CHAPTER 5
IMPLEMENTATION AND RESULTS

5.1. Implementation

In the implementation of this section, Orange will process data by going through several stages. Namely by explaining how data is processed, from sample data to actual data that passes through several testing processes. And shows different results in processing with the KNN and SVM algorithms. This process aims to find out how the results of the predictions are processed by the algorithm to show the level of accuracy in each of the algorithms.

This explanation will explain the evaluation metrics for classification, namely accuracy (CA), precision, recall and f1-score in the Confusion Matrix. The following explanation such as accuracy is defined as the degree of closeness between the predicted value and the actual value. Then Precision describes the accuracy between the requested data and the predicted results given. Meanwhile, Recall is a model of success in retrieving information. And finally the F1-score is about comparing the average precision and recall which is weighted, so if accuracy is right as a reference for performance but if our dataset has an approaching number of false negatives and false positives. However, if it is not close, we should use the f1-score as a reference.

In this process Orange will carry out stages such as using repeat train tests as much as 2x with 25% data set testing and 75% data testing. Then repeated 2x with 50% training data set and 50% training data. Then bring up the calculation results of accuracy, precision, recall and f1-score which measures the performance of each algorithm. After Orange will display the results of the predictions on the confusion matrix to show the results of right and wrong as an implementation

5.2. K-NN Result Prediction

According to the implementation's description, K-NN goes through multiple testing phases. The testing procedure employs 2x repeat train tests with training data sets of 25%, 50%, and 75% to generate accuracy, precision, recall, and f1-score findings. The outcomes are shown in the Test and Score section. The following technique determines the real and false outcomes in the confusion matrix.

On the Test and Score outcomes, the following is a random repeat train test data process 2x with 25% training data and 75% testing processed with K-NN.

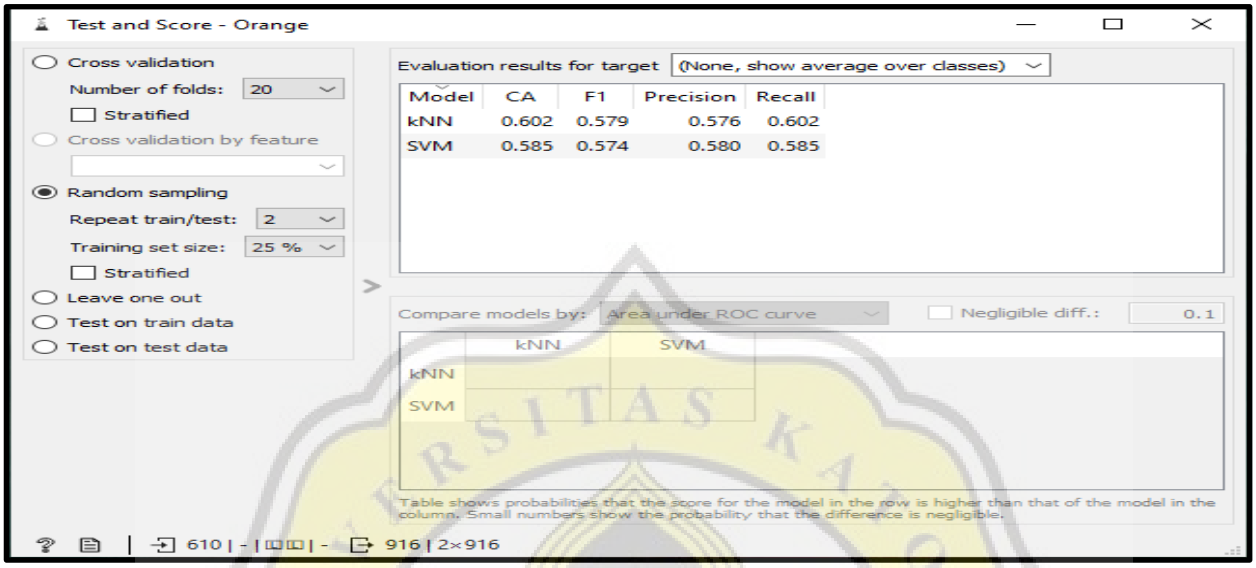

Figure 5.2 1 Test and Score 25 % Data Training and 75% Data Testing

The table shows the results of testing data processing using 25% training data and 75% testing data then addressing the value:

Table 5.2 1 Test and Score 25 % Data Training and 75% Data Testing

CA	f1-score	Precision	Recall
60 %	57 %	57%	60%

Then in the confusion matrix that has been carried out 2x the repeat train test process shows the correct and incorrect data values as follows:

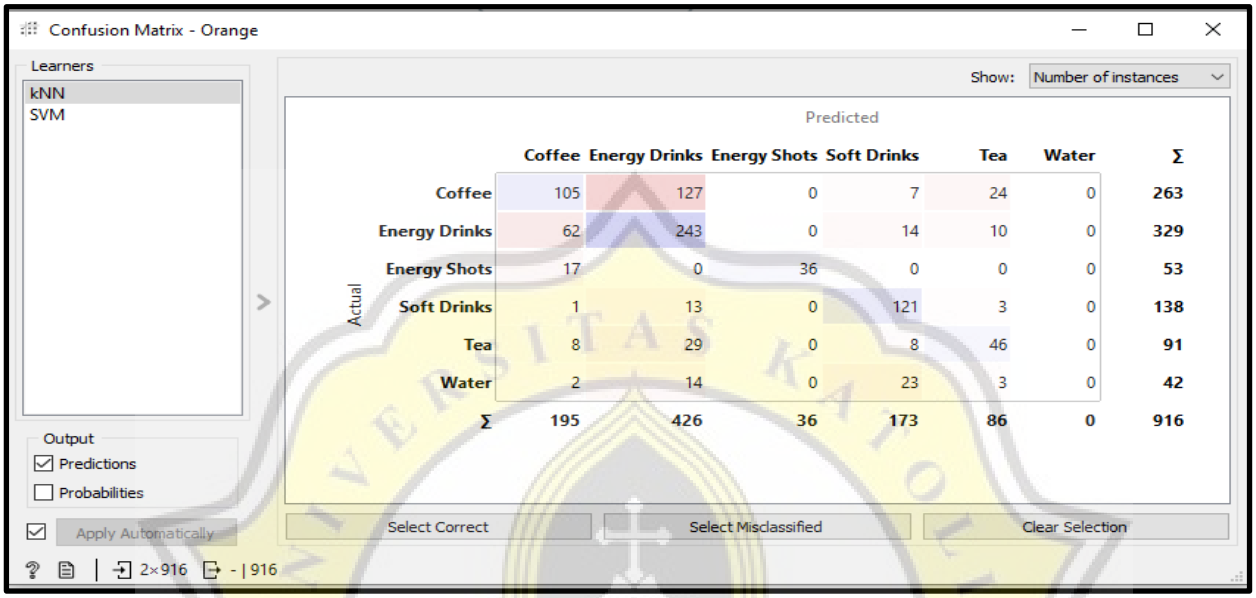

Figure 5.2 2 Confusion matrix 25 % Data Training and 75% Data Testing

In Figure 5.2.1 explains about shows the correctness and error data in the confusion matrix. The data is 458 which is then carried out 2x testing process then becomes 916 testing data. Then it is described that the value on coffee has 105 correct data out of 263 testing data. Then Energy Drink has 243 correct data out of 329 testing data. Then Energy Shots has 36 data from 53 data testing. Then Soft Drink has 121 correct data testing out of 138 data testing. Then Tea has 46 correct data out of 91 data testing. And finally Water does not have the correct data because it is considered not to have testing data.

Then the next process is to carry out the process with the Training data set as much as 50% training data and 50% testing data. The process is repeated 2 times. Then the application will produce results on the Test and Score as follows:

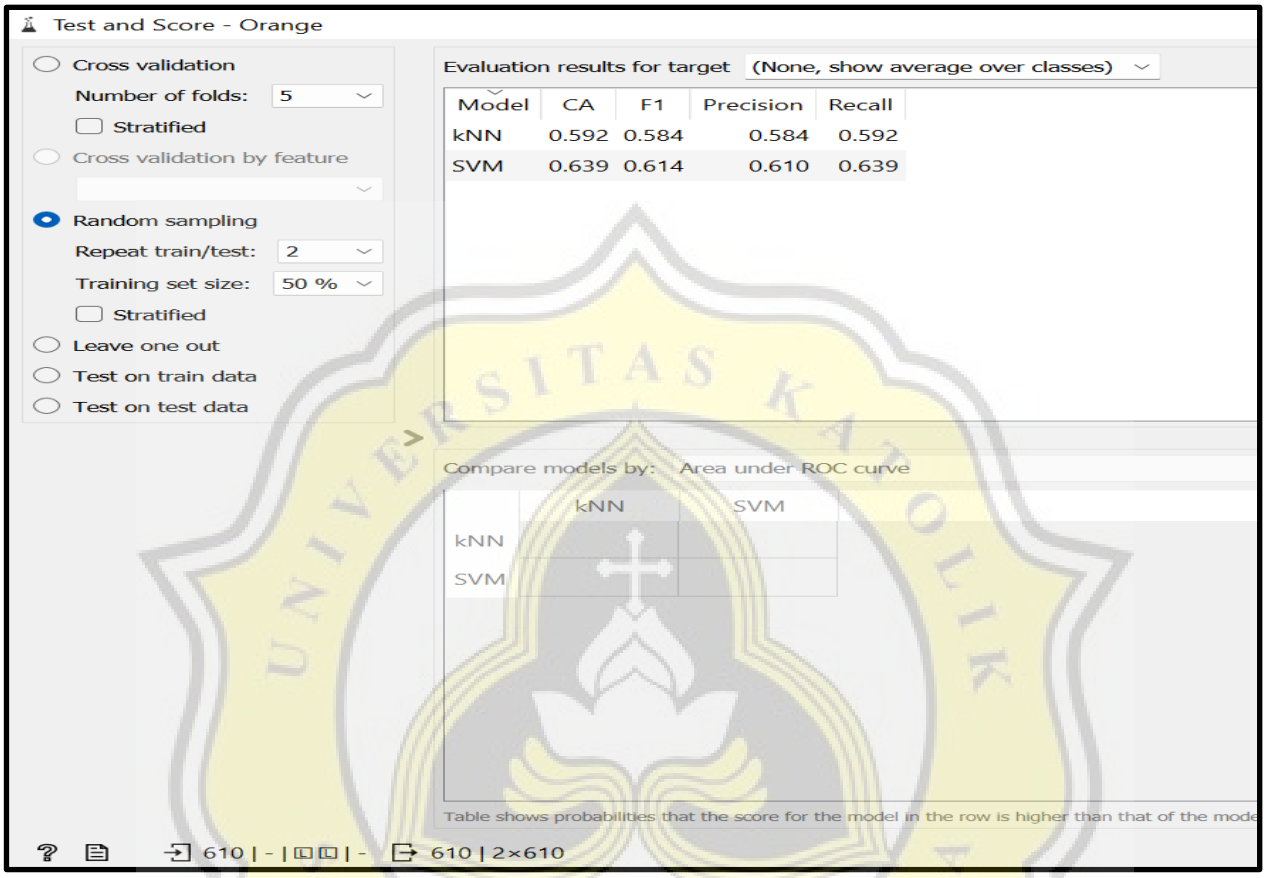

Figure 5.2 3 Test and Score 50 % Data Training and 50% Data Testing

The table shows the processing of testing data using 50% training data and 50% testing data then addressing the value:

Table 5.2 2 Test and Score 50 % Data Training and 50% Data Testing

CA	f1-score	Precision	Recall
59%	63%	58%	59%

Then in the confusion matrix that has been carried out 2x the repeat train test process shows the correct and incorrect data values as follows.

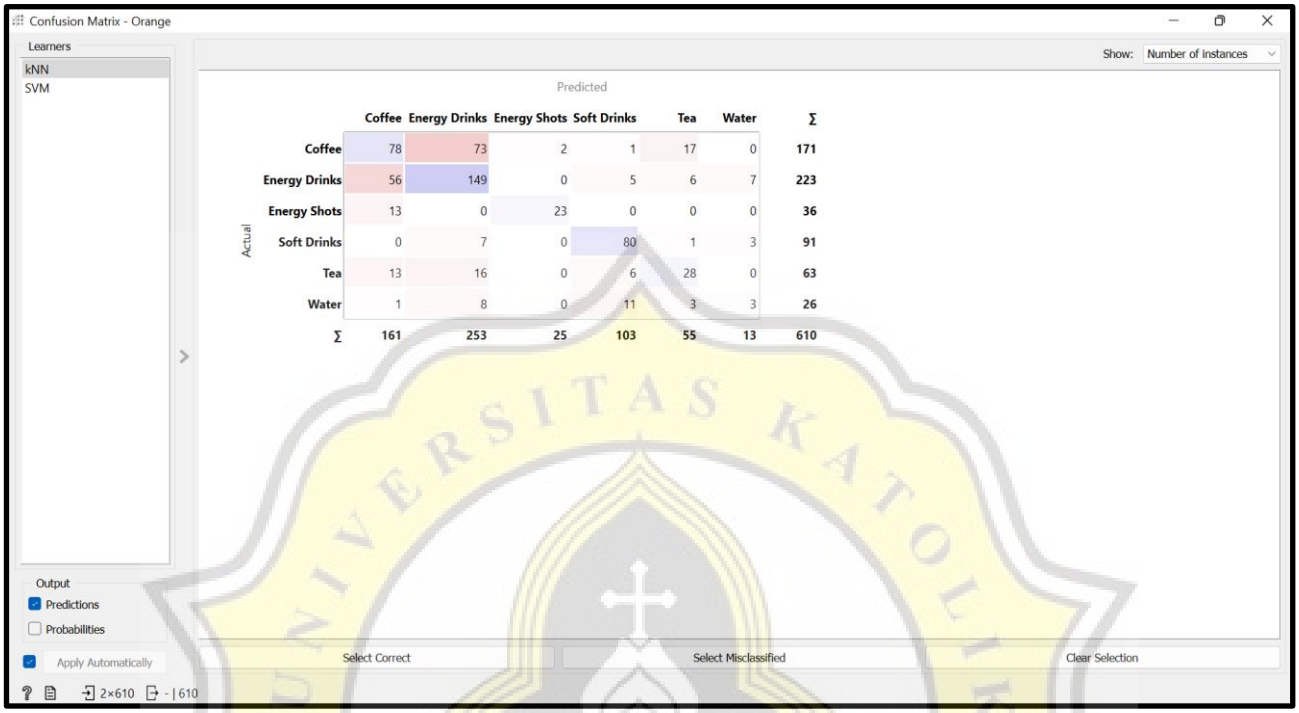

Figure 5.2 4 Confusion matrix 50 % Data Training and 50% Data Testing

In Figure 5.2.4 explains about shows the correctness and error data in the confusion matrix. The data is 310 which is then tested 2x and then becomes 610 testing data. Then it is described that the coffee value has 78 correct data out of 171 testing data. Then Energy Drink has 149 correct data from 223 testing data. Then Energy Shots has 23 correct data out of 36 testing data. Then Soft Drinks has 80 correct data out of 91 testing data. Then Tea has 28 correct data out of 63 testing data. And finally Water has 3 correct data out of 26 testing data.

Then the next last process is to carry out the process with the Training data set as much as 75% training data and 25% testing data. The process is repeated testing 2 times. Then the application will produce results on the Test and Score as follows:

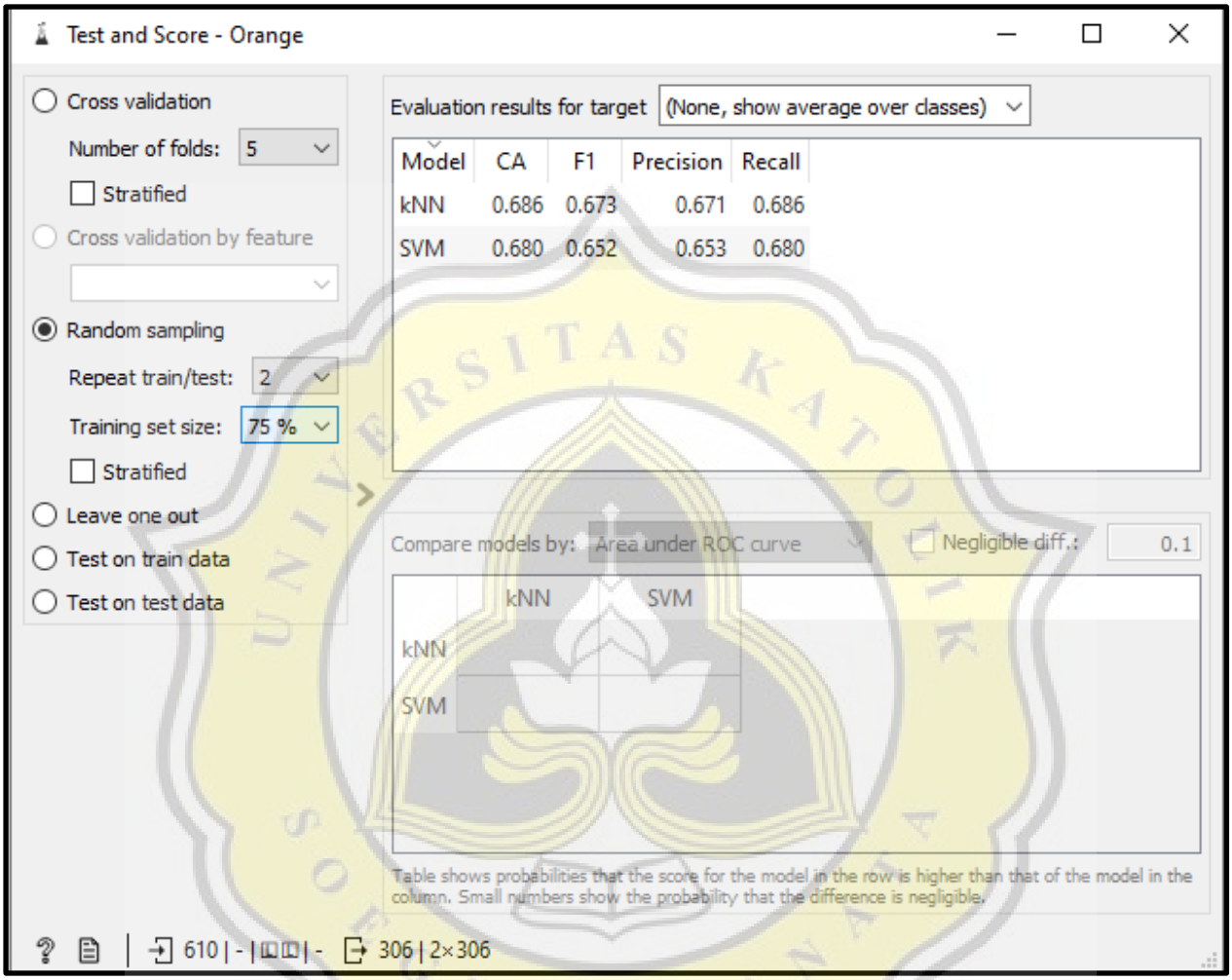

Figure 5.2 5 Test and Score 75 % Data Training and 25% Data Testing

The table shows the processing of testing data using 75% training data and 25% testing data then addressing the value:

Table 5.2 3 Test and Score 75 % Data Training and 25% Data Testing

CA	f1-score	Precision	Recall
68%	67%	67%	68%

Then in the confusion matrix that has been carried out 2x the repeat train test process shows the correct and incorrect data values as follows:

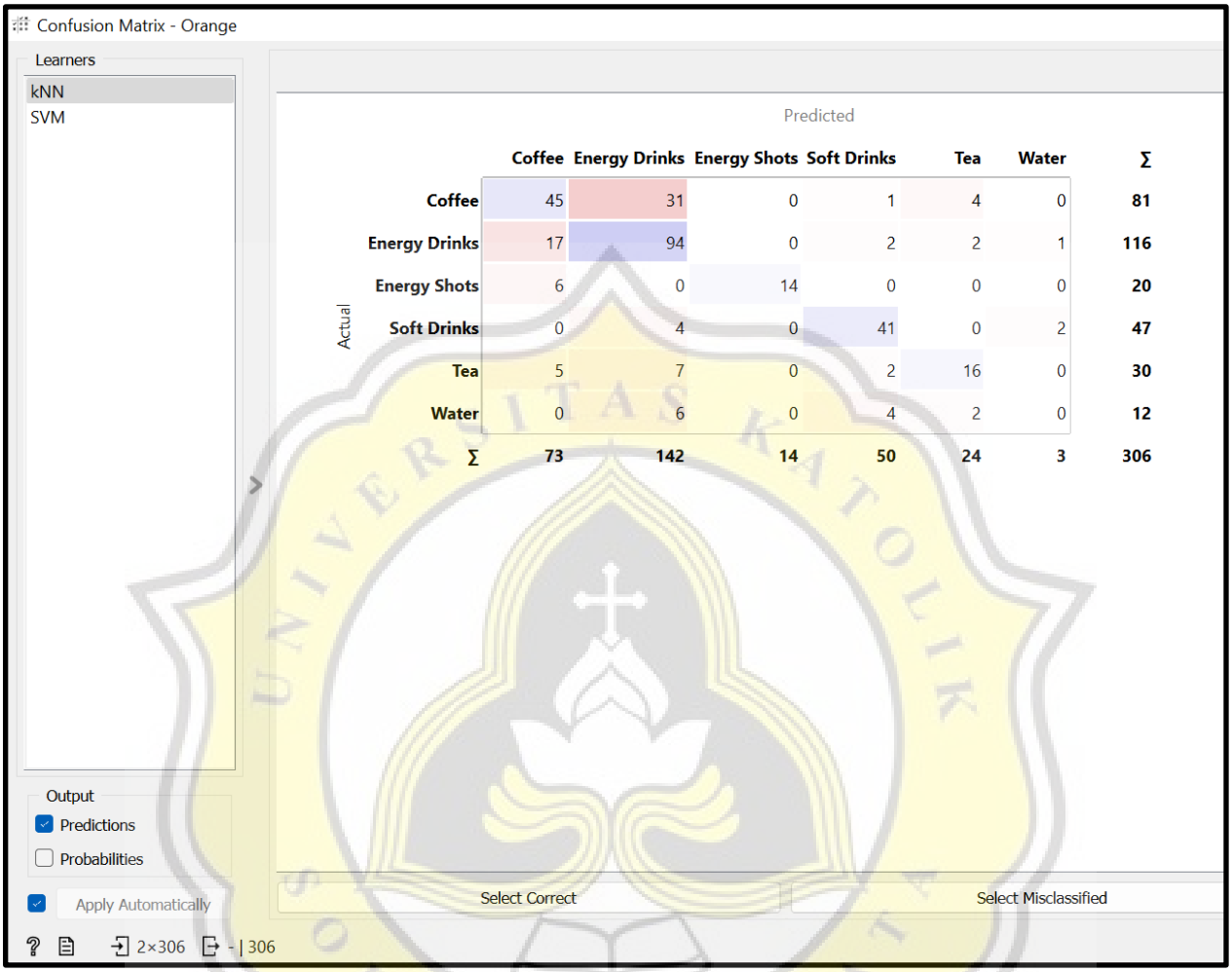

Figure 5.2 6 Confusion matrix 75 % Data Training and 25% Data Testing

In Figure 5.2.6 explains about shows the correctness and error data in the confusion matrix. The data is 153 which is then tested 2x and then becomes 306 testing data. Then it is described that the coffee value has 45 correct data out of 81 testing data. Then Energy Drinks has 94 correct data out of 116 testing data. Then Energy Shots has 14 correct data from 20 testing data. Then Soft Drink has 41 correct data from 47 testing data. Then Tea has 16 correct data from 30 testing data. And finally Water does not have the correct data because it is considered not to have testing data.

5.3. SVM Result Predicition

In the explanation of this implementation, SVM goes through several testing processes. The testing process takes place using 2x repeat train tests and training data sets of 25%, 50% and 75% to produce results for accuracy, precision, recall and f1-score. The results are displayed in the Test and Score section. And the next process determines the true and false results in the confusion matrix.

The following is a random repeat train test data process 2x with 25% training data and 75% testing processed with SVM on the Test and Score results.

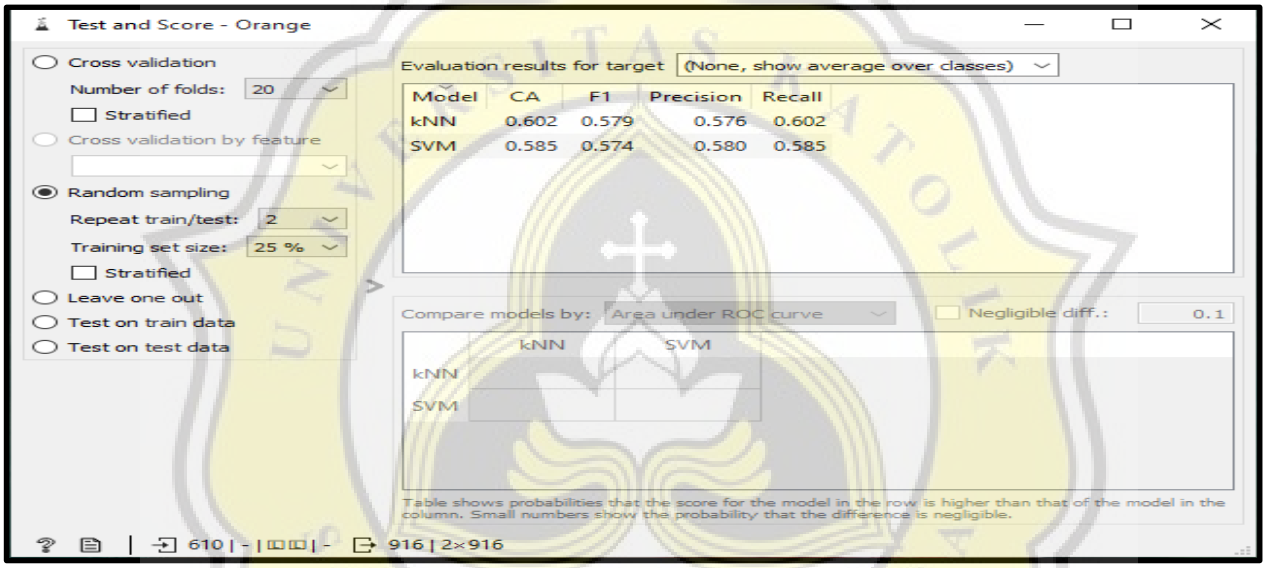

Figure 5.3 1 Test and Score 25 % Data Training and 75% Data Testing

The table shows the processing of testing data using 25% training data and 75% testing data then addressing the value:

Table 5.3 1 Test and Score 25 % Data Training and 75% Data Testing

CA	f1-scrore	Precision	Recall
58%	57%	58%	58%

Then in the confusion matrix that has been carried out 2x the repeat train test process shows the correct and incorrect data values as follows:

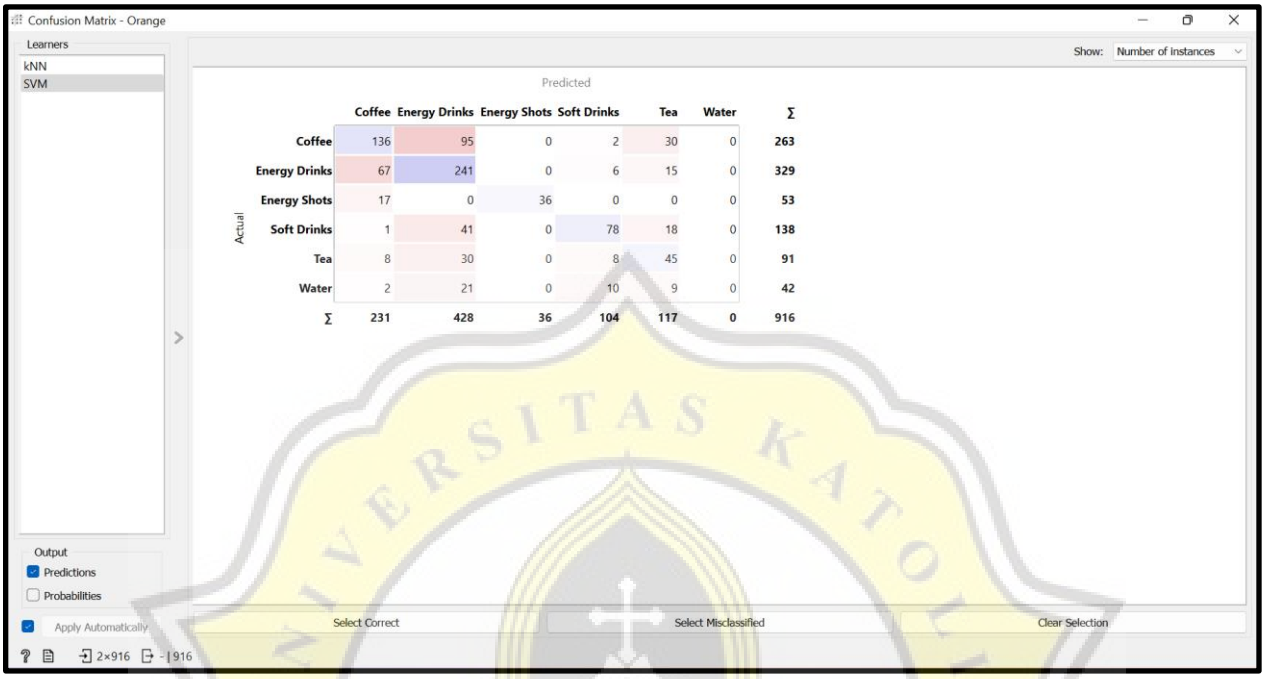

Figure 5.3 2 Confusion matrix 25 % Data Training and 75% Data Testing

In Figure 5.3.2 explains about shows the correctness and error data in the confusion matrix. The data is 458 which is then tested 2x and then becomes 916 data testing. Then it is described that the coffee value has 136 correct data out of 263 testing data. Then Energy Drinks has 241 correct data out of 329 testing data. Then Energy Shots has 36 correct data out of 53 data. Then Soft Drinks has 78 correct data out of 138 testing data. Then Tea has 45 correct data from 91 testing data. And finally Water does not have the correct data because it is considered not to have testing data.

Then the next process is to carry out the process with the Training data set as much as 50% training data and 50% testing data. The process is repeated testing 2x. Then the application will produce results on the Test and Score as follows:

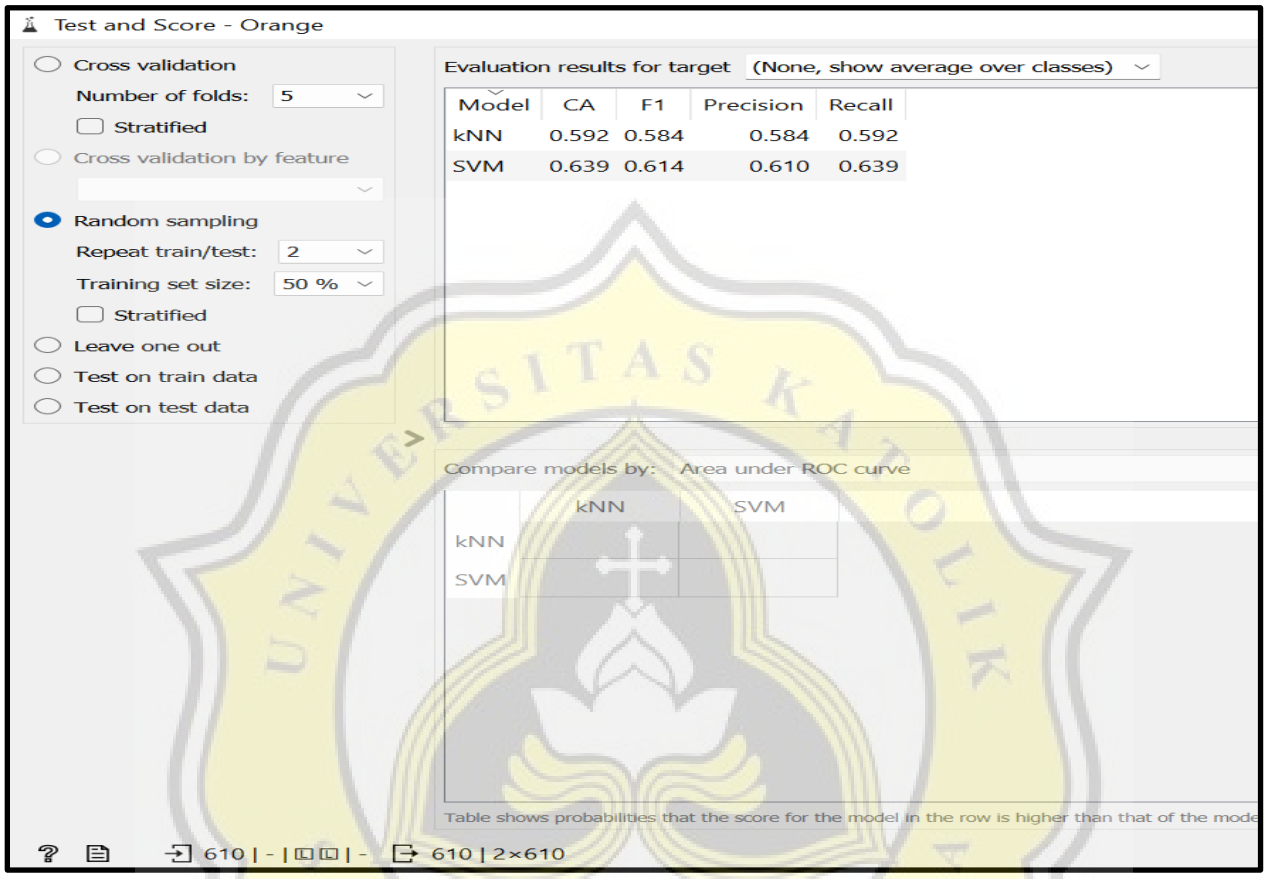

Figure 5.3 3 Test and Score 50% Data Training and 50% Data Testing

The table shows the processing of testing data using 50% training data and 50% testing data then addressing the value:

Table 5.3 2 Test and Score 50 % Data Training and 50% Data Testing

CA	f1-score	Precision	Recall
63%	61%	61%	63%

Kemudian pada confusion Matrix yang telah dilakukan 2x proses repeat train test menunjukan nilai kebenaran dan kesalahan data sebagai berikut:

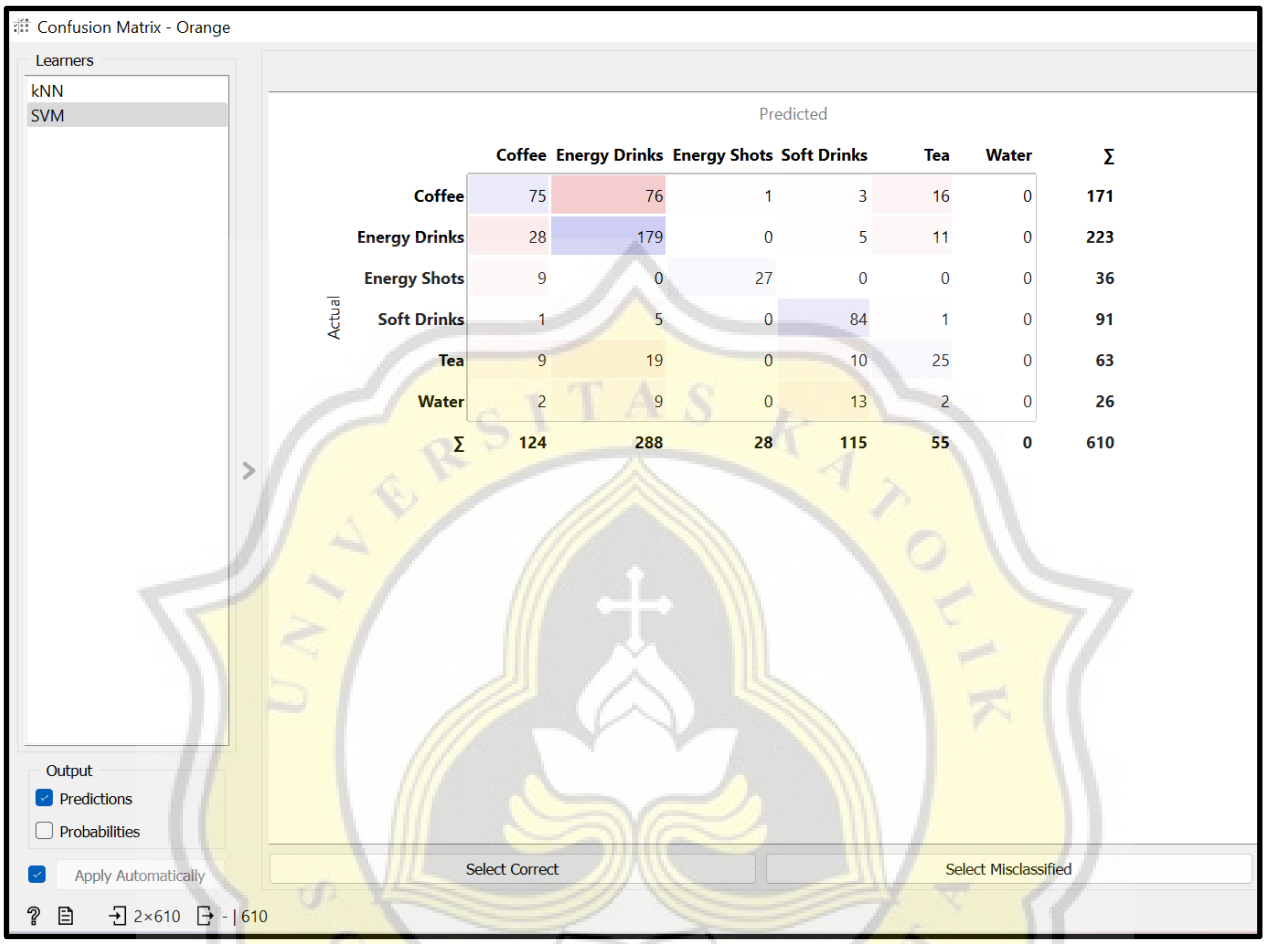

Figure 5.3 4 Confusion matrix 50% Data Training and 50% Data Testing

In Figure 5.3.4 explains about shows the correctness and error data in the confusion matrix. The data is 310 which is then tested 2x and then becomes 610 testing data. Then it is described that the coffee value has 75 correct data out of 171 testing data. Then Energy Drinks has 179 correct data from 223 testing data. Then Energy Shots has 27 correct data out of 36 testing data. Then Soft Drinks has 84 correct data from 91 testing data. Then Tea has 25 correct data from 63 testing data. And finally Water does not have the correct data because it is considered not to have testing data.

Then the next last process is to carry out the process with the Training data set as much as 75% training data and 25% testing data. The process is repeated testing 2x. Then the application will produce results on the Test and Score as follows:

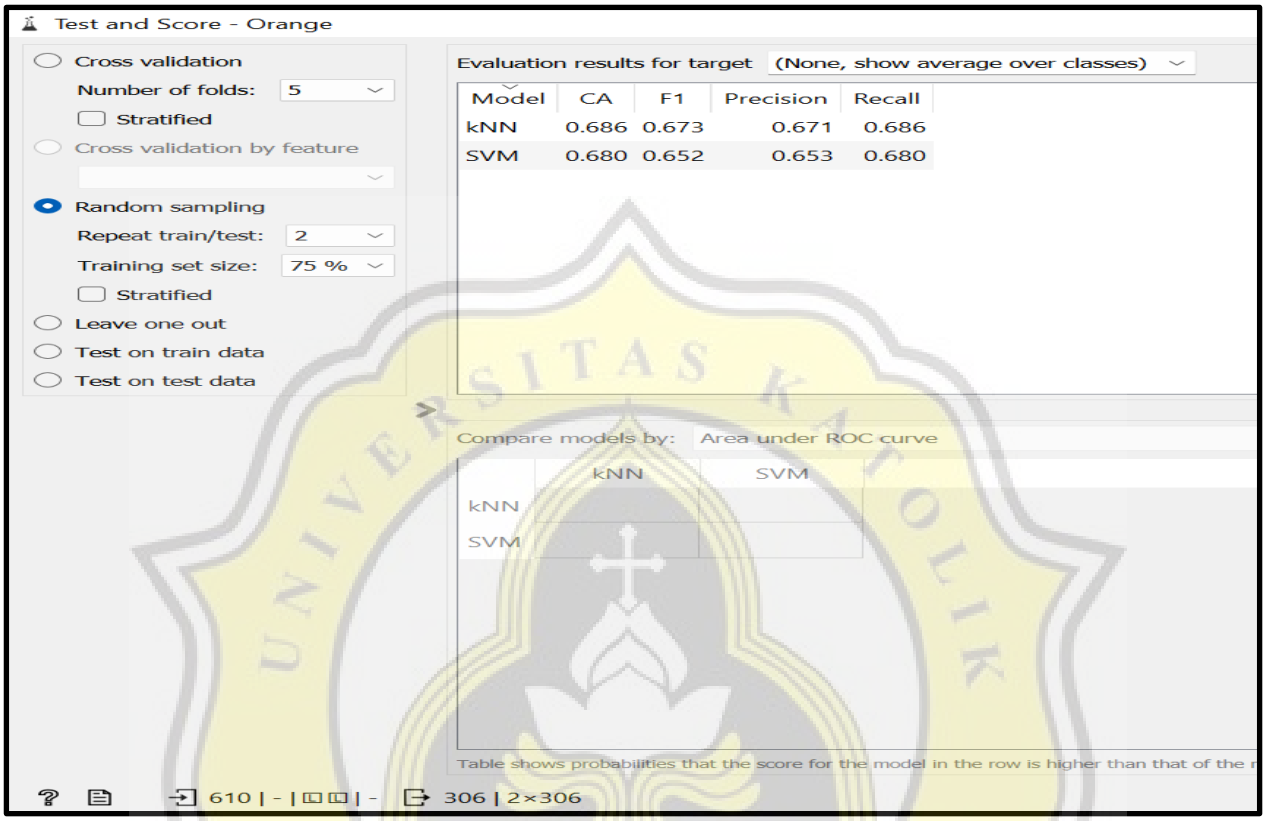

Figure 5.3 5 Test and Score 75% Data Training and 25% Data Testing

The table shows the processing of testing data using 75% training data and 25% testing data then addressing the value:

Table 5.3 3 Test and Score 75 % Data Training and 25% Data Testing

CA	f1-score	Precision	Recall
68%	65%	65%	68%

Then in the confusion matrix that has been carried out 2x the repeat train test process shows the correct and incorrect data values as follows:

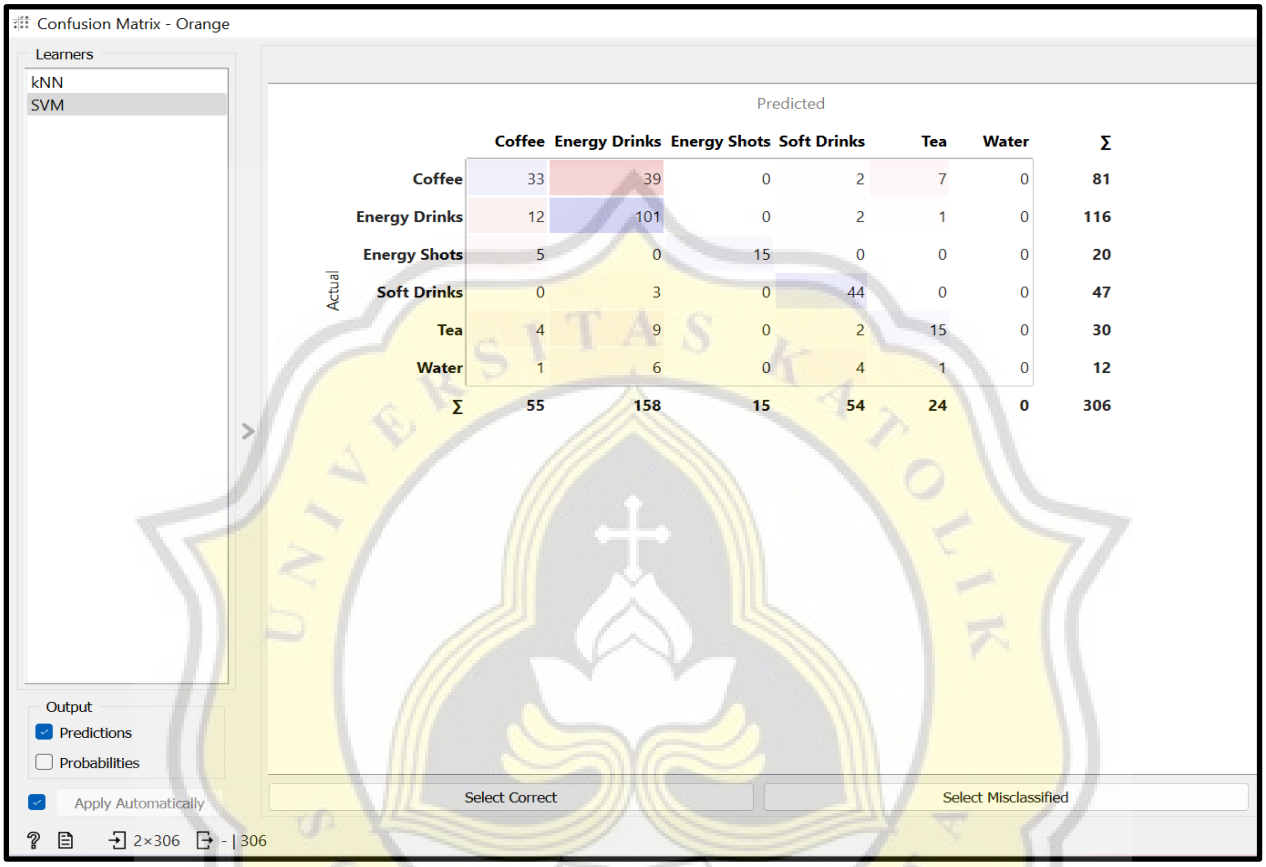

Figure 5.3 6 Confusion matrix 75% Data Training and 25% Data Testing

In Figure 5.3.6 explains about shows the correctness and error data in the confusion matrix. The data is 153 which is then tested 2x and then becomes 306 testing data. Then it is described that the coffee value has 33 correct data out of 81 testing data. Then Energy Drinks has 101 correct data out of 116 testing data. Then Energy Shots has 15 correct data out of 20 testing data. Then Soft Drinks has 44 correct data from 47 testing data. Then Tea has 15 correct data from 30 testing data. And finally Water does not have the correct data because it is considered not to have testing data.

5.4. Result

The results of the implementation of the two algorithms with Orange have a conclusion. Orange has deficiencies when determining detailed results on ca, f1-score, precision, and recall. Then there are many features whose functions are unknown and only show instantly. Therefore, there are many features that must be studied further according to the needs we want. Then when processing the application it only states that the data is true and false. So it can't be as detailed as using processing with coding. So we cannot determine where the TN, FN, TP, and FP are located. But the strengths of Orange are that it's easy to use and does it instantly.

The results of the processing above show that the two algorithms have results that are similar and slightly compared. . K-NN shows an accuracy rate of 0.602/60% and SVM shows a value of 0.585/58%. The following shows that K-NN has a higher level of accuracy .Because the two algorithms have the nature of "Supervised Algorithm" which means that the two algorithms must have training data. Then the last result shows that the K-NN algorithm is superior to SVM because some data shows the average accuracy of the "test and score" and "confusion matrix" results that K-NN has better results, even though it differs slightly from SVM.