CS3099 Junior Honours Software Project

Individual Report

Machine Learning Team 7

Kira Kim

University of
St Andrews

April 23, 2018

Summary

In this project, I have made several contributions from intra- and inter-group collaboration in design and implementation, team organisation and communication. My contributions allowed me to utilise both technical and group working skills.

Contributions

As the first scrum master of the project, I have made some guidelines from which the scrum master could manage a sprint and ways to keeping track of individual members progress by using tools available in GitLab. For example, in order to assist with the scrum sprints, milestones are used to assign individual tasks as issues. Considering that the collaboration had to be done remotely during holidays, this was an effective solution to update group members and tracking progress. For every milestone, a scrum master stored recordings of stand-up meetings which were also helpful for any members who were absent. We have continued using the tools and the general approach to the SCRUM process throughout.

I have also made a contribution to design decisions and implementations we made for our application. The idea such as the hierarchical object-oriented design of algorithms inherited from abstract class is reflected in our initial implementation design which helped progress towards the later design of pipelines. Implementation of random forest and receiver operating curves are done accordingly.

Later on, my area of implementation was around data preprocessing; from the motivation to improve our data cleaning stage, the methods are oriented towards the ability to handle various data types and data variables. Data.py is written to be flexible enough to accept data which to be used in various requests, including training and prediction. I have particularly communicated often with team members working on the server since various arguments can be called depending on the types of data to be cleaned in multiple training and prediction settings. Overall, the data cleaning stage made an integral part of the pipeline structure as the first step to be executed as shown in our pipeline diagram.

On behalf of the team, I was responsible for general communication with the Product Owner. We have asked questions whenever raised to the product owner, Peter, which helped clarify the user requirements: this was where our user stories were created. After completing the basic specification, the team has taken the requests from Peter to devise an advanced specification. Implementation and testing of these new functionalities were followed by the feedback and advice of the Product Owner which have been reflected our final design.

Conclusion

The project entailed various components. It has enabled me to broaden my knowledge of machine learning and data mining. Additionally, I was able to apply software development methodology and learn ways of working as a group and individually. Overall, I have thoroughly enjoyed the project with a potential to improve further given more time.