CHAPTER 5
IMPLEMENTATION AND RESULTS

5.1. AdaBoost Steps

The AdaBoost method has several prediction outputs on Orange data mining. In this first step, it can produce MSE, RMSE, MAE, and R2 outputs. The following are the steps in data processing in Orange data mining.

5.1.1. Importing CSV data

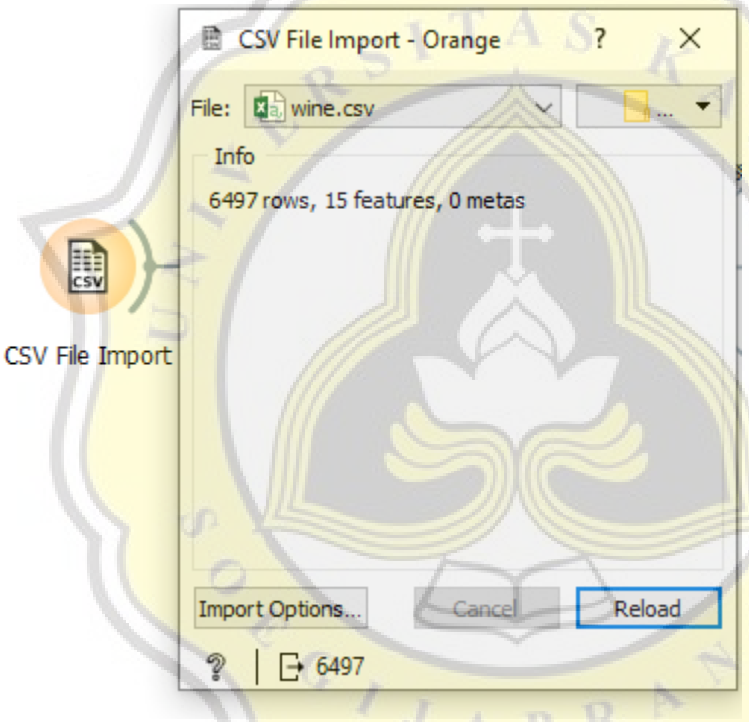

Figure 5.1 Import CSV AdaBoost

In this first step, the main thing to do is to import CSV data using the Orange widget. This data import is taken from the kaggle.com site or can be obtained from other sites. The data I took has 6000 datasets and 12 attributes.

5.1.2. Selecting Attribute

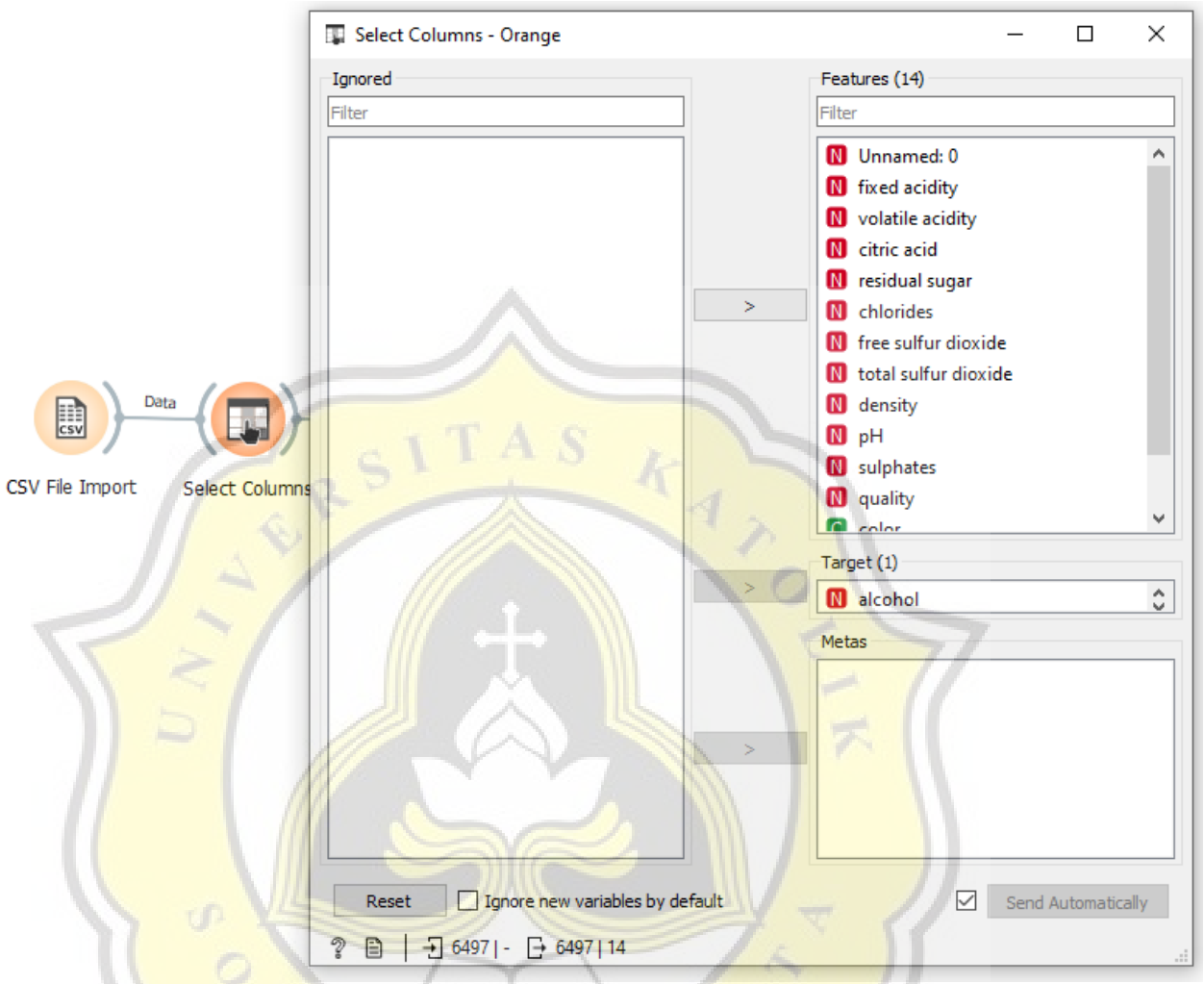

Figure 5.2 Set target AdaBoost

In this step, select the target attribute variable by selecting the target data to be used in the data column. In this step, choosing the target variable is very important because the selected attribute has a value that is modeled and predicted by other attributes. The target I chose is alcohol as the target variable.

5.1.3. Data Sampler

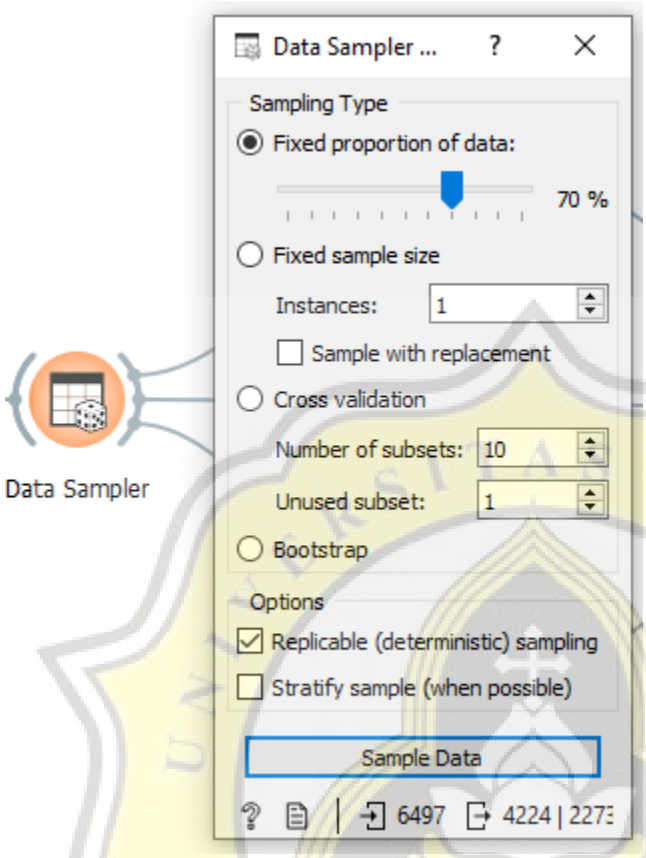

Figure 5.3 Data Sampler AdaBoost

In the next step, use the data sampler widget in the orange data mining application. This widget can divide into several data, namely testing data and training data. In this data sampler tool, the proportion of data used is 70% data. Training data has 4200 data and 1800 becomes testing data or becomes the remaining data.

5.1.4. Making Prediction

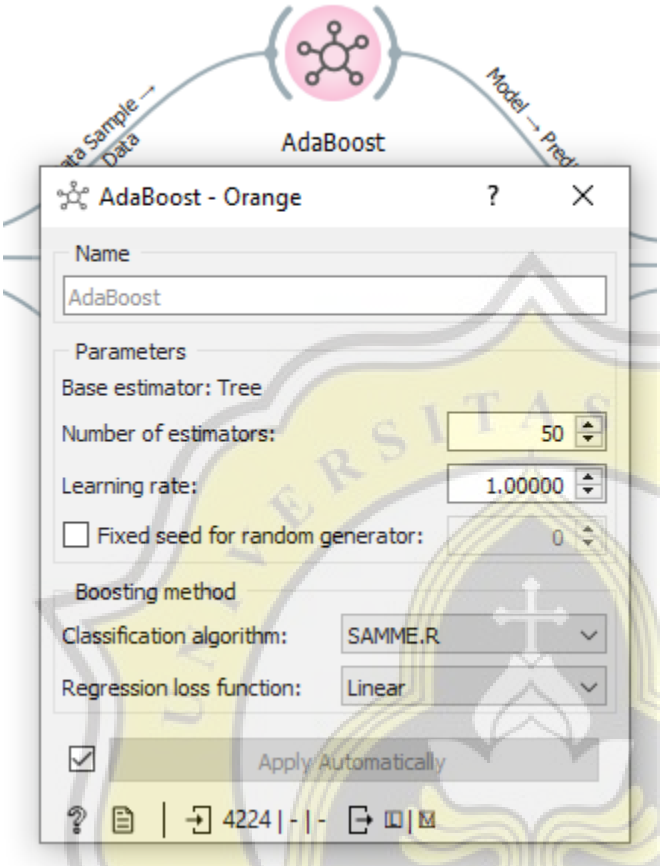

Figure 5.4 Making Prediction AdaBoost

In the next step 4200, training data will be calculated and then tested using widgets in Orange data mining. In this step, the AdaBoost method is used with parameters, a base estimator tree, a number of estimators of 50, and a learning rate of 1.00000.

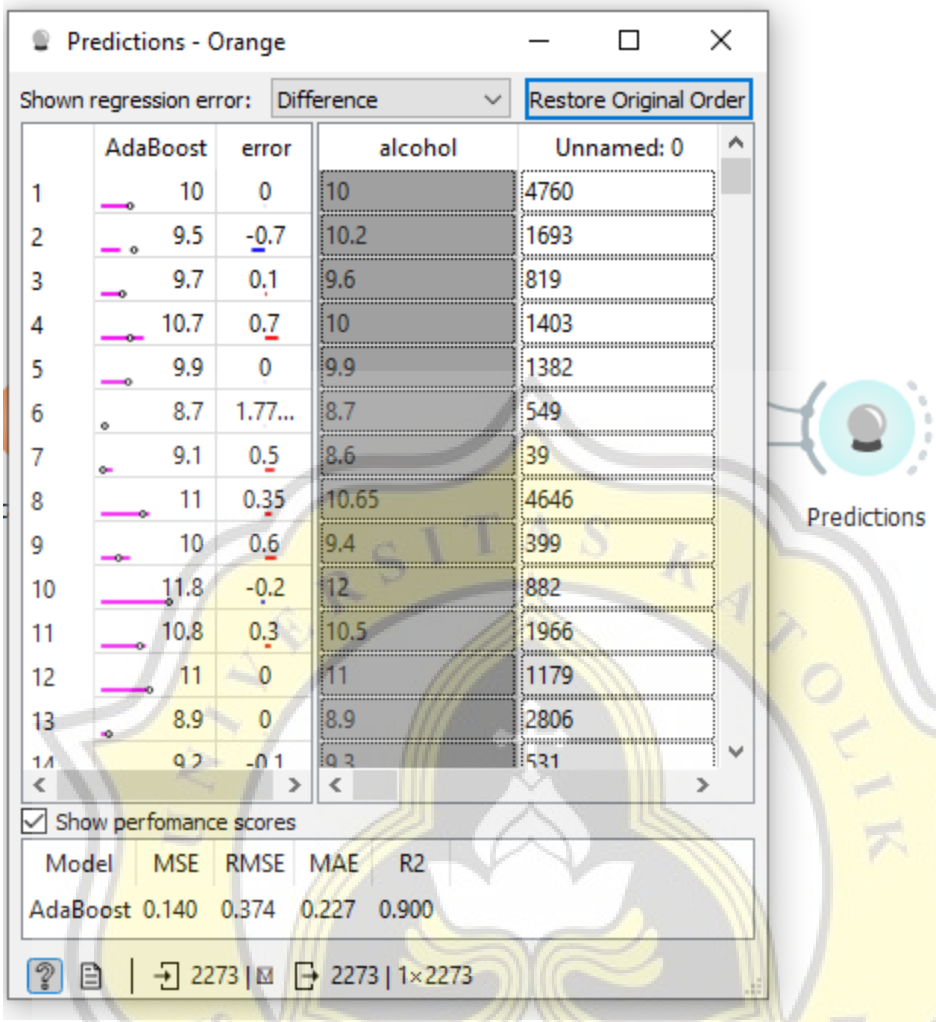

Figure 5.5 Output AdaBoost

Then from the use of orange data mining is to make predictions and the prediction output is MSE,RMSE, MAE, R2.

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

$$RMSE = \sqrt{MSE} \quad (2)$$

$$MAE = \frac{1}{n}\sum_{i=0}^{n}|target - prediction| \quad (3)$$

Indeks	explanation
MSE	Mean Square Error
N	Total Data
RMSE	Root mean Square error
MAE	Mean absolute error
∑	Summation

5.2. Random Forest Steps

The Random Forest method also has the same steps as the previous method. This method also makes prediction output and gets MSE, RMSE, MAE, and R2 values. Below are the steps in making output on Orange data mining.

5.2.1. Importing CSV

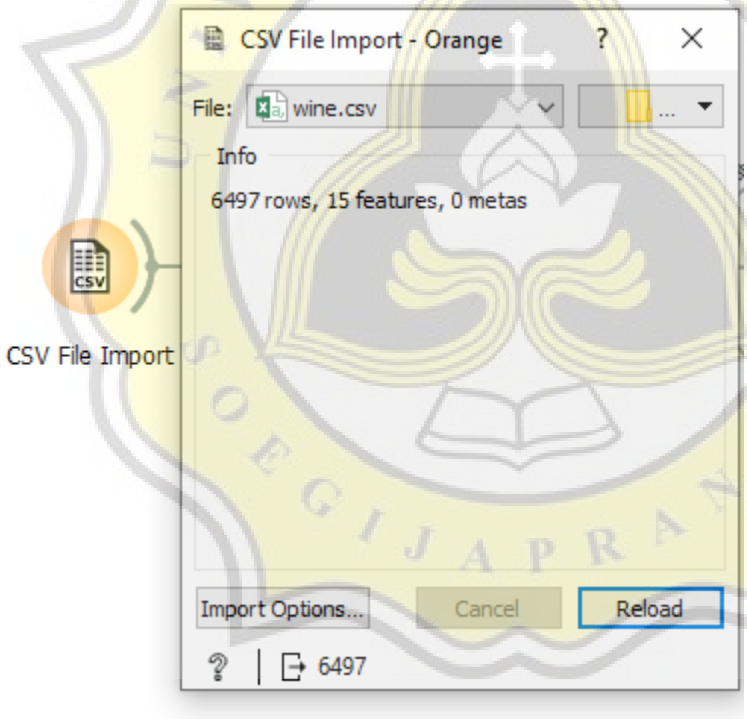

Figure 5.6 Import CSV Random Forest

Just like the previous method, the first step in processing data with the random forest method is to import data with CSV files using tools in Orange data mining. In the CSV file, there are 6000 datasets and 12 attributes. This CSV file can be obtained on the kaggle.com site or can also be obtained from other sources.

5.2.2. Selecting Attribut

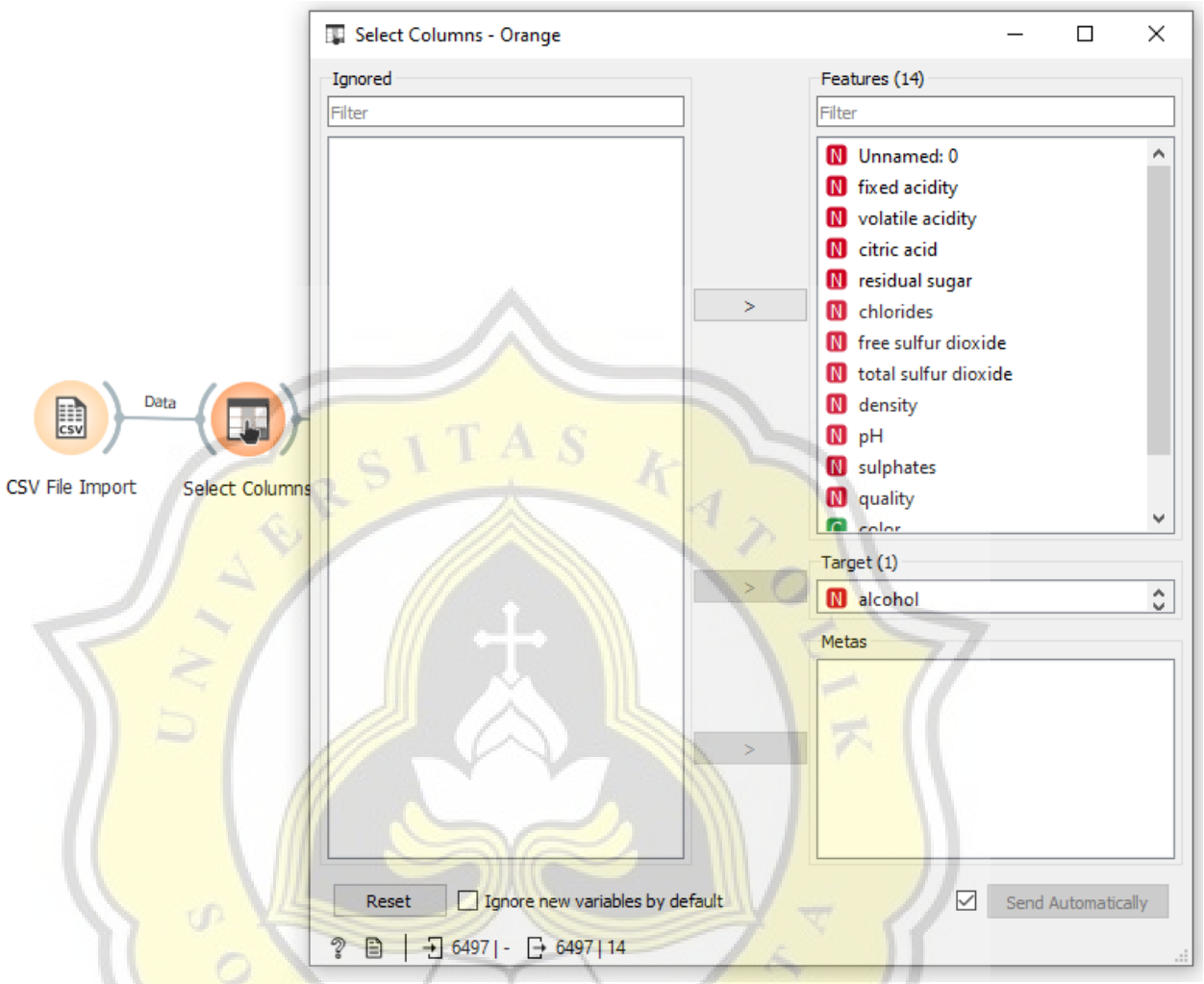

Figure 5.7 Set target Random Forest

After the data has been successfully imported, the next step is to select the target variable. The target variable to be used in this method is the same as the target variable used in the previous method, namely alcohol as the target selection in this method.

5.2.3. Data Sampler

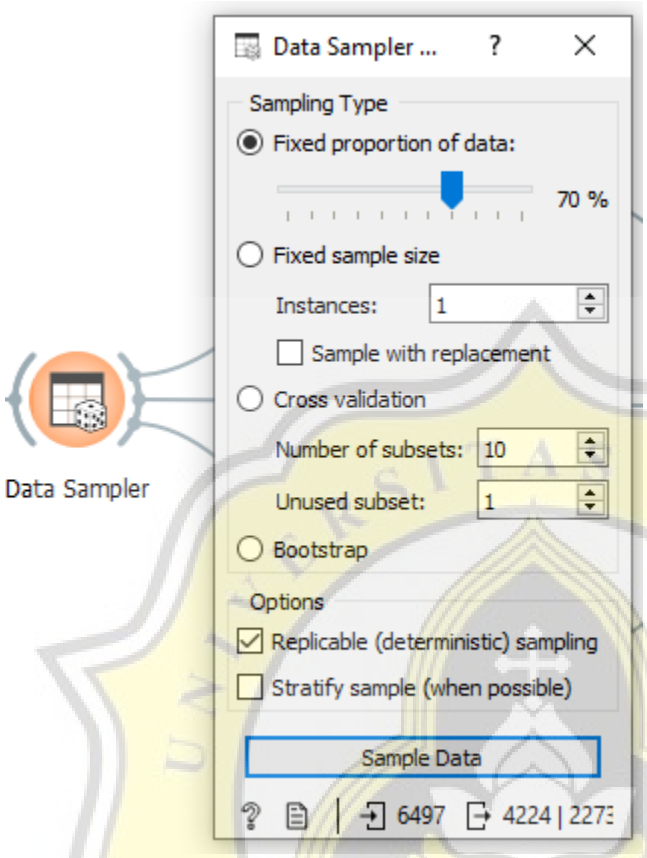

Figure 5.8 Data Sampler Random Forest

Similar to the AdaBoost method step, this step uses the data sampler widget in the Orange data mining application. This widget functions to divide two types of data, namely testing data and training data. The proportion of data used is 70% data. The training data has 4200 training data and 1800 testing data.

5.2.4. Making Prediction

Figure 5.9 Prediction Random Forest

The next step after performing the steps on the data sampler is to make predictions. In this step, 4200 training data will be trained and processed and will be tested using the random forest method. In this method, data training will be carried out in the form of configuring the number of trees, training replication, and also class balancing.

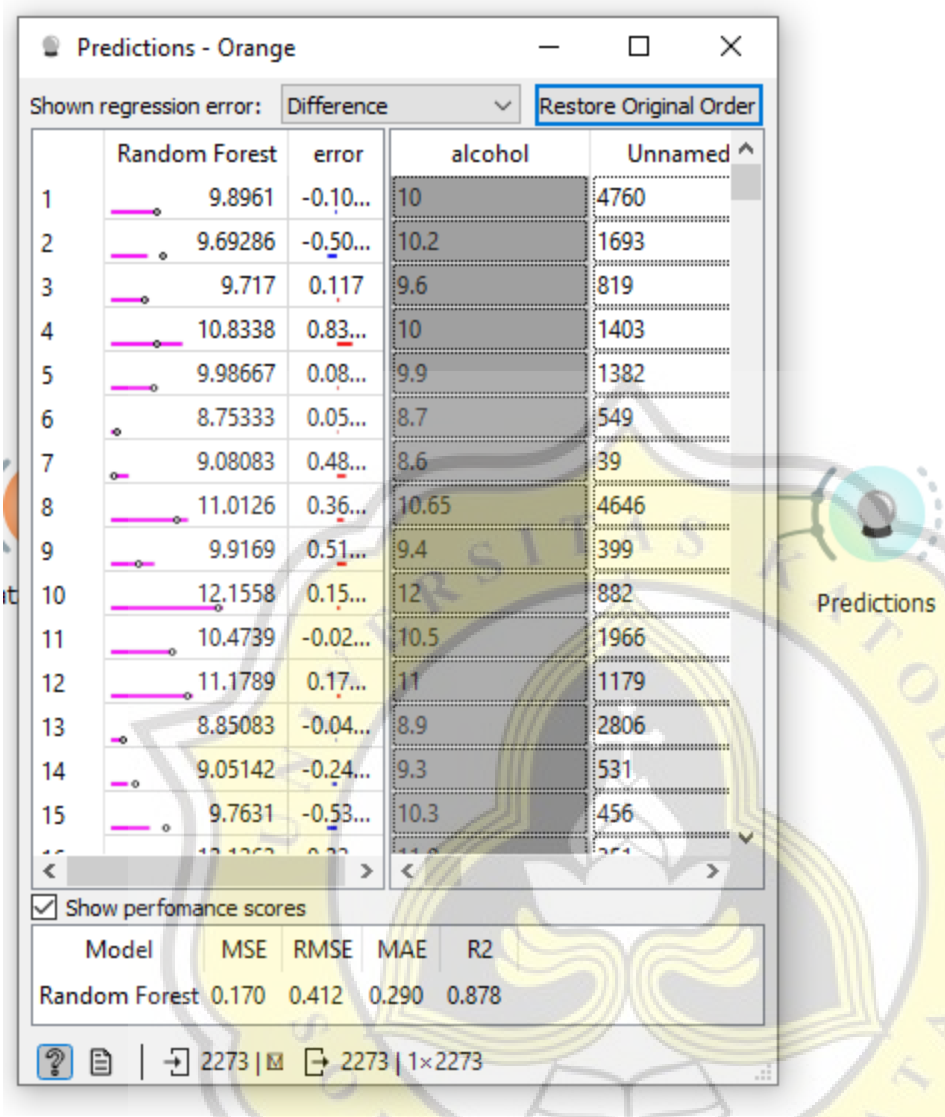

Figure 5.10 Output Random Forest

Then from the use of orange data mining is to make predictions and the prediction output is MSE, RMSE, MAE, R2. The following is an example of the MSE, RMSE, MAE formulas.

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

$$RMSE = \sqrt{MSE} \quad (2)$$

$$MAE = \frac{1}{n}\sum_{i=0}^{n}|target - prediction| \quad (3)$$

Indeks	explanation
MSE	Mean Square Error
N	Total Data
RMSE	Root mean Square error
MAE	Mean absolute error
∑	Summation

5.3. Results

After doing research using the Orange data mining application and getting the output results of the two methods above. The author compares the two methods above by doing different configurations and with total sampler data for testing, namely 60%, 70%, 80%, and 90% to get fewer errors and the accuracy of the two methods above.

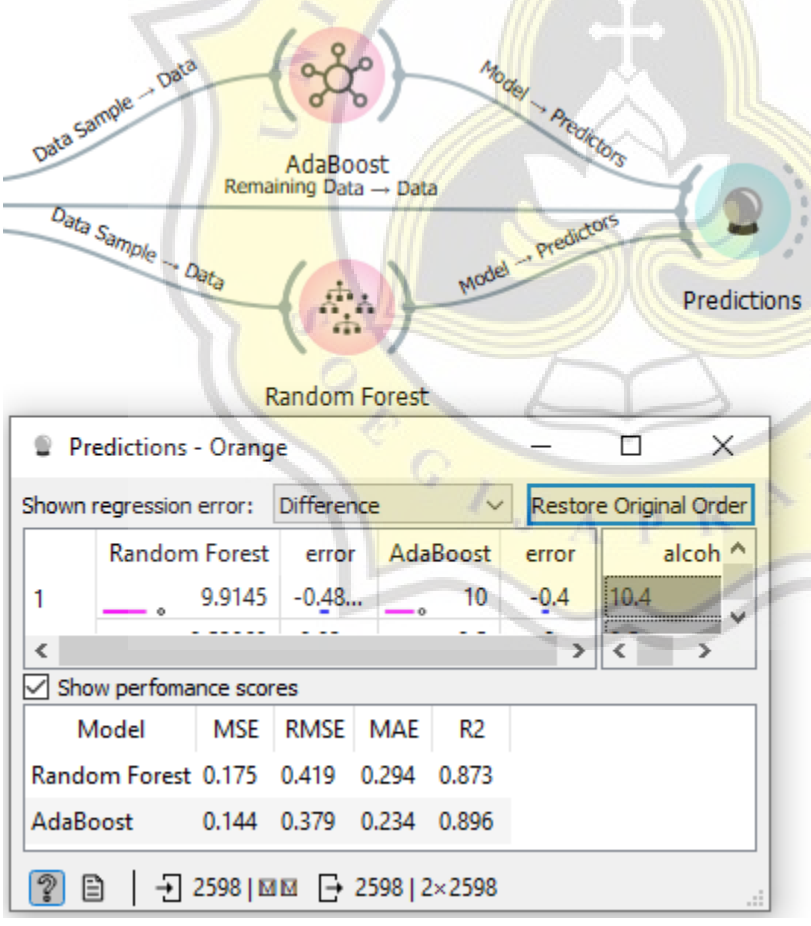

Figure 5.11 Output AdaBoost and Random Forest

DATA SAMPLER				
60%	***MSE***	***RMSE***	***MAE***	***R2***
AdaBoost	0.144	0.379	0.234	0.896
Random Forest	0.180	0.424	0.303	0.870
70%	***MSE***	***RMSE***	***MAE***	***R2***
AdaBoost	0.136	0.368	0.218	0.903
Random Forest	0.177	0.421	0.293	0.874
80%	***MSE***	***RMSE***	***MAE***	***R2***
AdaBoost	0.115	0.339	0.196	0.919
Random Forest	0.151	0.389	0.267	0.894
90%	***MSE***	***RMSE***	***MAE***	***R2***
AdaBoost	0.112	0.335	0.191	0.920
Random Forest	0.151	0.388	0.262	0.892

Table 5.1. Data Sampler

The table above is a comparison of sampler data, in this study comparing 60%, 70%, 80%, and 90% of total sampler data. And from the comparison of the total sampler data, it has MSE, RMSE, MAE, and R2 outputs. in these outputs, it can be determined that the less the output value, the better the method used.

Figure 5.12 Graph comparison

In the MSE comparison graph 60%, 70%, 80%, and 90% of the AdaBoost method has the least output compared to Random Forest. In this data sampler, it means that the best output value is AdaBoost compared to RandomForest on each data tested on each data sample.

Table 5.2. Result

Sample Data	Model	RMSE	MAE
60%	AdaBoost	0.379	0.234
	Random Forest	0.424	0.303
70%	AdaBoost	0.368	0.218
	Random Forest	0.421	0.293
80%	AdaBoost	0.339	0.196
	Random Forest	0.389	0.267
90%	AdaBoost	0.335	0.191
	Random Forest	0.388	0.262

The table above is a table to compare the value of the AdaBoost and Random Forest outputs. And produces the output of RMSE and MAE prediction values. And in this sample data, the data used is 60%, 70%, 80%, and 90%.

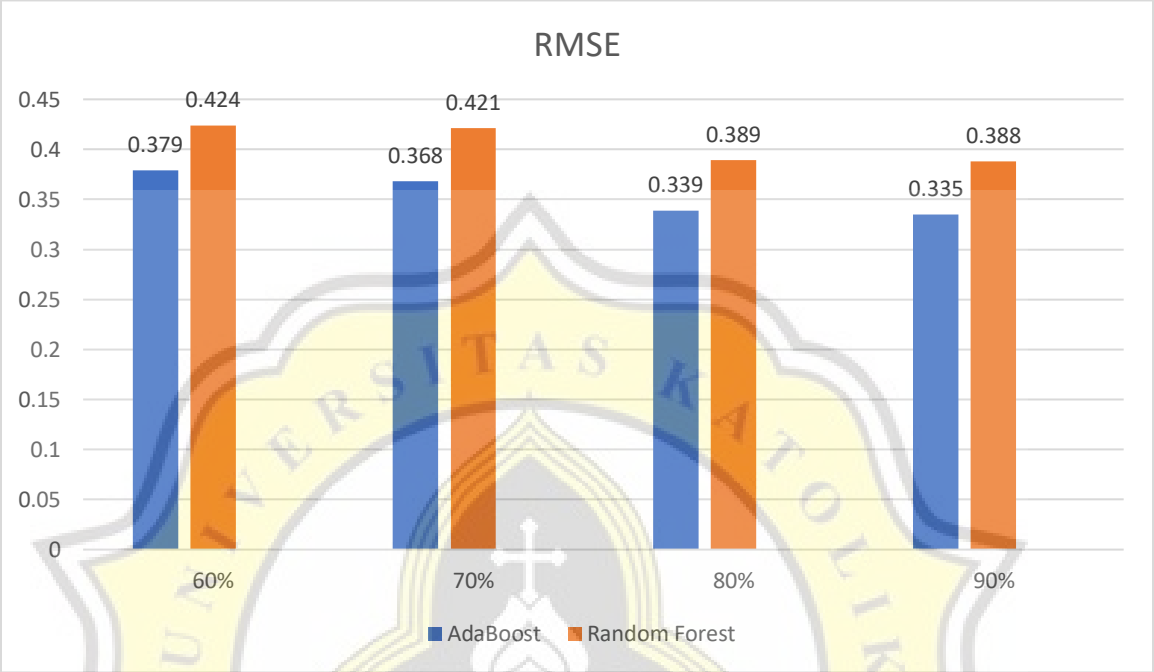

Figure 5.13 RMSE graph

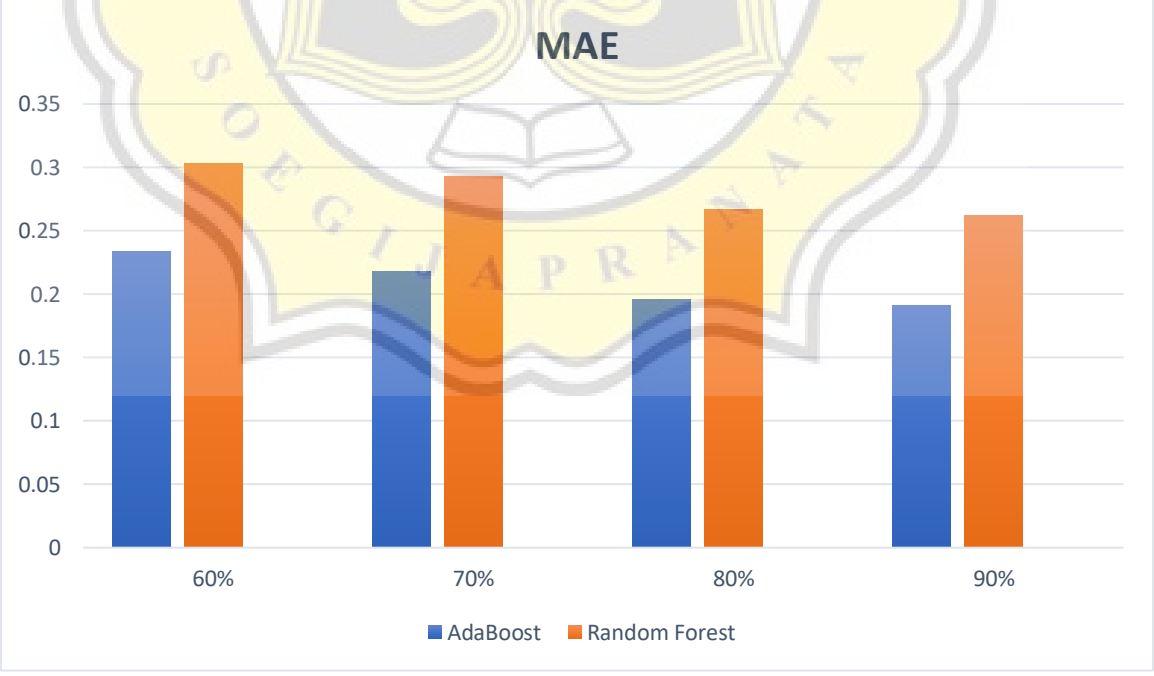

Figure 5.14 MAE graph

Based on the chart above by predicting the output of MSE, RMSE, MAE, and R2 the AdaBoost method has less error rate than the Random Forest method. In this case, the most suitable method used in this research is AdaBoost. The less the output number in this method, the more suitable this method is to be used in sample data.

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

$$RMSE = \sqrt{MSE} \quad (2)$$

$$MAE = \frac{1}{n}\sum_{i=0}^{n}|target - prediction| \quad (3)$$

Random Forest functions in orange data mining to create a decision tree or can be called a decision tree. The Random Forest algorithm also increases randomness in the data while growing the tree and also combining it to get accurate results. And also of course using too much, will affect the level of accuracy obtained and become a more optimal result.

Random Forest is a classification algorithm. Random Forest works by building decision trees and combining them to get more accurate results and also stable results. This set of decision trees is trained with the bagging method. Bagging is a method to improve the overall result. It increases the randomness of the data to grow the tree. And this results in better output data.