

CHAPTER 5

IMPLEMENTATION AND RESULTS

5.1. Implementation

The first step is to enter the CSV file in the Orange application. The import option allows the user to inspect the contents of the raw data set fields. From the image below, we can see the car price data has 19,237 rows, 14 features, and 4 metas.

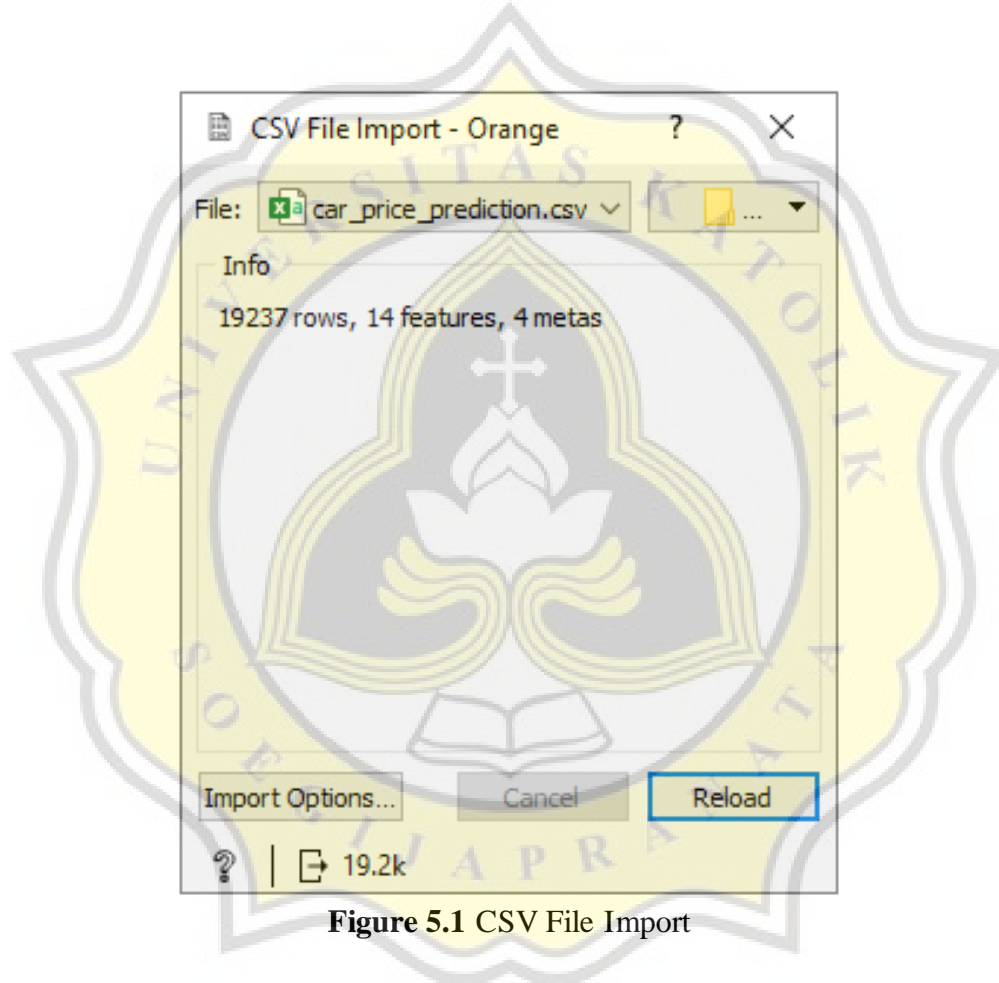


Figure 5.1 CSV File Import

After importing the CSV file, choose Select Transform. In the image below, it can be seen that there are 13 features consisting of ID, manufacturer, production year, category, leather interior, fuel type, cylinders, gear box type, drive wheels, doors, wheels, color, airbags, and column price that are the target for data processing. On the other hand, there are metas, which consist of levy, model, engine volume, and mileage.

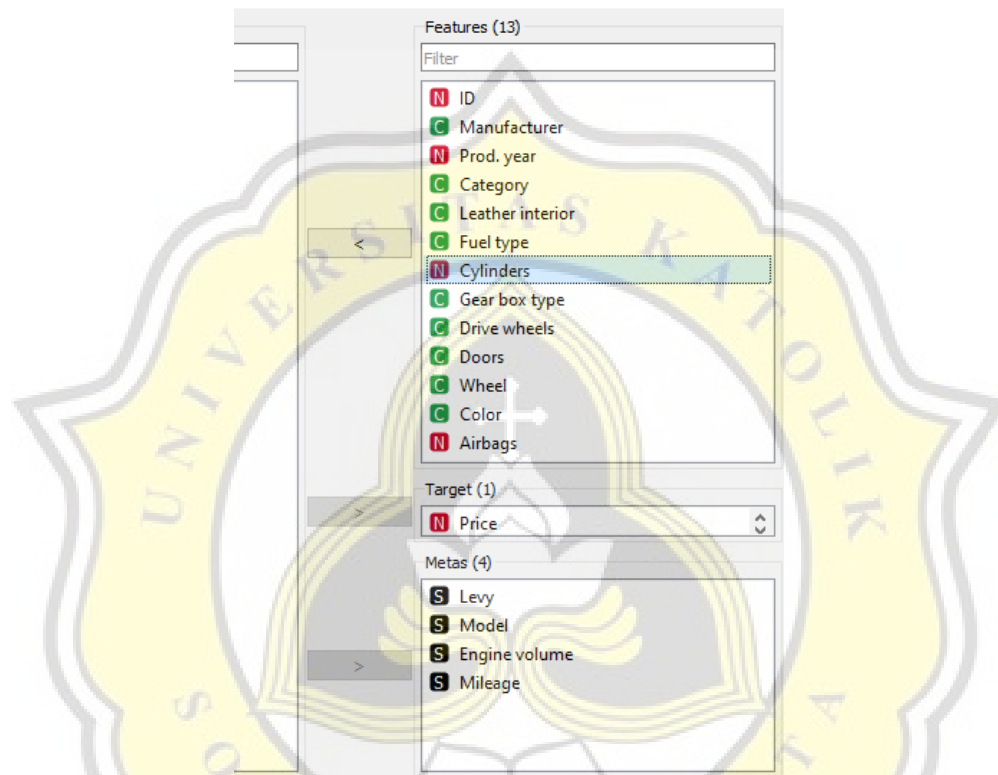


Figure 5.2 Select Target Price

After that, make a data sampler. Establish the preset percentage of data in the data sampler. In order to choose columns, attach the imported CSV file to the data sampler. The Fixed Ratio of Data in this study is set to 70%, 80%, and 90%.

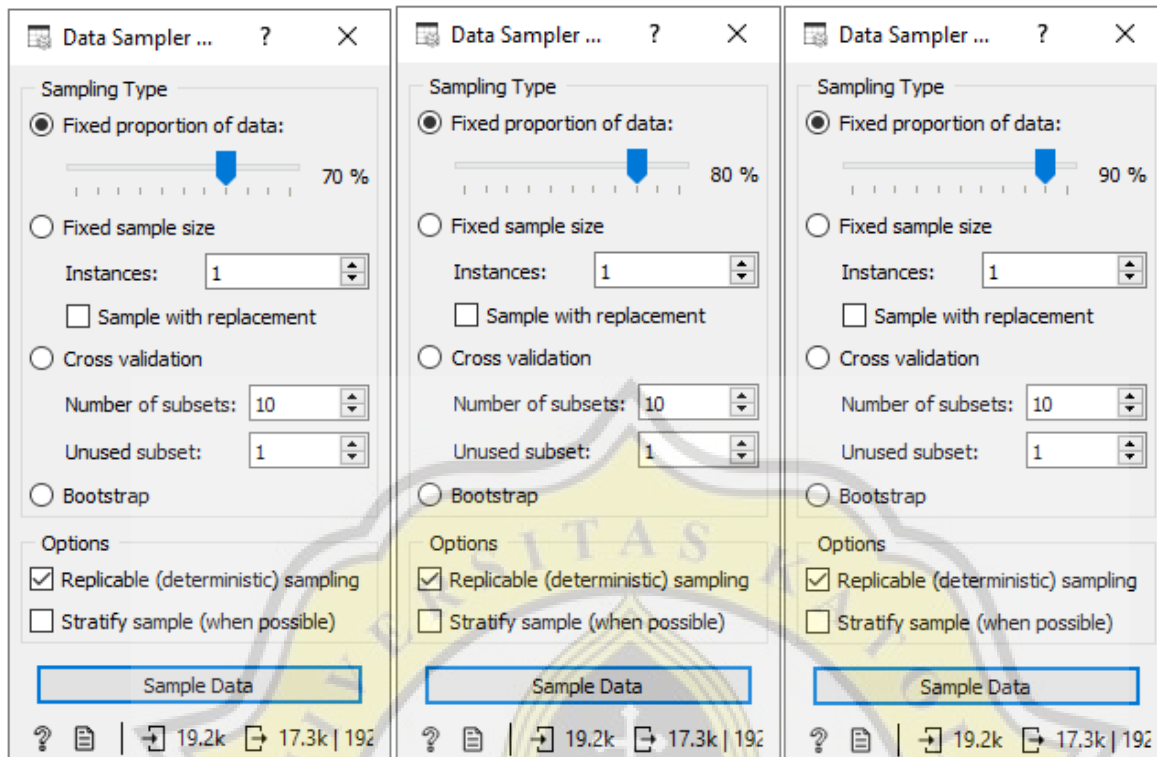


Figure 5.3 Set Ratio Data to 70%, 80%, and 90%

Then click the model, what we use here is a model that can do the regression process. this time we use Neural Network and KNN as a model to get training data. We will later use these two models as comparisons which are effective in predicting car prices.

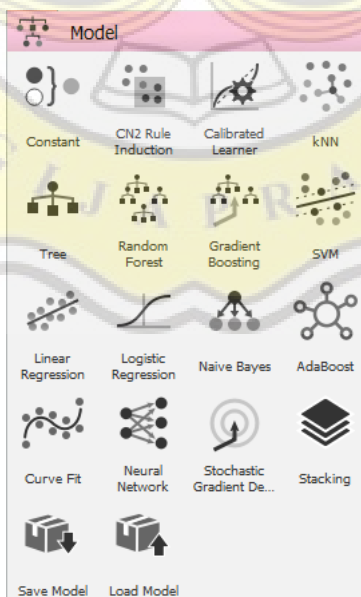


Figure 5.4 Select Model Neural Network and KNN

After selecting the model, in the Evaluation stage, the Orange application can display several results according to what we need. Because we aim to predict car prices, what we need are predictions. We can see the results in the evaluation section, the model with the lowest MSE value is the most effective model in predicting prices.

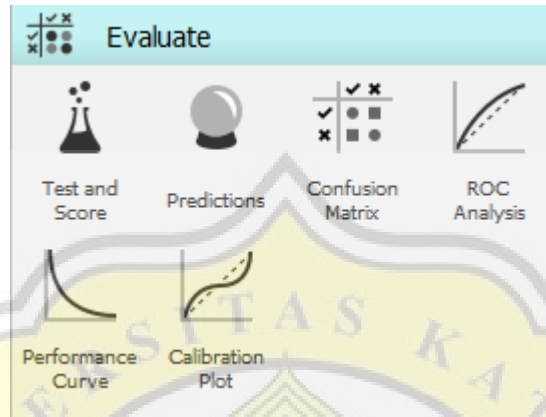


Figure 5.5 Select Prediction from Evaluate

Connect everything after completing the previously described setup to enable the use of the predictions. Make sure that nothing is missed in each process because it will affect the end result. Especially in the Data Sampler and Remaining data sections.

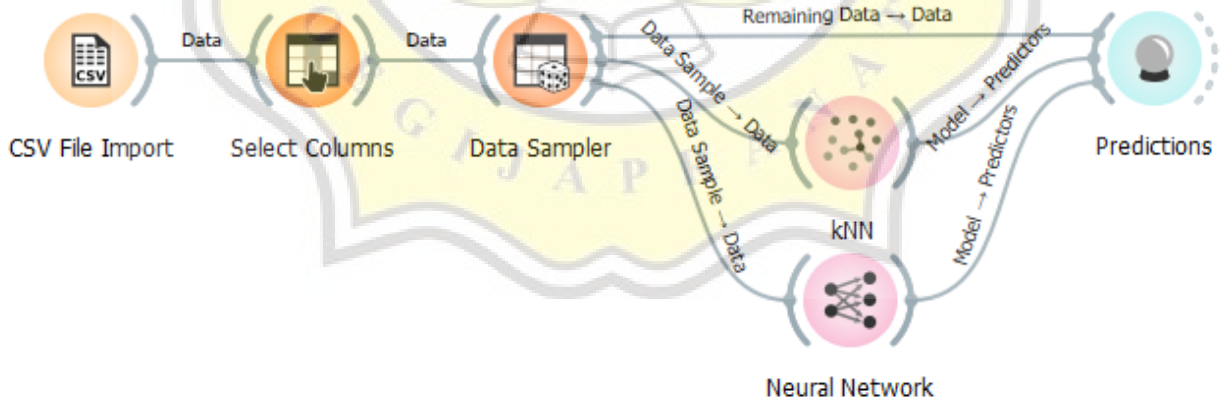


Figure 5.6 Connect all Steps

Another thing to do is to change the link between the Sample Data and Predictions. The Data Sample will be connected to Data by default; change it to remaining data being connected to Data. This needs to be done so that the prediction process can run properly.

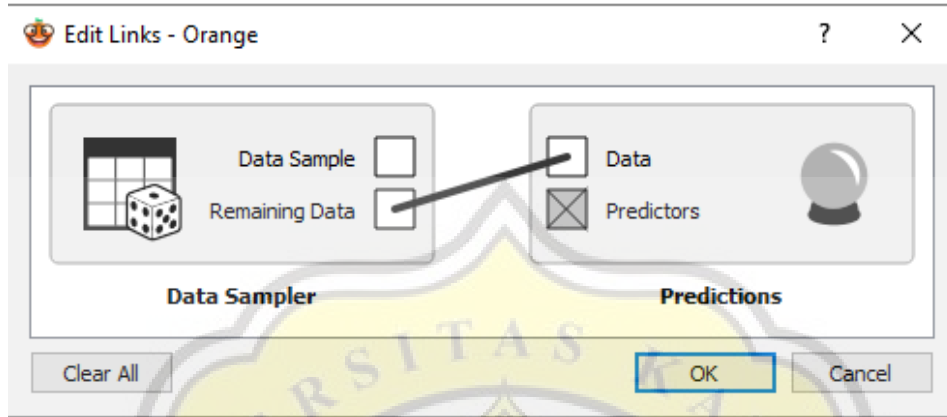


Figure 5.7 Connect Remaining Data to Data

5.2. Results

After conducting an experiment with 90% sampler data, the predicted results are as shown in the image below. The data will be processed using the SME, RMSE, and MAE formulas. From this formula, you can see which model has the lowest error value.

Neural Network	kNN	Price
24881	12654	16935
3174	12168	11604
4726	13817	7840
13800	2446	2195
5308	14445	706
33269	9046	47042
23501	44181	24709
42909	14552	27598
15674	6539	1333
15255	15367	15681

Figure 5.1 Car Price Prediction Results

$$MSE = \frac{1}{N} \sum_{i=0}^N (target - prediction)^2$$

$$RMSE = \sqrt{MSE}$$

$$MAE = \frac{1}{N} \sum_{i=0}^N |target - prediction|$$

Figure 5.2 MSE, RMSE, and MAE formulas

Model	MSE	RMSE	MAE
Neural Network	193344601.112	13904.841	9236.884
kNN	271856171.021	16488.061	9989.058

Figure 5.3 Calculation results with the MSE, RMSE, and MAE formulas

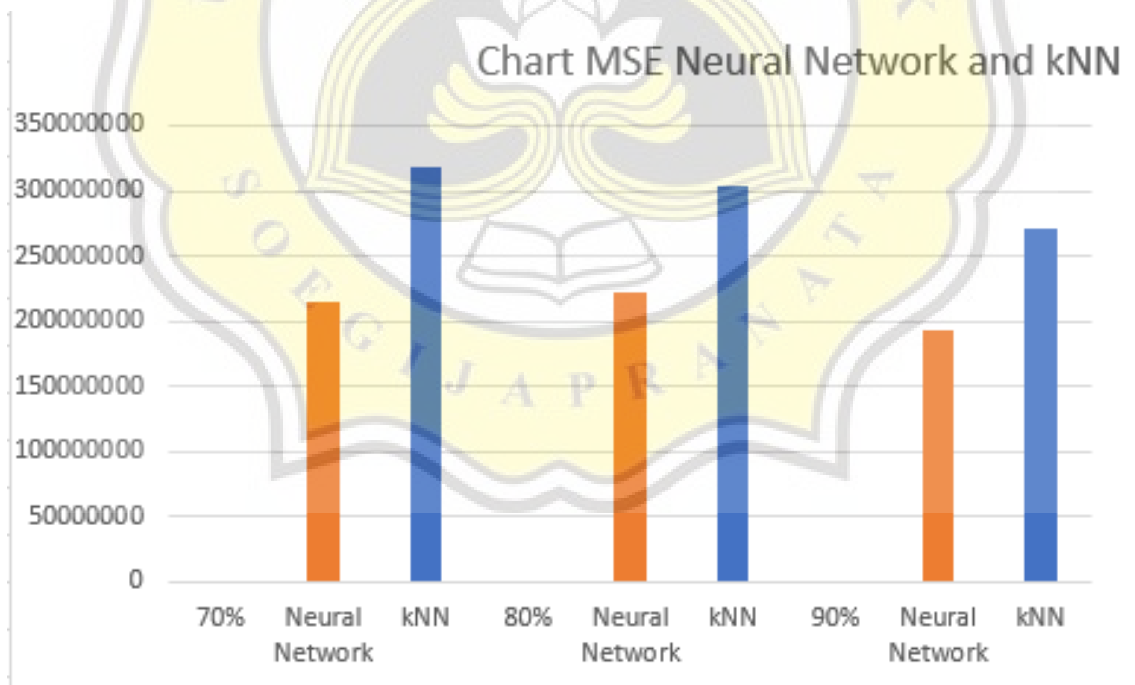


Figure 5.4 Result Prediction with 70%, 80%, and 90% data training.

Based on the results of the graph above, it can be concluded that the Neural Network model has a lower error value than the KNN mode. This conclusion was obtained after three trials with a comparison of the variance between the training data and the testing data. The comparison variances are 70%-30%, 80%-20%, 90%-10%. Of the three variants, the Neural Network produces the lowest value.

