Individual Project Report

Team Name: AI Predictors

Project Title: Intelligent Dengue Predictor

Team Member Name: Ng Kwee Boon

Team Member Matriculation Number: A0010308B

Team Member Contact (Mobile/Email): kweeboon.ng@u.nus.edu

1. My personal contribution to this project.

Project Ideas, discussion and material/data gathering:

- Discussion of different project ideas with team members and reached a common decision to pursue the Dengue Cases prediction project that would allow the team to practice and experience the different stages of the ML application development and implementation processes.
- Researched the Dengue information on the web and discussed with team members who also noted the vast resources on this topic and consolidated other information gathered by team member in a power point slide.
    - Example:
        - Dengue cases data from data.gov.sg
        - NEA dengue cases weekly reports available as pdf.
        - Past research paper on Dengue prediction techniques and approaches.
        - etc

Software integration:

- Validates and integrates the software solutions by bringing together the different software sub-modules together and integrate them to a unified single software system (web application).
- Collaborates and resolves any interface issues among the sub-modules with all team members.
- Checks project deliveries from team members for completeness.
- Manages the software project repository in GitHub and its layout.
- Installation and setup of the development environment and application execution.

Application Architecture/Design:

- Provides the software structure for the Dengue Cases Prediction web application in Python FLASK.
- Discussed with team members on the Python packages/libraries to use for web framework, visualization, etc and reach a common agreement.
- Tests the final application installation and execution and derived the installation and

execution guide.

2. What I have learnt that is most useful?

I have learnt the following:

a. At the start of the project, team members were having different project ideas and did not reach a common consensus on what to do for the project. The team wants to do something that is useful for the community and at the same time provides the opportunity for each of us to experience the development of a ML application/system. At the end we found that Dengue is a problem that still exists all over the world and it is still worth the efforts to try different ways to provide a solution to aid in the prevention on the spread of the virus.

   Therefore, it is important to identify the following at the start of the project:
   - What kind of problem I'm trying to solve?
   - What kind of question I'm trying to answer?

b. At first, we were concerned on the availability of the data for the Machine Learning and how we can gather the data. Fortunately, there are various resources from Data.gov.sg and NEA web sites that provide the information we needed. From the search for data, each of the team member learned the following:
   - Data gathering process from different sources.
     - NEA weekly Dengue Cases Report in pdf format – initially I was trying to use an OCR lib in Python to extract the data but realized that it is not perfect as not all the texts can be extracted completely. Moreover, after the extraction from the pdf, the data still need to be identified to determine the day and number of cases which is a more tedious process. At the end, each team member shared the burden to extract the information from different period manually and put them into csv data file format.
     - Weather data extraction from data.gov.sg web site – I learnt from a team member using a script to query the data.gov.sg web site and extracted the weather data over a period in json format. The knowledge was shared and was a useful tool to have for future project.

c. It is important to do versioning no matter how small the project might be:
   - ☐ With team member working on the implementation of the Application in Python and development of ML model in parallel and sometime overwriting each other files in the Git Repository. It is important to practice version control process and use the **GitHub** correctly. With proper versioning of the data, it allows reversal to an older version if necessary.
   - ☐ During the project, I was not familiar with the version control processes of **GitHub** and learnt about what is pull, push, fetch, merge, pull request, commit, main branch and development branches, etc. It will help in the long run for a bigger project and allow all team members to collaborate in the development of the project efficiently and effectively.

d. From the raw data that all the team member gathered, we discussed and studied the features available from the raw data. Some new input features were created from the existing raw data features and were used to determine the performance of the

training model. It was a repetitive and tedious process whereby some experiments were carried out, and the determination of the right ML algorithm to use, which created a kind of which come first problem. Fortunately, I have experienced team members who are familiar with these processes, and they guided and helped me in the understanding of the steps to use in the Model algorithm selection and training. E.g. As the data is time series, some ML algorithms that are suitable for time series data were used in the training and compared the results, before landing on using LightGBM as the ML algorithms to train the model. (In boosting, learners are trained one by one sequentially. Each subsequent one paying most of its attention to data points that were mis predicted by the previous one. And continue until satisfied with the results.)

e. In the team, we have people from diverse background collaborate and work together. I learnt that we could leverage on each other experiences and knowledges and collaborated to enable this project to completion and success.

f. Developing a web-based application using a simple Python web framework (FLASK). This ability allows me to be able to develop a web-based application from Front end to Back-end processing and will be a useful skill in the future for any web-based project that I might participate in. It initiated in me an interest and continued learning journey in web application development.

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

a. I can apply the knowledge of Machine Learning and Web based application development in future MTech project and in other workplace-initiated project for process improvement and optimization.

- ☐ As I am a Technical Project Lead, I often must gather and process a lot of data on project budget, project approved cost, project spending or hours used, project team performance in product development and delivery (e.g., number of defects vs number of deliveries), the trend/forecast on performance within the quarter, etc, and to generate project performance reports for management decision making and risk identification/mitigation.

- ☐ I would like to leverage the knowledge and skills that I have acquired in the project work, to initiate a project within my workplace that would help the project team and myself to process and handle the information mentioned above and to determine the trend accurately and efficiently and for easy management reporting and decision making.