Individual Reflective Account
Kristian (Chris) Litsis
Student number:2076356
wg20498@bristol.ac.uk

In the second term, we were given a team mini-project as part of the teaching block. The intended work was to replicate a real Data Science project in an actual employment situation. Our team consisted of 4 members, and we were given Problem B. This project was a leading-edge data analytic for level 2 financial data provided by HSBC bank. There were several deliverables with the goal to construct a model that could create a simple trading strategy to generate profit.

Overall, the project was very exciting and with many challenges to be solved. I was very keen on this area of the subject because it is related to my master thesis that I selected, and therefore wanted to learn more about trading strategies and models. In addition, it would help me to understand the challenges better and find the best techniques to use in the future.

Firstly, we divided the work equally among all the team members. This was a little difficult in the beginning because we could not organise the work as we were very limited in accessing and reading the large data. Data preprocessing was required because the format of data was not a solid JSON as it contained some errors and empty values. Therefore, we run some scripts using regular expressions, and we converted them in a correct format that could be accessed. The size of the dataset was one of our major challenges because it was very large and could not be loaded into the memory. This made us possible to explore and use new technology, and we managed to use Spark to make all the processing and reading quicker. Therefore, we collected all the data together and significantly reduced the size by removing the outliers outside the two deviations from the mean that we detected when we studied the data distributions.

The next challenge was to understand the problem definition better. We had to learn about the marketing definitions, bids and asks in the Limited Order Book and different machine learning algorithms used for trading assets. This would help us to compare different techniques and identify advantages and their limitations. After some studies, we decided to create some new features that would help train the model with additional inputs. We looked in two main approaches; the first one consisted of a Long Short-term memory model and the second one a Reinforcement learning agent. The approach that we took was predicting the price of the given asset as a time-series forecast. Both of these models were new to use, so further studying was necessary, and this was a challenge taking into consideration that we had a very limited time. We investigated a lot of different papers and resources from the latest scientific models and, with the help of teaching assistants and the professors, managed to create a simple, realistic solutions.
My focus was to implement the Long short-term memory model and make predictions of the fluctuation of the price. The model made accurate predictions in the testing data, and now we had to implement some trading strategy to generate profit. A lot of evaluation was made in order to compare different implementations. We first compared the performance in errors made in the predictions of the bid and ask price. After the successful predictions, we implemented different strategies for automated trading. This enables us to learn how the traders think and what actions they make to generate profit. In the end, we managed to create some profit, and we were delighted with the results of the study.

Overall, I want to emphasise that this project was essential because it gave us an idea about how real work is in a big company and prepared us through learning and exploration. Each team member performed great, and this was made possible through crowdsourcing ideas and solutions. The collaboration would be much better if we were not in a pandemic as there were some minor problems in communications due to week internet connections and different time zones. However, we solved this by using an agile methodology (Scrum), and we created a kanban desk in Trello. Each deliverable was divided as a user story, and each member was assigned. Lastly, we created a Github repository for version control of the code to keep track of the changes and the progress of the project by analysing different commission versions.