**Weekly Report** ---*Friday, 26 March 2021*

Student Name: Siqi Sun
Student ID: a1752383

- **What I have done**

1. Apply requirement engineering for our project and elicited functional requirements with teammates.
2. Analysed non-functional requirements for the benchmark framework.
3. Get the knowledge of 'over-fitting' and 'under-fitting' theory about labelled data.
4. Communicated with team members and divide proposal into four parts and make sure everyone got assigned part.
5. Get familiar with the approach for Cross-Project Defect Prediction (CPDP).
6. Online research on the TextCNN and BiLSTM.
7. Uploaded Meeting Summaries (Wiki Pages) and Meeting Minutes (PDF version).

During our this-week group meeting, we discussed functional requirements and non-functional requirements for the benchmark framework and set the tone of the proposal. Our Proposal will be written in six parts:

- Introduce the research question -*Siqi Sun*
- Why do you think this research question is important? -*Siqi Sun*
- What does previous research tell you about your question? -*Siqi Sun*
- What data will you need to answer the research question? -*David Wu*
- How are you going to analyse this data? - *Tianlei Qi*
- Requirement and design(charts and description) - Fan Zhang

I chose the first three parts that mainly focuses on the introduction of research question and the importance of research question.

- **What I plan to do**

1. Learn how to extract features from labelled data bases and source code within the domain of data science.
2. Do online research on the previous, relevant topics and write the reflections as a preparation for proposal writings.

3. Hands on other open source relevant GitHub repositories and focus on deep learning models.

4. Research on knowledge transferring the data from model A to model B.

- **The issues I've encountered**

1. Don't know how to properly evaluate the performance of cross-domain vulnerability detection method.

Details: I've looked up research papers and find out the performance can be detected by existing tools. Yet, I'm still wondering how exactly these tools work. I may explored that in the next week.

2. Since we are planning to create a benchmark system, I'm not sure how to give an accurate mark for end-users' private system.

Details: My initial though is that by detecting the vulnerabilities of users' system, our benchmark system will provide a report and mark in based on the detecting frequency .Yet, these data are dynamic and updatable. It vary from time to time. Therefore, how to make sure the latest data get involved is our key question.

3. Don't know how to exactly protect end-user's source code and data.

Details: When it comes to users test their system, they may upload their source code to our benchmark system, we may need to encrypt their source code. However, as an open-source project contributor, the transparency of the whole project is essential. We may need to trade off these factors.

4. Don't know how to make sure the benchmark system to support very large-scale.

Details: If the system is very large-scale, it may cause the benchmark system breakdown. Thus, how to make benchmark system compatible with users' large-scale systems is very important.