



Figure 4.1: Flowchart Orange

The RMSE method will be used in this project to determine which algorithm is the best. The root mean square error (RMSE) is used to calculate a model's error rate when predicting a numerical value. The lower the RMSE value, the more accurate the model's prediction.

CHAPTER 5 IMPLEMENTATION AND RESULTS

5.1 Implementation

First, csv file inserted into orange. The contents of raw dataset columns can be viewed in the import option. CSV File Import is linked to Select Columns.

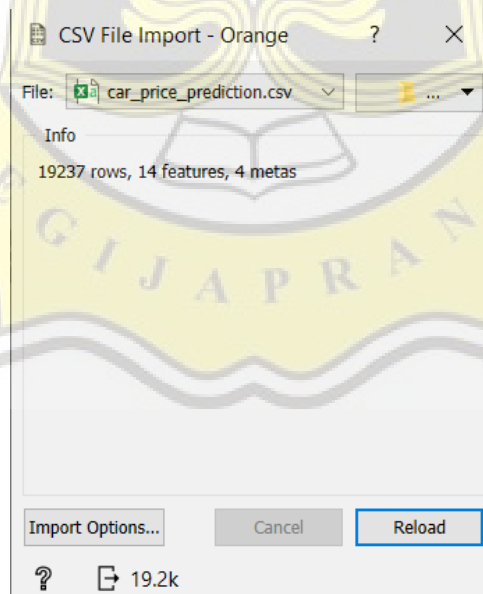


Figure 5.1: CSV file import

After input a csv file, Select Transform will be selected, followed by Select Columns. Select Columns is placed next to the CSV file. By default, the target will be empty. Following that, Select Columns attribute prices will be targeted.

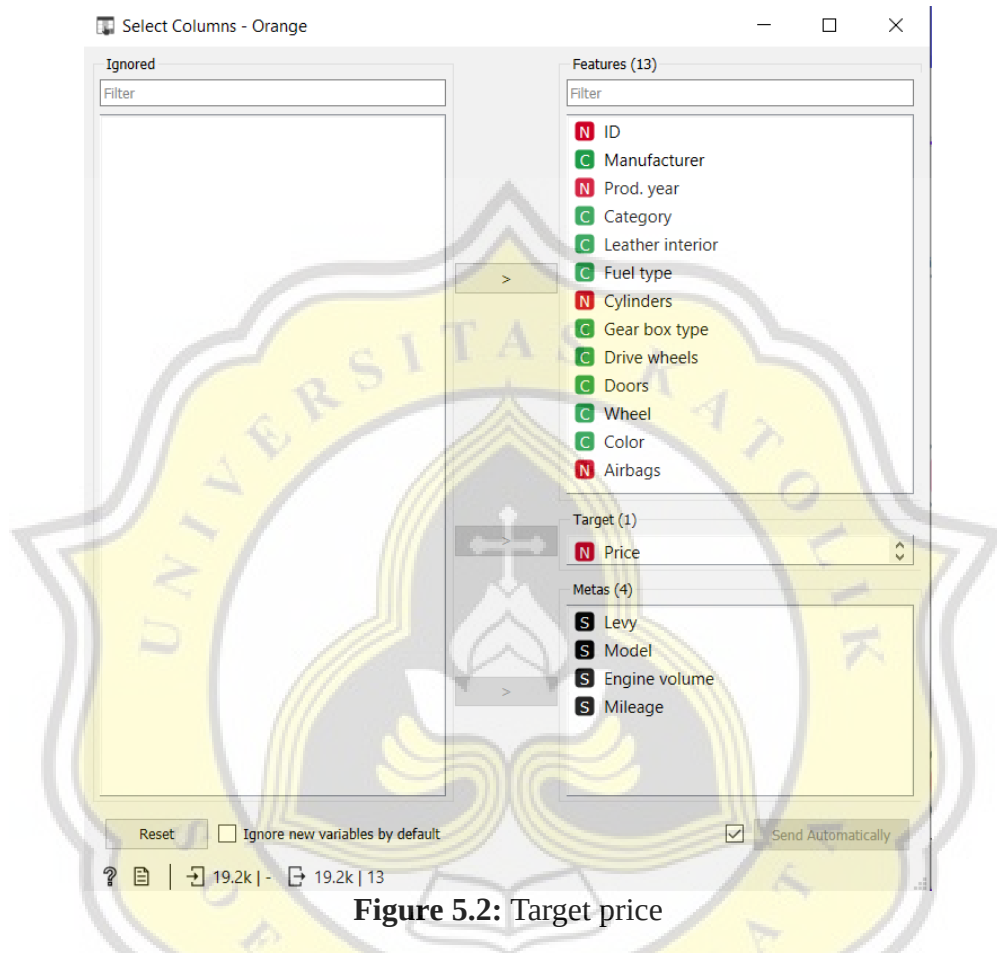


Figure 5.2: Target price

After that, Transform is opened, and then the data sampler is created. Select Columns are placed next to Sampler Data. In Data Sampler, the fixed proportion of data is determined by the author. Then the imported CSV file is linked to the sample data to select fields. In this study, the proportion of fixed data was set to 70%, 80%, and 90%.

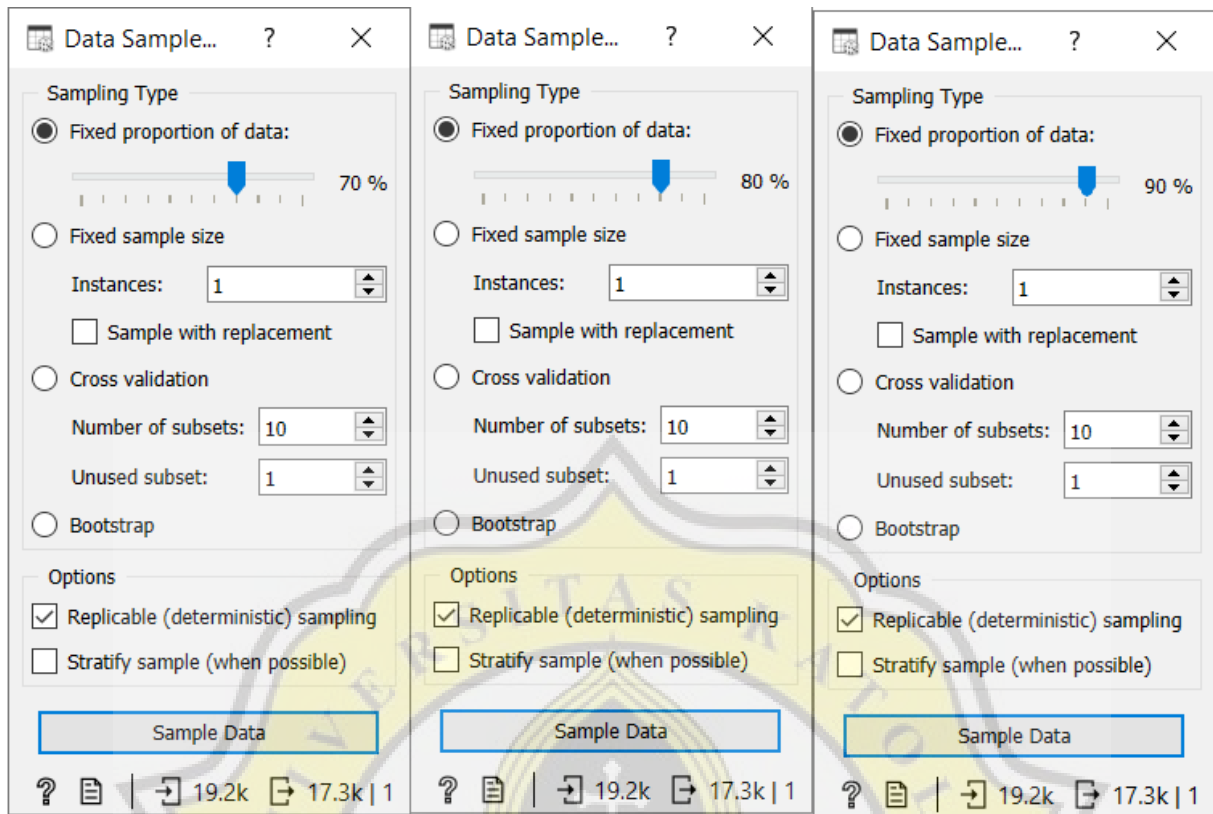


Figure 5.3: Fixed proportion of data

A model is chosen based on the input target. Following that, kNN and Random Forest are chosen. It is placed next to Data Sample and Data Sample connects to kNN and Random Forest.

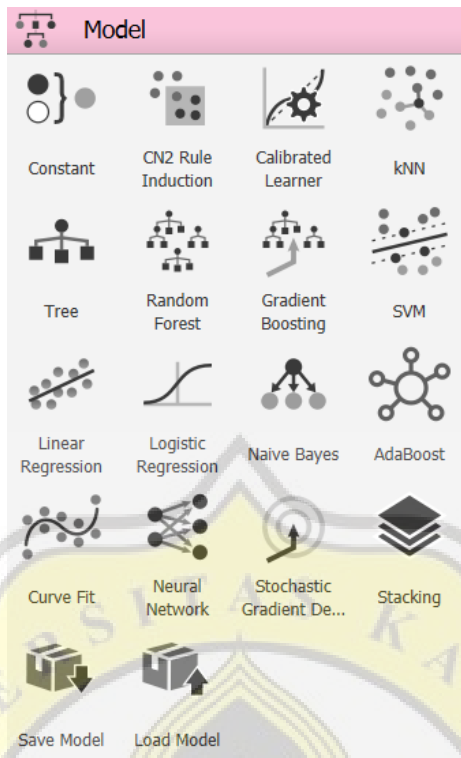


Figure 5.4: Model

After selecting the model, Evaluate is selected. In Evaluate need to choose Predictions. Predictions are placed next to kNN and Random Forest. KNN and Random Forest are then linked to predictions.

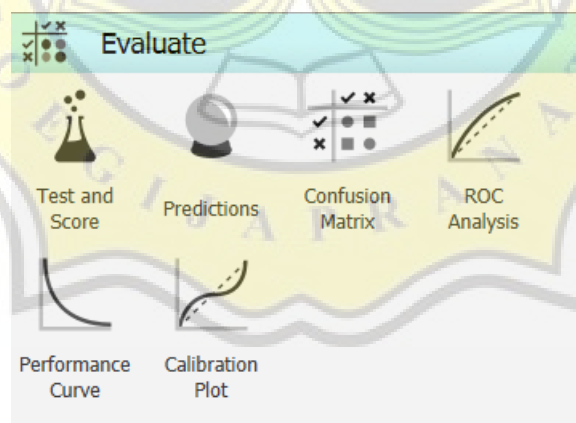


Figure 5.5: Predictions

To find out the target data set, Data Sampler is connected to Predictions. After completing the setup described earlier, everything is wired to enable predictive use. To see the comparison of results between the algorithm and the target, there is one more step.

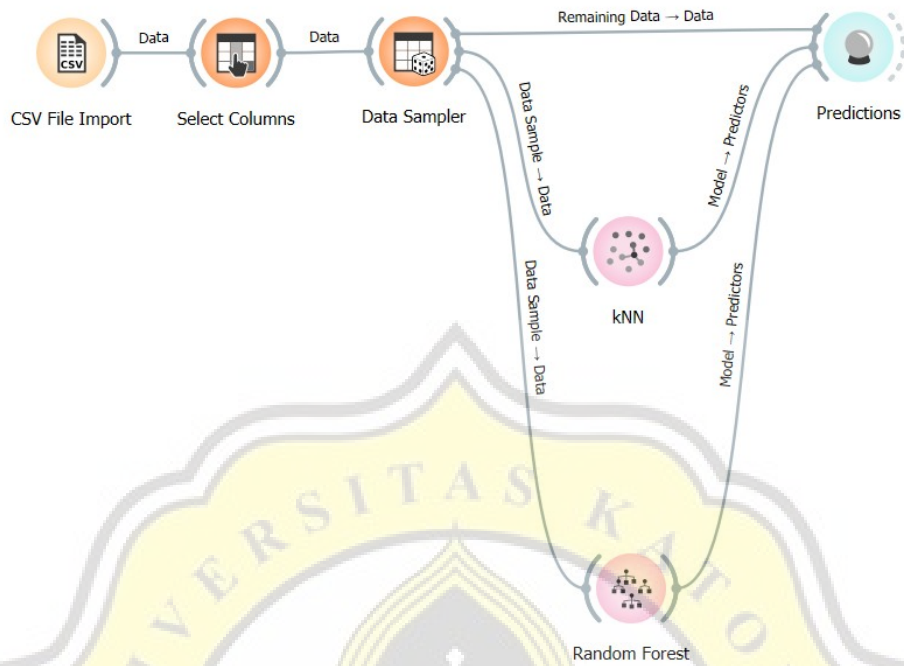


Figure 5.6: Connect all

One more thing to do, the connection between Data Sampler and Predictions must be changed. The default setting is that Sample Data will be connected to Data. It is changed to Remaining Data connected to Data.

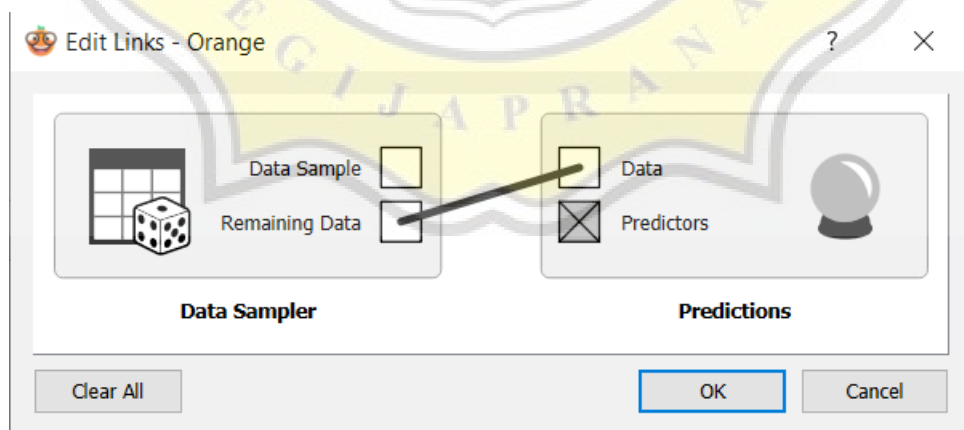


Figure 5.7: Connection between Data Sampler and Predictions

5.2 Results

In Predictions, we can see in detail which algorithm has the closest result to the target. Shown regression is changed from difference to (none). To know the average of algorithm, Show performance score will be checklisted.

	kNN	error	Random Forest	error	Price
1	12654	-4281	16699	-236	16935
2	12168	564	9010	-2594	11604
3	13817	5977	13981	6141	7840
4	2446	251	2768	573	2195
5	14445	13739	572	-134	706
6	9046	-379...	4042	-430...	47042
7	44181	19472	27195	2486	24709
8	14552	-130...	69591	41993	27598
9	6539	5206	6335	5002	1333
10	15367	-314	15320	-361	15681
11	1653	-181...	14173	-5584	19757
12	16276	-32	17197	889	16308
13	14180	-5264	24496	5052	19444
14	6492	2101	4200	-191	4391
15	6774	-129...	31640	11883	19757
16	34552	10552	17104	-6896	24000
17	2365	-4064	3667	-2762	6429
18	33744	-7966	40969	-741	41710
19	12500	-4000	13105	-3395	16500
20	12382	2290	10882	790	10092
21	20196	3888	6680	-9628	16308
22	27894	-936	27507	-1323	28830
23	33650	33180	4328	3858	470
24	15869	1907	17145	3183	13962
25	44227	24344	13680	-6203	19883
26	5488	5174	3594	3280	314
27	36989	17859	17626	-1504	19130
28	20717	4096	19709	3088	16621

Figure 5.8: Result of Predictions

The RMSE KNN results 70% are 17819.453, 80% are 17437.003, and 90% are 16488.061. The results of the RMSE Random Forest 70% is 142286.83, 80% is 127456.71, and 90% is 85347.085. The results of 70%, 80%, and 90% of the data show that the K-Nearest

Neighbors (KNN) algorithm has predictions that are closer to the target than the Random Forest algorithm.

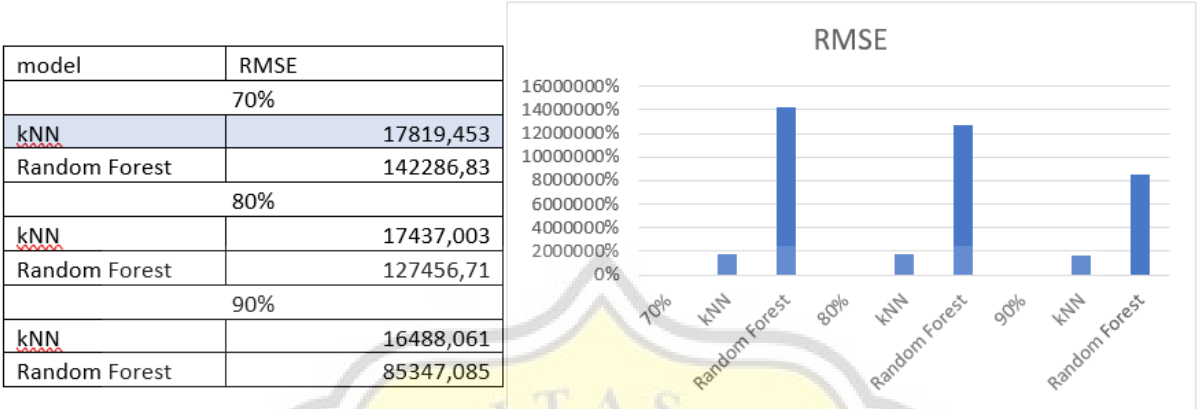


Figure 5.9: Result

