

Individual Project Report

Team Name: [AI Predictors](#)

Project Title: [Intelligent Dengue Predictor](#)



Team Member Name: Lu Zhiping

Team Member Matriculation Number: A0229959E

Team Member Contact (Mobile/Email): lu.zhiping@u.nus.edu

1. My personal contribution to this project.

My role in the project is to be the model trainer, use my past experiences to do multiple experiments on different model training, I work with team members that consolidate the dataset the most, discuss the data size that can let the model perform the best, and select the features that can highly describe the phenomenon that we are researching for.

In the project, I have done the following:

- Select the best model, I have tried:
 1. Linear Regression
 2. ElasticNet
 3. Decision Tree Regressor
 4. RandomForest Regressor
 5. Gradient Boosting Machine Regressor from Scikit-learn
 6. LightGBM

With hyperparameter tuning using

- GridSearch in Scikit-learn
- RandomSearch in Scikit-learn
- HalvingGridSearchCV in Scikit-learn experiment

After the model exploration and hyperparameter tuning in Jupyter Notebook, I have written a Python module called `model`, inside the `model` module, there are two Python files:

- `train_model.py`
- `predict_model.py`

`train_model.py` is the Python file that runs each time when the dataset is updated, receiving csv format dataset, returns model file in `.pkl` binary, stores in repository, for `predict_model.py` there is a `CaseReasoner` class, which will read the model `.pkl` binary, and receive the incoming vector, to produce the prediction. My team member's function will call the `CaseReasoner` to do further processing.

2. What have I learnt that is most useful?

I have learnt that, data define the maximum where the model can go, and the algorithms just trying to approach this maximum as much as it can, for Machine Reasoning, the critical part is how we define our data, and how we select the feature that can describe the phenomenon the most, in our case, our data coming from public, the limitation of public data is, it could be very limited to describe what we are researching, we need to refer to other studies and researches how to feature engineering the feature, one take away is: the model perceive part play the most important part, how much it can become knowledge is depends on it.

3. How can I apply this knowledge and skills in other situations or my workplace?

The main inspirations from this project is in my workplace, especially CyberSecurity domain, carefully select the sensor is critical, as the dengue predictor highly depends on the sensor data, to generate knowledge about it, similar rules apply to CyberSecurity, a carefully selected sensor to perceive software's behaviors highly influence the knowledge generation afterward.