WEEKLY INTERNSHIP REPORT

11rd July – 17th July

HUYNH THANH QUAN

# Plan of this week

* Collecting the total of insertion lines and deletion lines data from GitHub
* Normalizing the value of each feature
* Dividing training data into training set and test set
* Learning and testing once
* Writing weekly internship report
* Preparing for business trip in Numazu on 20th July
* Preparing for interview on 17th July

# Detail schedule of this week

## Wednesday

* + Discussed how to evaluate the commit has bug or not
  + Oyu-san had suspicion on script which might has bug
  + Reviewed the script for fetching commits from GitHub
  + Debugged and corrected the code to fetch all commits instead of fetching only commits that contain ‘FIX’ in message
  + Labelled non-bug commits that contain ‘fix’ in their messages
  + Normalized the range of each feature

## Thursday

* + Continue to normalize the range of each feature
  + Plotted the histograms and aggregated statistics for better insight information
  + Split dataset into training set, test set
  + Attended the Japanese training class

## Friday

* + Researched Precision, Recall and F-Score
  + Researched how to evaluate bug and non-bug commit
  + Learned to use Support Vector Machine (SVM) algorithms from Scikit-learn library
  + Built SVM model current data set (only has None label and non-bug label)

## Tuesday

* + Attended the Sprint Review and Retrospective
  + Import latest update into MongoDB from CSV files

# What I have learned

**Technical**

* Referencing how to evaluate commits as bug or non-bug from research paper
* Evaluating commits as bug or non-bug by tracing back the previous commits
* Plotting data visualizing (histograms, scatter plot, 2D, 3D)
* Normalizing data and data scaling
* Statistic information to get insight information
* Building support vector machine model to predict the data
* Learning precision, recall, f-score
* Debug and code refractor

**Work Schedule**

* Plan user stories in Agile Scrum development to accomplish the project
* Organize the stories of project into smaller applicable tasks
* Evaluate the time to spend, effort, human resources to finish the task
* Estimate how many percentage of task can be done in 1 Sprint (= 1 week)
* How to communicate or ask for helps when encountering the troubles

**Social**

* Japanese vocabulary (day of week, months, numbers, greetings and welcoming, how to buy food in convenience stores…)
* Use body languages to express the thoughts instead of using verbal communication
* Visit and explore Hokane and Akihabara
* Catch the bus and train
* Ask how to change train lines
* Organize the time (dinner, sleep, go shopping…)

# Plan of next week

* To evaluate AI system by current data set
* To create training set
* To evaluate AI system by latest data set
* Self-introduction presentation in July 18th
* HR interview in July 19th
* Business trip to Numazu July 20th

HUYNH THANH QUAN