CHAPTER 3
RESEARCH METHODOLOGY

3.1 Gethering Data From Kaggle

Dataset car price was taken from kaggle. The dataset format is CSV. This dataset has 19237 rows and 18 columns. Dataset has attributes including ID, price, manufacturer, model, production year, fuel type, and drive wheels. Then create sampler data. In the sampler data, determine the fixed proportion of data.

3.2 Processing Data

In the orange application, select data, then select CSV file import. After that, choose Select Transform, then Select Columns. Following that, target the attribute price in select columns. Then create a data sampler. In the data sampler, determine the fixed proportion of data. Then connect the imported CSV file to the data sampler to select columns.

3.3 Modelling

This is the process of determining which algorithm is used. Here, the algorithms used are K-Nearest Neighbors (KNN) and Random Forest (RF). Sample data is fed into the algorithm to get predictions. In the orange application, select kNN and Random Forest. Correlate training data with datasets that can be found in the Data sampler.

3.4 Result Reporting

In order to be able to evaluate and report results, processed data is entered into the predictions to be able to see comparisons between the model and the data. In the orange application, select evaluate then select Prediction. Then correlate Data sampler to Prediction. In edit links, change the connected link from data sample to remaining data.