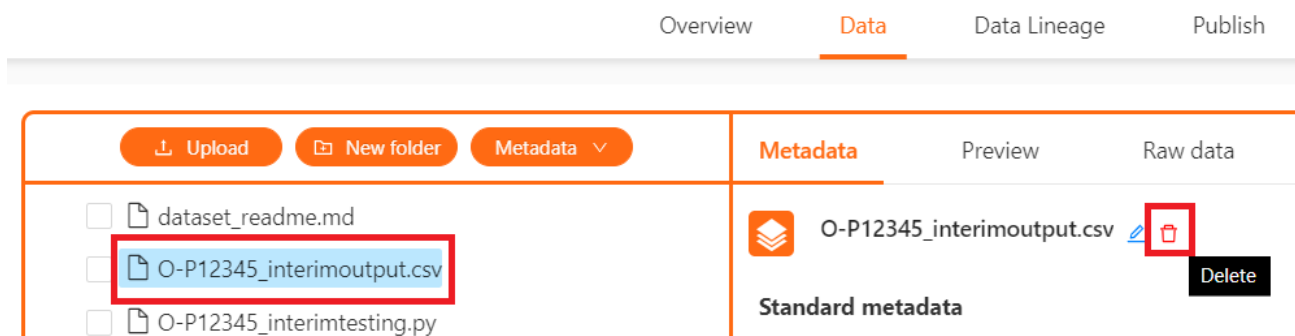


7. Check your uploaded files integrity in tab “Data”





8. In order to delete uploaded file, click on the file name and click the “bin” icon on the right



9. After the deliverables are uploaded. Please make sure you change the dataset status to **immutable** (BUT DO NOT publish the dataset) and submit.

This will freeze the dataset and you will not be able to update the content of the dataset afterwards. Please only do so when you are ready to submit.

Mutable submission after deadline will NOT be accepted.

☒ To publish ☐ Publishing ☐ Published

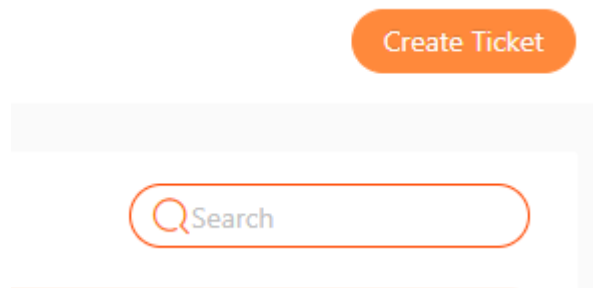
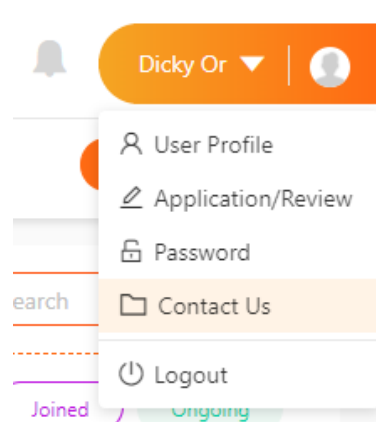
Dataset content: ☐ Mutable ☒ Immutable

Dataset publish: ☐ Publish ☒ Not to publish

Submit

9. Support

Platform helpdesk & technical support



1. Contestant can raise ticket or email to bip@hkstp.org for technical support anytime during their model submission period.
2. Enquiry will be responded within 48 hours.

10. Structure of Python Scripts

Contestants have to define directory path variables at the beginning of the finaltraining.py, finaltesting.py and the interimtesting.py.

```
# importing os.path module
import os.path

# directory path where you keep the input training and testing data sets
input_path = 'path of your input directory'

# directory path where you keep your predictor model and prediction (testing) output
output_path = 'path of your output directory'

# working directory
working_path = 'path of your working directory'

# Names of training and testing data sets provided in this competition
training_csv = "CoolingLoad15months.csv"
testing_csv = "CoolingLoad7days.csv"

# Names of your team's cooling load prediction model and the prediction output

YourTeamName_finalmodel = "YourTeamName_finalmodel.pkl"
YourTeamName_output = "YourTeamName_output.csv"

training_data = os.path.join(input_path, training_csv)
testing_data = os.path.join(input_path, testing_csv)
predictor_model = os.path.join(output_path, YourTeamName_finalmodel)
predictor_csv = os.path.join(output_path, YourTeamName_output)
```