Reflection

In the pre-processing of the data, we have learnt different ways to represent the vast amount of data provided, including plotting, statistical summary, and correlations. These methods provide us with medians to present our data more concisely and intuitively. This allows us to give interpretation and insight into the data that otherwise cannot be spotted by simply looking at the data. During the process, we have also improved our skills in extracting useful information from graphs, for example, skewness, outliers, and anything that spots out, while giving real-world explanations that can lead to those out-of-expectation results.

Throughout the processing of modelling, we have discovered many different algorithms that were previously not known. For instance, Lasso Regression, Ridge Regression, and Elastic Net Regression. This might look similar, but it has different properties in detail. This provides us with more options for fitting our model in the future. We have also learnt more about how to choose the values of hyperparameters. For example, we have discovered that there consists a rule of the number of nodes in the hidden layer equals to the two third of the input size. We can use this as a starting point for testing the number of hidden nodes and number of layers when we are fitting future datasets; we have also discovered the Akaike Information Criterion (AIC), which is a mathematical method of evaluating the model while taking overfitting into account.

However, in the report, we have only recorded the results of the model fitting but did not write much about the process of choosing algorithms and their hyperparameters. The record of model testing and hyperparameter tuning can be a very important note for us so that we can have expectations on how the result would change with respect to different hyperparameter values which will accelerate the speed of finding the best model.