CHAPTER 5
IMPLEMENTATION AND RESULTS

5.1 Implementation

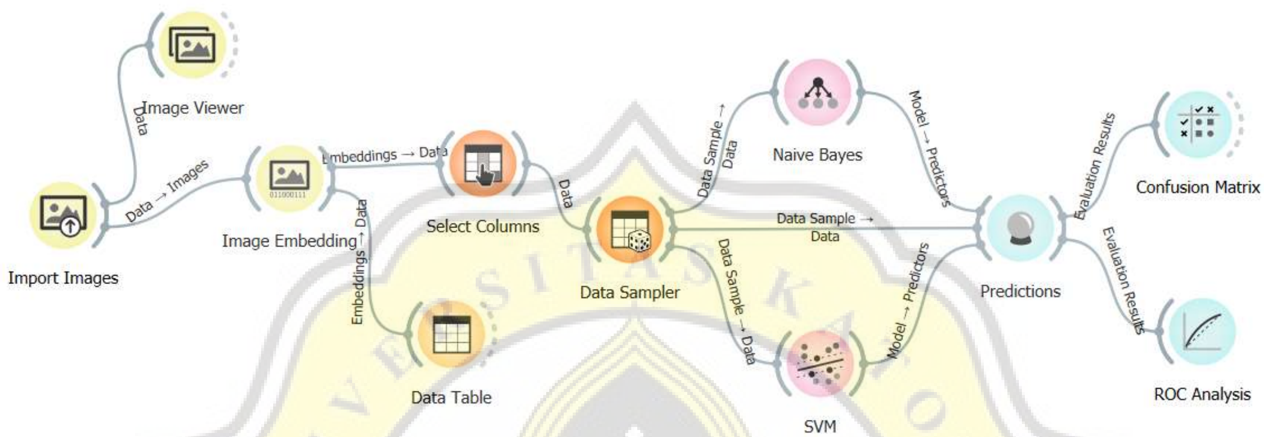

Figure 5.1 example of the implementation using Orange Data Mining software

This project created with Orange Data Mining software and implemented based on Figure 5.1 . The purpose of this project is to compare the accuracy, precision, and recall of the Naive Bayes and SVM algorithms. Starting from collecting dataset from the Kaggle.com website, a collection of photographs of human faces with and without masks in a range of brightness, image size, gender, and object tilt, then divide it into the folders named 1 (With Mask) and 0 (Without Mask). And then upload a folder containing image data using the Import Images widget in the Orange Data Mining software's Image Analytic add-ons.

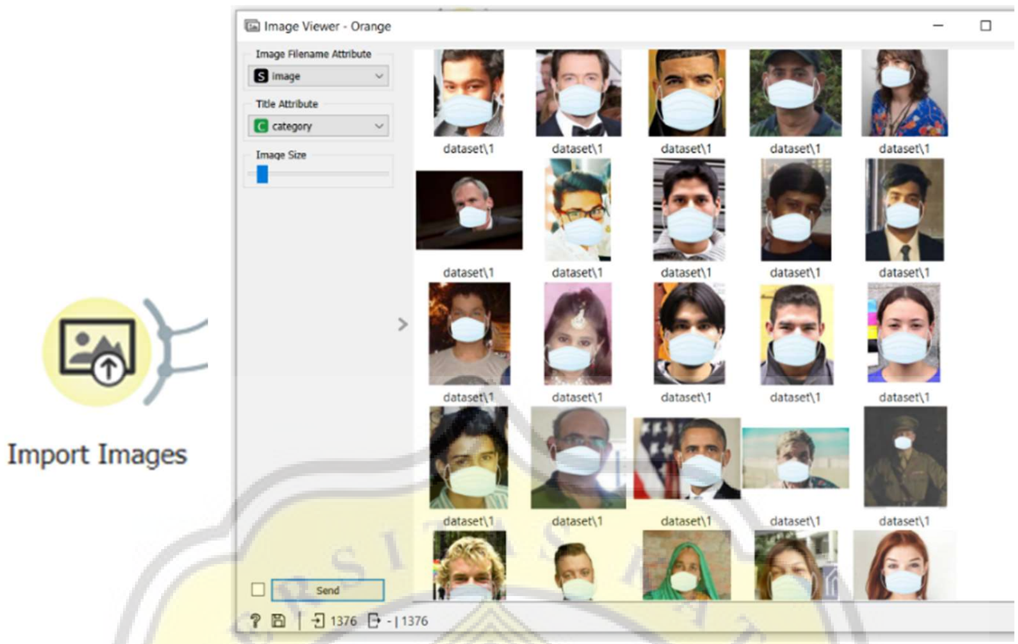

Figure 5.2 example of the dataset that has been uploaded

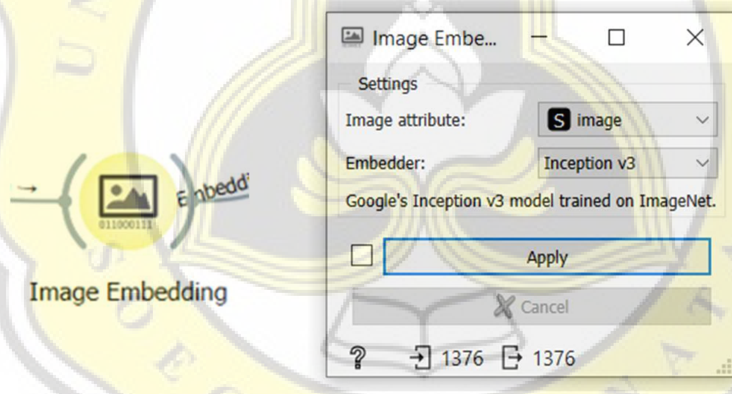

Figure 5.3 example of Image Embedding

Images uploaded through the Image Import widget can be viewed in the Image Viewer widget. The image viewer displays a preview of the uploaded images, image attributes, title attributes, size and number of images, as shown in Figure 5.2. After that, you can analyse photos locally or upload data to a remote server using the Image Embedding widget inside the image analytics options. Google's deep neural network for image identification is called InceptionV3. The ImageNet data collection is used to train it. The embedder that will be used is inceptionv3, based on Figure 5.3. Make sure that the imported data is accurate and that no data is missing using the Data Table widget. Example of Data Table's widget are on Figure 5.4. The data table will show the category, image name, image size, and other information, as well as the total quantity of uploaded data and whether any data is missing, as shown in the figure 5.4.

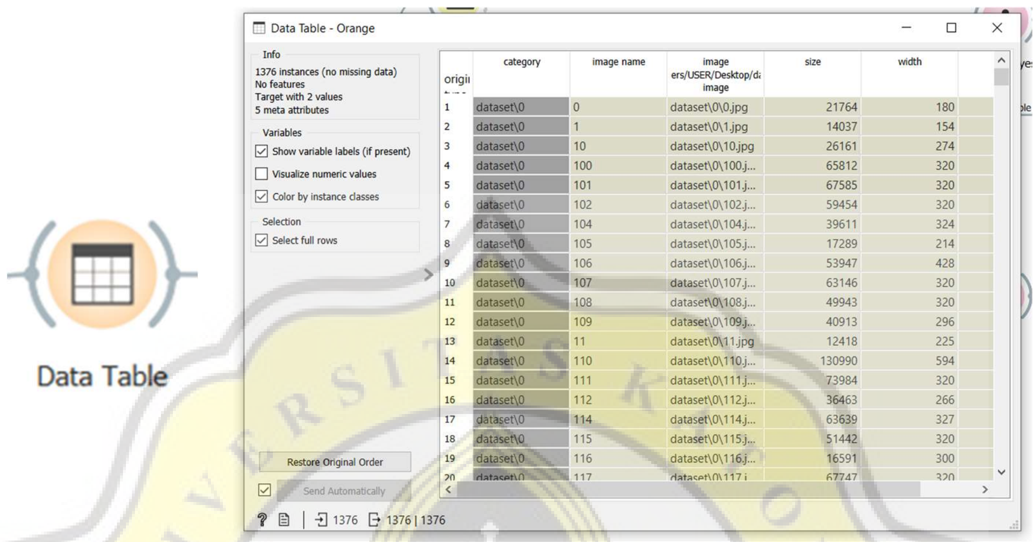

Figure 5.4 example of Data Table

Select the Target named Category in the Select Column widget to create your data domain by manually, as seen on figure 5.5.

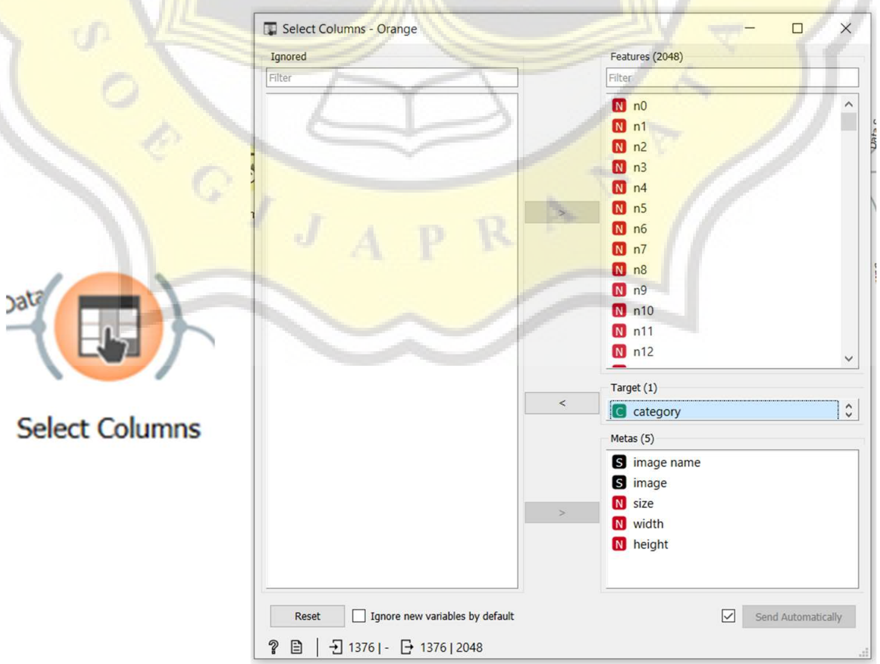

Figure 5.5 example of Select Columns

Furthermore, with the Data Sampler, the data will be divided into two, training and testing. In the figure 5.6 below, data that will be split into two will be used for training and testing purposes. While data testing is used to evaluate the performance of the algorithm, data training involves sending data to the algorithm for testing.

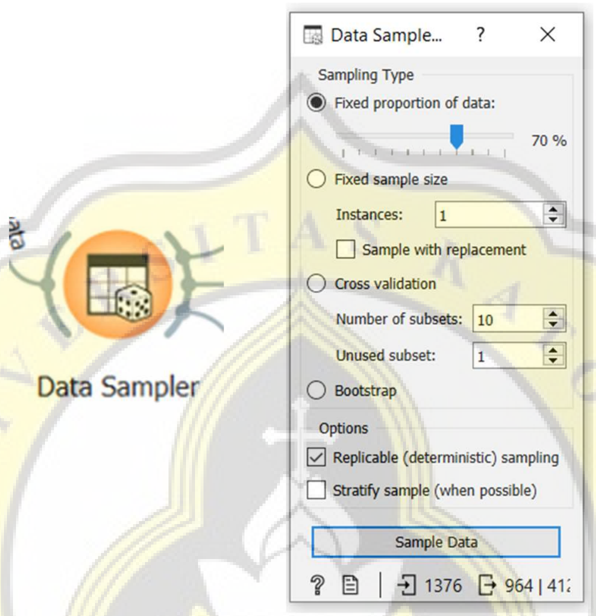

Figure 5.6 example of Data Sampler

Implementation of Naive Bayes and SVM algorithms, SVM and Naive Bayes algorithms are in the Model options, as seen on Figure 5.7. Additionally, there are a lot of models available in Orange Data Mining. SVM and Nave Bayes are two classification algorithms.

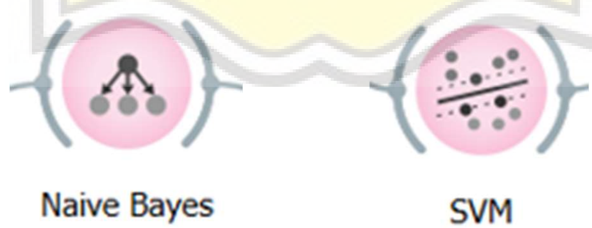

Figure 5.7 example of Naive Bayes and SVM Model

Based on figure 5.8, SVM has several kernels, including Linear, Polynomial, RBF, and Sigmoid. In this section, we'll compare every kernel from the SVM to maximize test results.

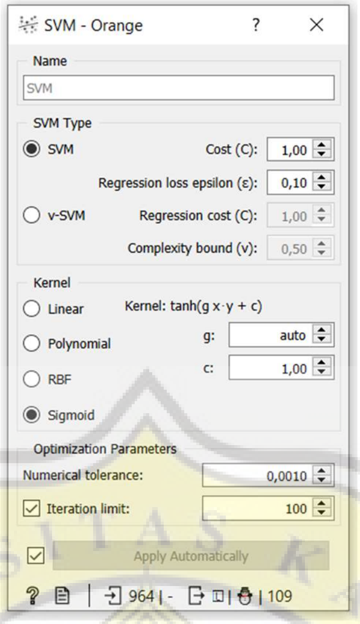

Figure 5.8 example of SVM Kernels

5.2 Results

The dataset being tested will be tested in stages. The 1376 existing datasets will be divided by approximately 25% for the first experiment, approximately 50% for the second trial, and all accessible datasets will be tested in the third experiment to ensure that the test is as accurate as possible. The first stage has 331 datasets, the second stage has 732 datasets and the third stage has 1376 datasets. Results of the first stage for Accuracy, Precision, and Recall are displayed on the Prediction widget, as seen on Table 5.1. Using the ROC and Confusion Matrix to evaluate the effectiveness of classification models.

	AUC	CA	F1	Precision	Recall
Naïve Bayes	1000	0,991	0,991	0,992	0,991
SVM Sigmoid	1000	0,996	0,996	0,996	0,996
SVM RBF	1000	1000	1000	1000	1000
SVM Polynomial	1000	1000	1000	1000	1000
SVM Linear	1000	1000	1000	1000	1000

Table 5.1 the result of the first stage at Prediction widget

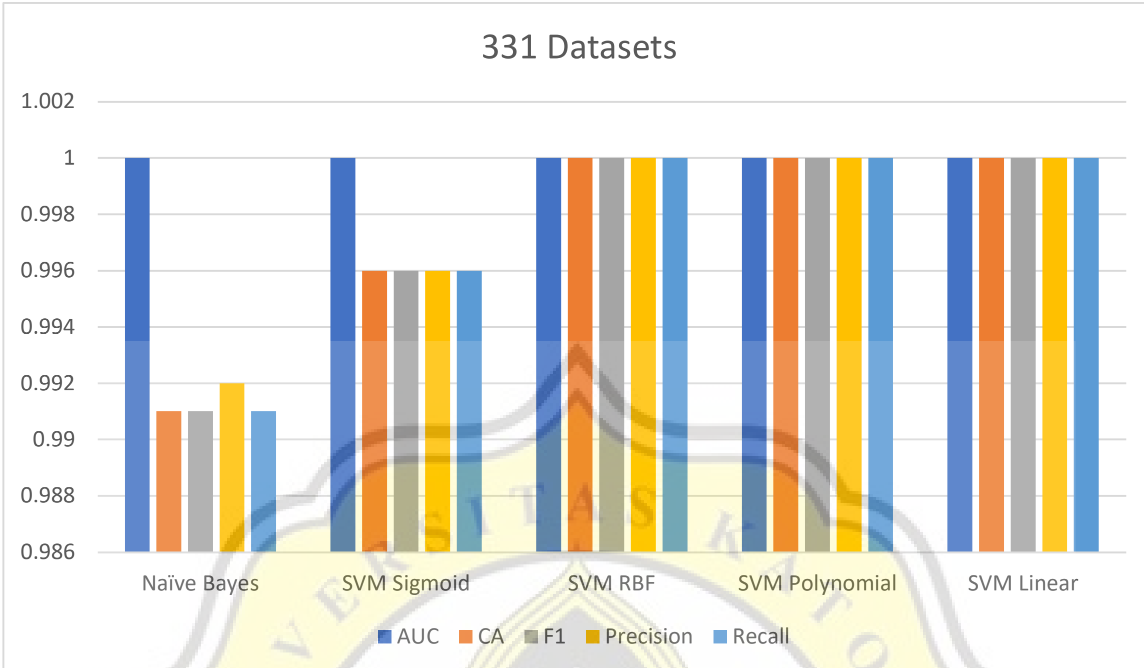

Figure 5.9 the chart of Prediction widget's result with 331 datasets.

After testing the first stage's classification findings, classification accuracy (CA) is the accurate proportion of data records classified correctly. The percentage of positive anticipated cases that are likewise positive correct on the actual data is known as precision or confidence. The percentage of real positive situations that are accurately indicated positively is known as recall or sensitivity. The F1 score is a weighted average of recall and precision, with a highest score of 1 and a worst score of 0. The percentage contributions of precision and recall to the F1 score are identical. According to figure 5.9, it can be seen that SVM is much superior to Naive Bayes because the Kernel RBF, Polynomial, Linear SVM has CA, F1, Precision, and Recall of 1000. But the Sigmoid Kernel SVM results are slightly lower than other kernels, namely CA 0.996, F1 0.996, Precision 0.996, and Recall 0.996. While Naive Bayes itself has the lowest value, namely CA 0.991, F1 0.991, Precision 0.992, and Recall 0.991.

In the second stage, there is a decrease in the accuracy rate in the algorithm testing. Results of the second stage for Accuracy, Precision, and Recall are displayed on the Prediction widget, as seen on Table 5.2 .

	AUC	CA	F1	Precision	Recall
Naïve Bayes	1000	0,984	0,984	0,985	0,984
SVM Sigmoid	1000	0,994	0,994	0,994	0,994
SVM RBF	1000	1000	1000	1000	1000
SVM Polynomial	1000	1000	1000	1000	1000
SVM Linear	1000	1000	1000	1000	1000

Table 5.2 the result of the second stage at Prediction Widget.

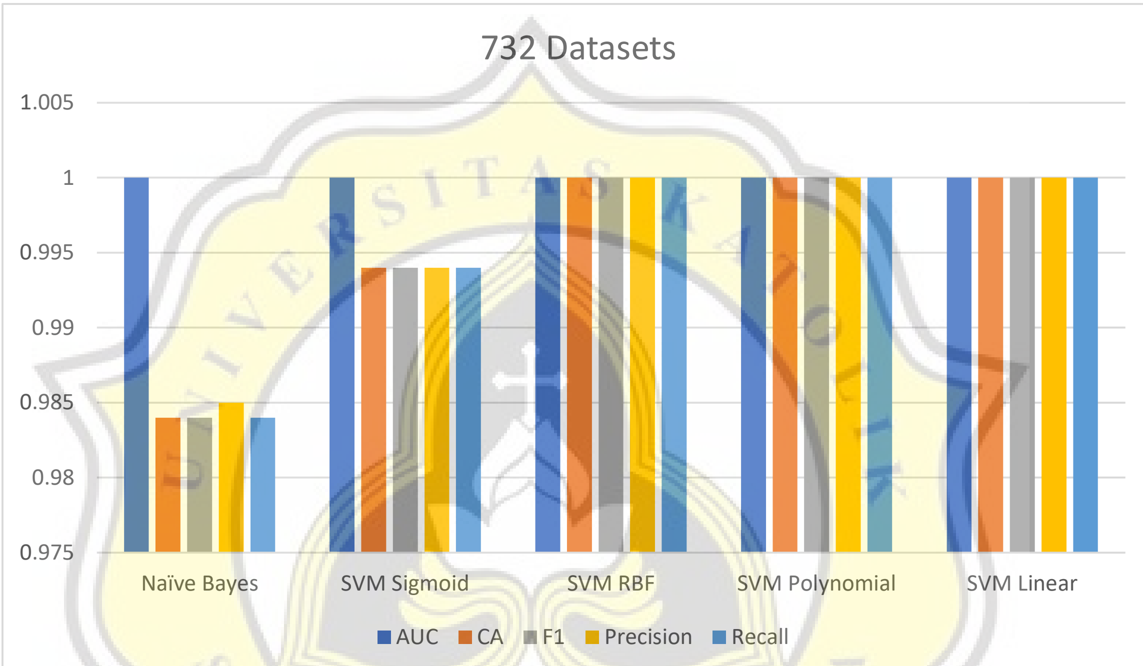

Figure 5.10 the chart of Prediction widget's result with 732 datasets.

Figure 5.10 shows that SVM is significantly better than Naive Bayes since it has CA, F1, Precision, and Recall values of 1000 for the Kernel RBF, Polynomial, Linear SVM. But the Sigmoid Kernel SVM results are slightly lower than other kernels like the first stage, namely CA 0.994, F1 0.994, Precision 0.994, and Recall 0.994. While Naive Bayes itself has the lowest value, namely CA 0.984, F1 0.984, Precision 0.985, and Recall 0.984.

Using the 1376 Dataset in the final phase. The Prediction widget displays the results of the second stage for Accuracy, Precision, and Recall, as shown in Table 5.3.

	AUC	CA	F1	Precision	Recall
Naïve Bayes	1000	0,97	0,97	0,972	0,97
SVM Sigmoid	1000	0,993	0,993	0,993	0,993
SVM RBF	1000	1000	1000	1000	1000
SVM Polynomial	1000	1000	1000	1000	1000
SVM Linear	1000	1000	1000	1000	1000

Table 5.3 the result of the third stage at Prediction Widget.

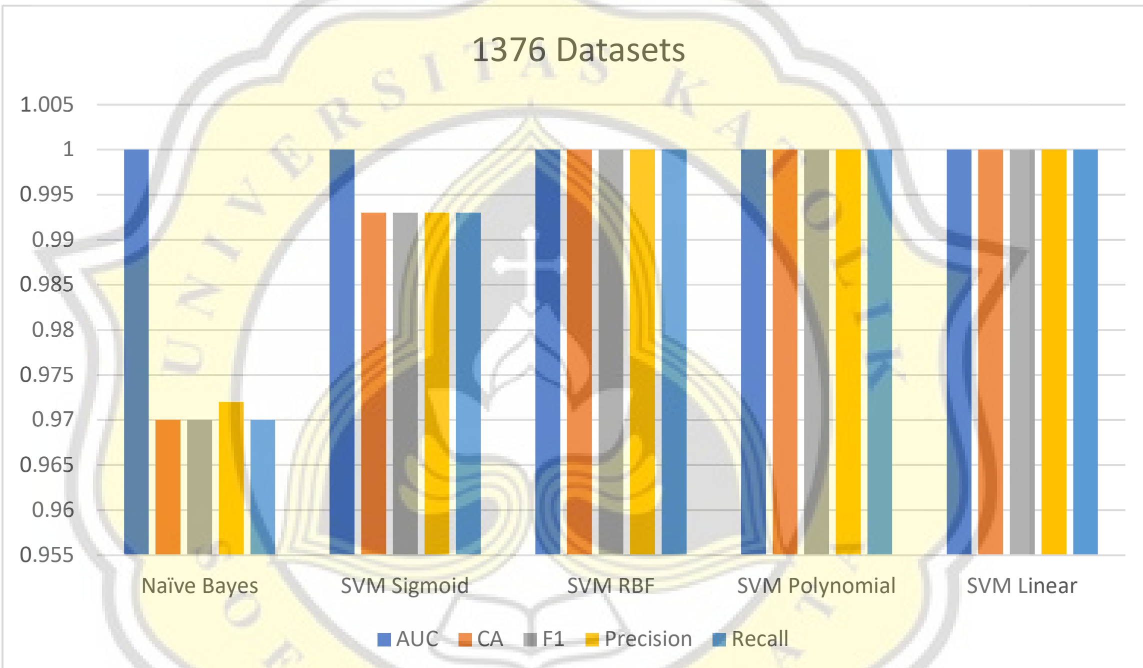

Figure 5.11 the chart of Prediction widget's result with 1376 datasets.

Figure 5.11 shows that SVM is significantly better than Naive Bayes since it has CA, F1, Precision, and Recall values of 1000 for the Kernel RBF, Polynomial, Linear SVM. But the Sigmoid Kernel SVM results are slightly lower than other kernels like the first stage, namely CA 0.993, F1 0.993, Precision 0.993, and Recall 0.993. The value of Naive Bayes is the lowest, namely CA 0.972, F1 0.972, Precision 0.972, and Recall 0.972. From the following three trials, it was determined that the accuracy values of the Naive Bayes and Sigmoid SVM algorithms decreased with the more amount of dataset examined in the algorithm.

Predicted

Actual	dataset\0	dataset\1	Σ
dataset\0	92	0	92
dataset\1	2	138	140
Σ	94	138	232

Figure 5.12 Naive Bayes’s confusion matrix at first stage test

Predicted

Actual	dataset\0	dataset\1	Σ
dataset\0	92	0	92
dataset\1	0	140	140
Σ	92	140	232

Figure 5.13 SVM’s confusion matrix at first stage test

According to first stage’s the data table's confusion matrix, figure 5.10 and figure 5.11, Naive Bayes predicts 92 True Positives and 138 True Negatives, while predicting 0 True Positives and 2 False Negatives. Model SVM predicts 92 True Positives and 140 True Negatives, 0 False Positives and 0 False Negatives.

		Predicted		
		dataset\0	dataset\1	∑
Actual	dataset\0	279	0	279
	dataset\1	8	226	234
	∑	287	226	513

Figure 5.14 Naïve Bayes's confusion matrix at second stage test

		Predicted		
		dataset\0	dataset\1	∑
Actual	dataset\0	279	0	279
	dataset\1	0	234	234
	∑	279	234	513

Figure 5.15 SVM's confusion matrix at second stage test

According to second stage's the data table's confusion matrix, figure 5.12 and figure 5.13, Naive Bayes predicts 279 True Positives and 226 True Negatives, while predicting 0 True Positives and 8 False Negatives. Model SVM predicts 279 True Positives and 234 True Negatives, 0 False Positives and 0 False Negatives.

		Predicted		
		dataset\0	dataset\1	∑
Actual	dataset\0	491	0	491
	dataset\1	29	444	473
	∑	520	444	964

Figure 5.16 Naïve Bayes's confusion matrix at third stage test

		Predicted		
		dataset\0	dataset\1	∑
Actual	dataset\0	491	0	491
	dataset\1	7	466	473
	∑	498	466	964

Figure 5.17 SVM's confusion matrix at third stage test

According to the data table's confusion matrix, figure 5.14 and figure 5.15, Naive Bayes predicts 491 True Positives and 466 True Negatives, while predicting 0 True Positives and 29 False Negatives. Model SVM predicts 491 True Positives and 466 True Negatives, 0 False Positives and 7 False Negatives. From this, it can be inferred that the SVM algorithm predicts less False negative and makes more accurate predictions than the Naive Bayes algorithm.

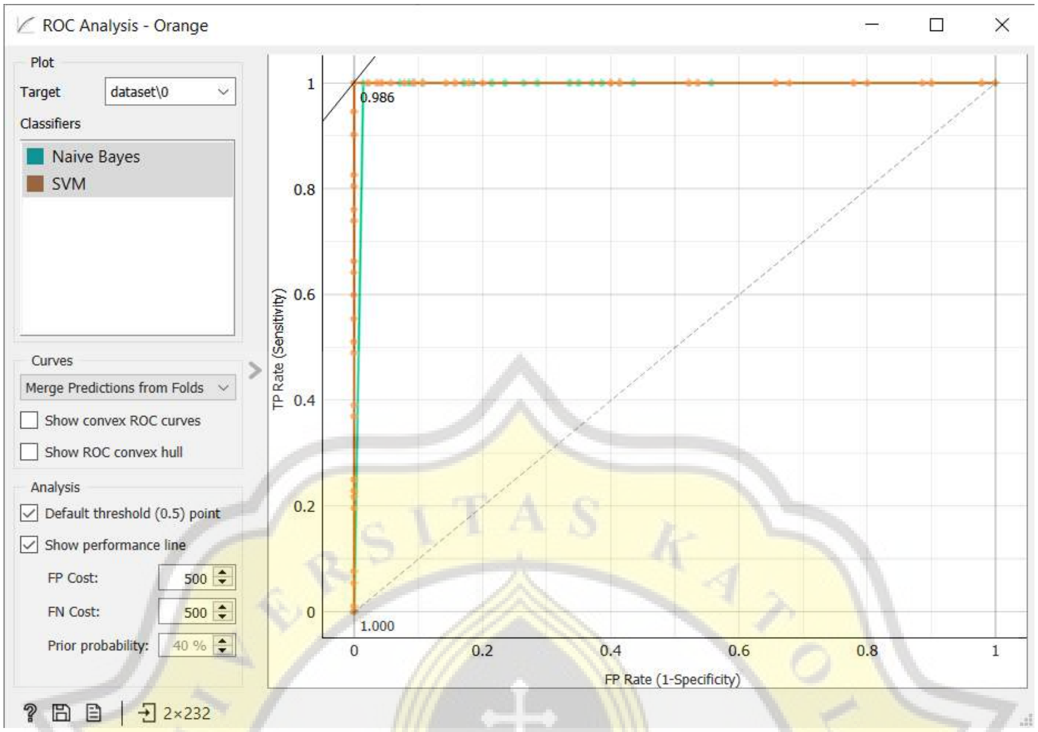

Figure 5.18 ROC Analysis graph of Naive Bayes and SVM at first testing

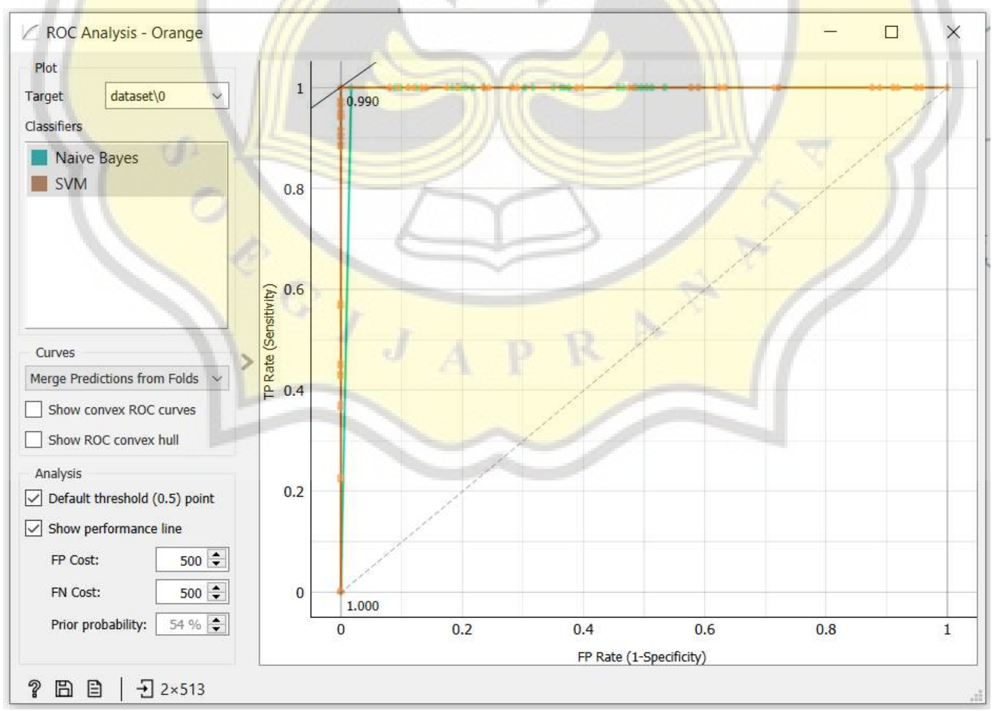

Figure 5.19 ROC Analysis graph of Naive Bayes and SVM at second testing

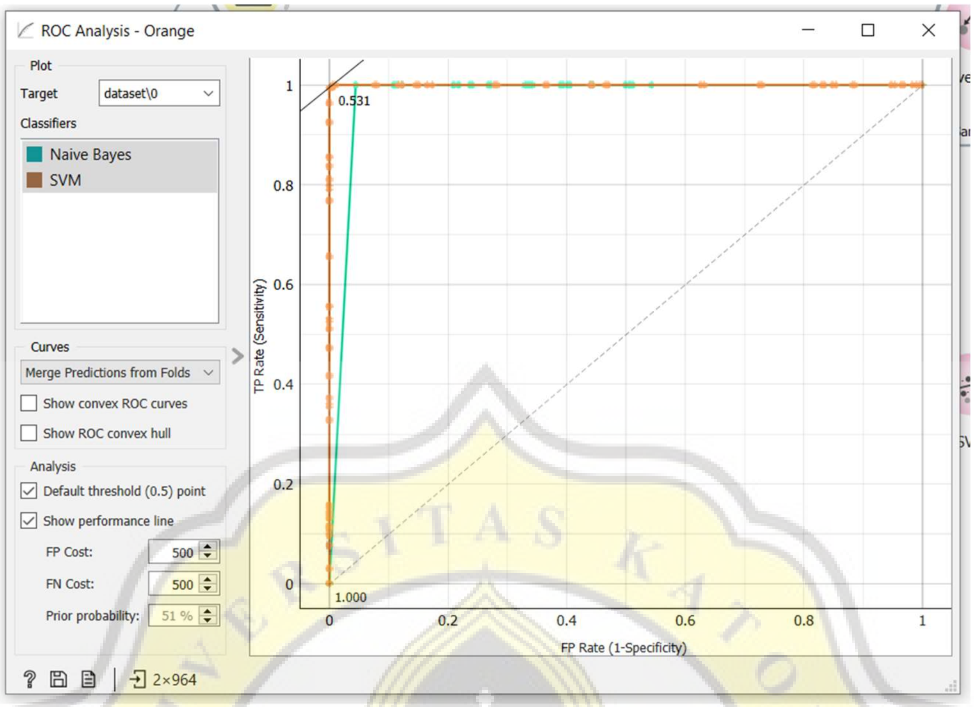

Figure 5.20 ROC Analysis graph of Naive Bayes and SVM at third testing

From figure 5.16, 5.17 and 5.18 above, it can be seen that Naive Bayes is in the blue line and SVM is in the brown line. The SVM algorithm is more accurate here because the lines are perfect and pass through the True Positive line, while Naïve Bayes algorithm are closer to False Positive line..