MITSUBISHI ELECTRIC
Changes for the Better

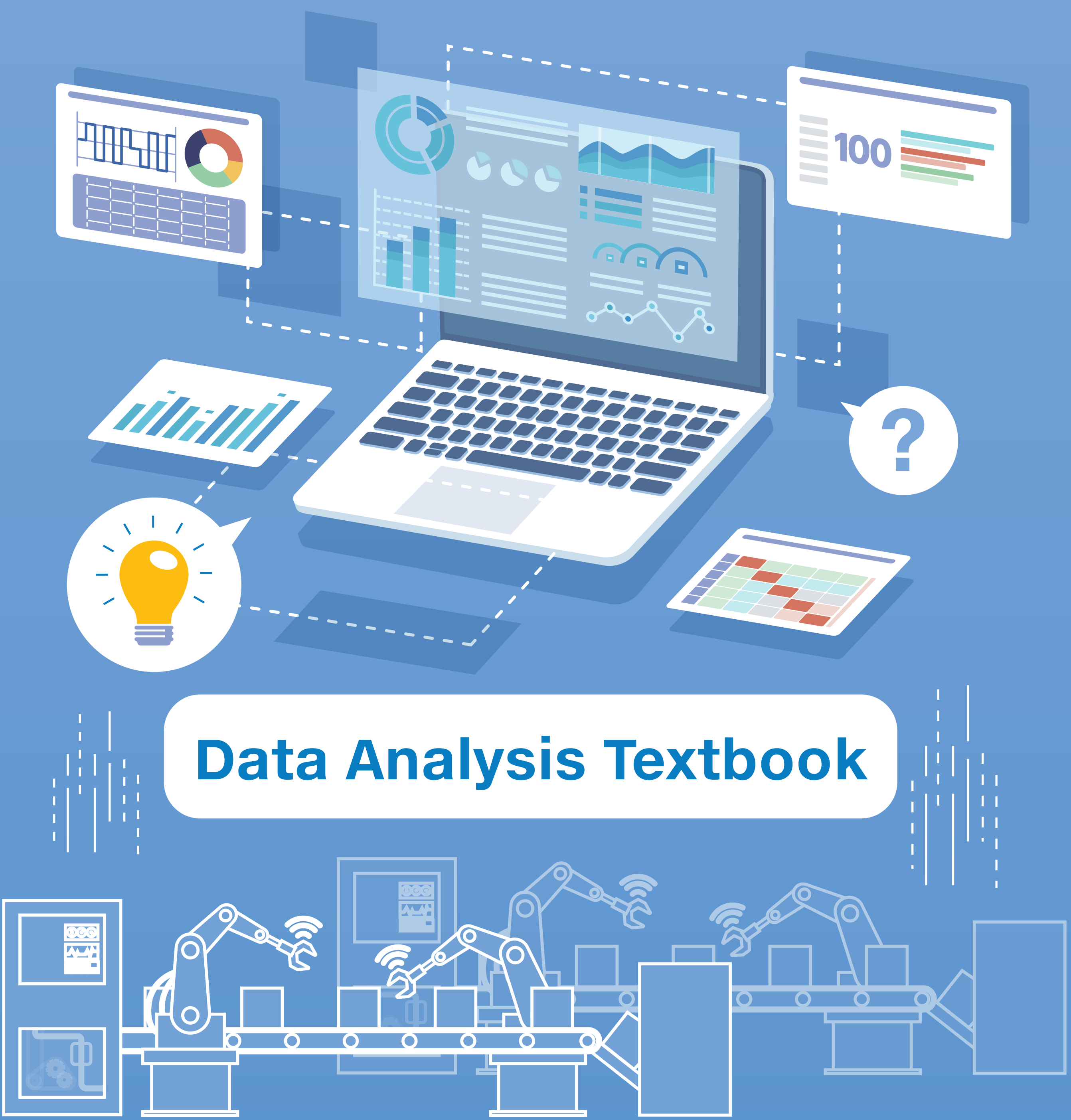
FACTORY AUTOMATION

Mitsubishi Electric Data Science Tool

MELSOFT MaiLab

e-F@ctory

Maisart



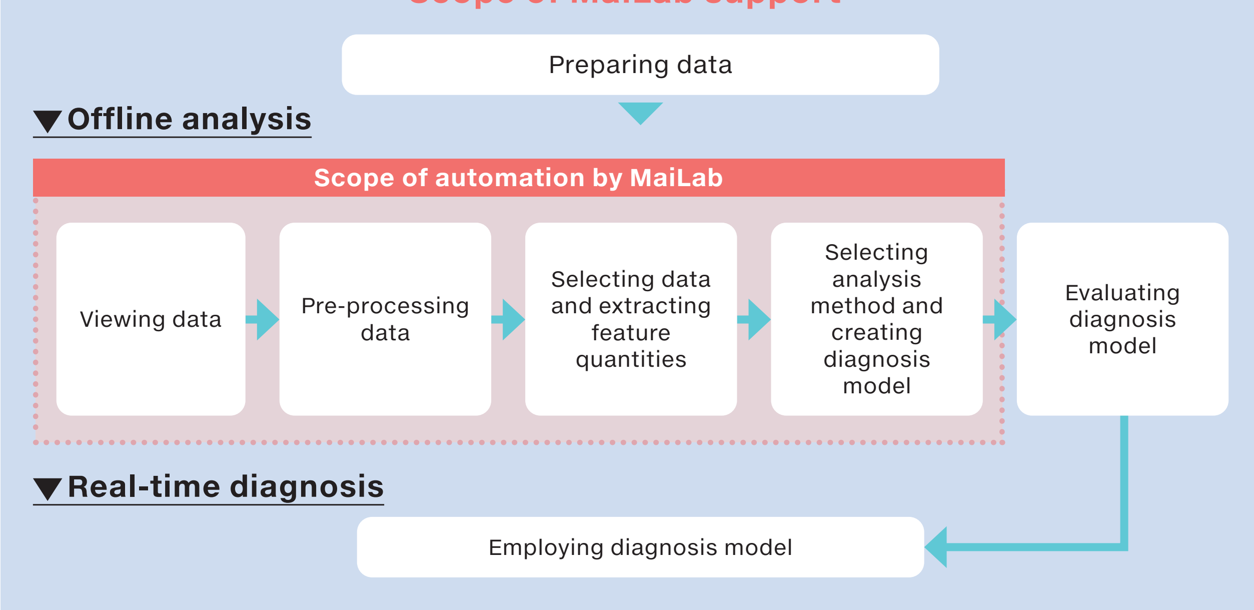
Data Analysis Textbook

chapter 1

Introduction

MailLab is software that serves as your data scientist and strongly supports you in utilization of data for solving problems. It performs data analysis, which is considered difficult and troublesome, completely automatically, and automatically generates the optimum diagnosis model for solving problems. Since the automatically generated diagnosis model can be used as is in the MaiLab environment, there's no need to construct a separate operating environment. This document explains procedures for easily using and experiencing MaiLab and methods for more effective utilization.

Scope of MaiLab support



Structure of this document

- **chapter 2 — Try MaiLab in action!**: Using sample data, you can experience going through the operation of MaiLab.
- **chapter 3 — Using MaiLab in various ways**: Go to this chapter for how to perform data analysis and diagnosis model creation completely automatically.
- **chapter 4 — Creating an original diagnosis model**: Go to this chapter for how to perform from data analysis to diagnosis model creation based on user judgments instead of completely automatically.
- **chapter 5 — Improving the accuracy of the diagnosis model**: Explains concepts and methods for improving the automatically created diagnosis model.
- **chapter 6 — How to create diagnosis models with higher accuracy**: Explains how to proceed with data analysis and the concepts involved. You can deepen your knowledge of data analysis and utilize MaiLab more effectively.

index

Terms

Term	Explanation
AutoML function	Function in which the user just inputs data and Machine Learning is automatically performed to create a diagnosis model.
Diagnosis model	Learning results created using learning. During real-time diagnosis, predictions are made based on the learned content by inputting operation data.
Offline analysis	Indicates all phases including visualization, pre-processing, and machine learning performed on data. Offline analysis is performed before starting operation, and a diagnosis model is created as the results.
Real-time diagnosis	Indicates the phase of performing predictions during operation using the diagnosis model.
Objective variable	Variables within the data that are prediction subjects. When performing supervised learning, these variables are necessary as data used for offline analysis.
Explanatory variables	Variables within the data that are used for prediction of objective variables. The diagnosis model receives the explanatory variables during learning and makes predictions based on the learned content.

Trademarks and Registered Trademarks

- “EDGECROSS” is registered trademark of Edgecross Consortium,a general incorporated association.
- QR Code is either a registered trademark or a trademark of DENSO WAVE INCORPORATED in the United States, Japan, and/or other countries.
- The company names, system names and product names mentioned in this document are either registered trademarks or trademarks of their respective companies.
- In some cases, trademark symbols such as ‘™’ or ‘®’ are not specified in this document.

chapter 2

Try MaiLab in action!

In this chapter, sample data will be used and the following procedures will be explained.
· Creation of a diagnosis model using the automatic analysis of the AutoML (Automatic Machine Learning) function.
· Real-time diagnosis using the diagnosis model

2.1 Before starting

Check the outline of MaiLab operating procedure and terms.

2.2 What to prepare

Prepare the sample data.

2.3 Using MaiLab

Create a diagnosis model using the AutoML function and execute the created diagnosis model.

- Create data set.
- Create task.
- Create AI.
- Execute task.

Check in the preview display.
3.1.2 Visualizing the created data set

Using AutoML with your own data
Chapter 3 Using MaiLab in various ways

Chapter 5 Improving the accuracy of the diagnosis model

Chapter 4 Creating an original diagnosis model

2.1 Before starting

The procedure for creating a diagnosis model using the AutoML function and executing the created diagnosis model is shown below. Use sample data and perform the following procedure.

1 Add a data set

Step 1. Data set creation

Import the data to be used for analysis to create the diagnosis model into MaiLab. A group of imported data is called a "data set". Data sources which can be imported are CSV-format and TSV-format text files.

2 Create the AI

Step 2. AI creation

The data's regularity and rules are derived through learning using the data set. A model that enables diagnosis of unknown data is called "AI" in MaiLab. In MaiLab's AutoML function, when what you want to do (objective) is selected, an "AI" is automatically created.

3 Create the task

Step 3. Task creation

Settings for performing diagnosis of unknown data using the created AI are called a "task". Define the data input/output methods and threshold values for OK/NG judgments of diagnosis results.

4 Execute it

Step 4. Task execution

Execute the task and monitor the diagnosis status of unknown data. The AI operation can be checked in the GUI using the monitor function. (Whether data are flowing according to the settings, good or bad judgment status, etc.)

About the created diagnosis model

In this chapter, sample data will be used to create a diagnosis model for inferring wine quality.
The sample data are the scientific data related to quality for 4535 bottles of white wine published by UCI. Each line is the data for 1 bottle of wine, and consists of 11 items indicating component values and 1 item for quality, for a total of 12 items. Wine quality is categorized into 3 classes: Class_0, Class_1, and Class_2.

Sample data structure

Component values (explanatory variables)

Objective variable

Tartaric acid concentration	Acetic acid concentration	Citric acid concentration	Residual sugar concentration	Chloride concentration	Free sulfurous acid concentration	Sulfurous acid concentration	Density	pH	Sulfate concentration	Alcohol by volume	Wine quality
7	0.27	0.36	20.7	0.045	45	170	1.001	3	0.45	8.8	Class_1
8.1	0.28	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1	Class_1
7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.4	9.9	Class_1
8.1	0.28	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1	Class_1
7	0.27	0.36	20.7	0.045	45	170	1.001	3	0.45	8.8	Class_1
6.3	0.3	0.34	1.6	0.049	14	132	0.994	3.3	0.49	9.5	Class_1
8.1	0.27	0.41	1.45	0.033	11	63	0.9908	2.99	0.56	12	Class_0

Reference data: http://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/
The above data has been processed for the explanation.

By performing supervised learning using wine quality as the objective variable and the 11 items of component values as the explanatory variables, a diagnosis model to infer quality from the component values will be created completely automatically.

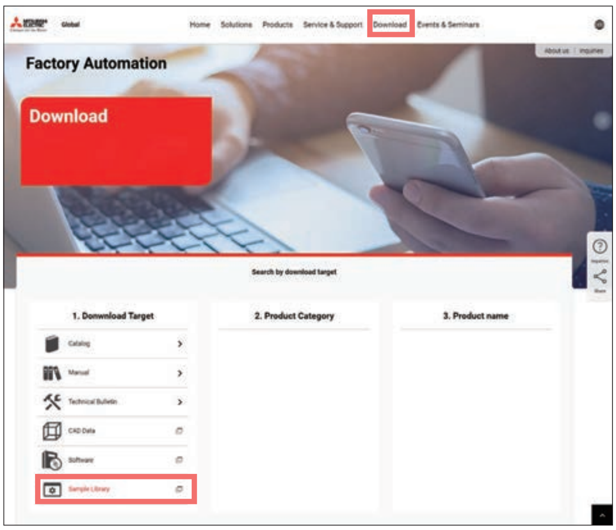
2.2 What to prepare

Step 1. Sample data preparation

Download the sample data from the MITSUBISHI ELECTRIC FA Global Website and unzip the downloaded file.

■ MELSOFT MaiLab Data Analysis Textbook Sample Data download procedure

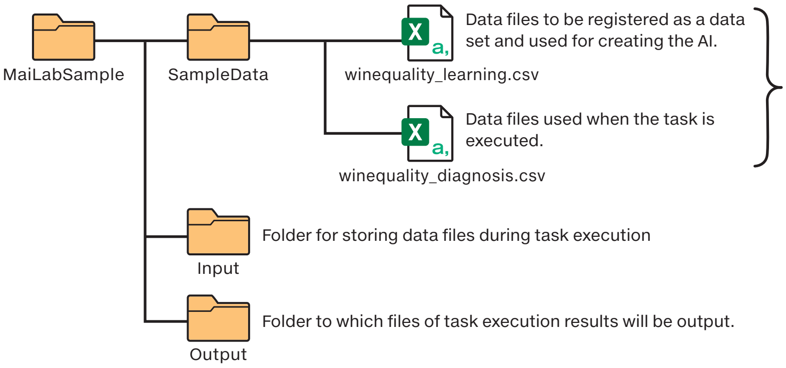
MITSUBISHI ELECTRIC FA Global Website top page
https://www.mitsubishielectric.com/fa/
Download → Sample Library



Sample data	Download file
MELSOFT MaiLab Data Analysis Textbook Sample Data	MaiLabSample.zip

Step 2. Check the folders and files

Check that the unzipped folder structure and files are as shown below.

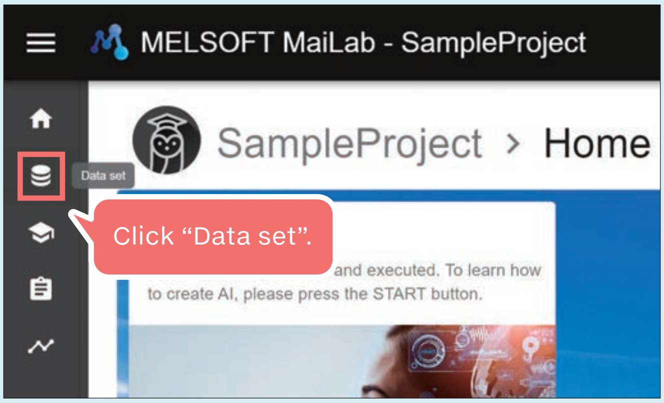


Divide the data for the 4535 bottles. Freely select the data for 3614 bottles and store them in “winequality_learning” and store the data for the remaining 921 bottles in “winequality_diagnosis”.
In “winequality_diagnosis”, delete the objective variable “wine quality”.

2.3 Using MaiLab

Step 1. Data set creation

1

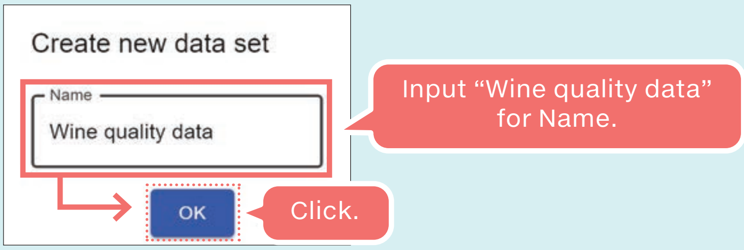


Click “Data set” in the side bar.

2

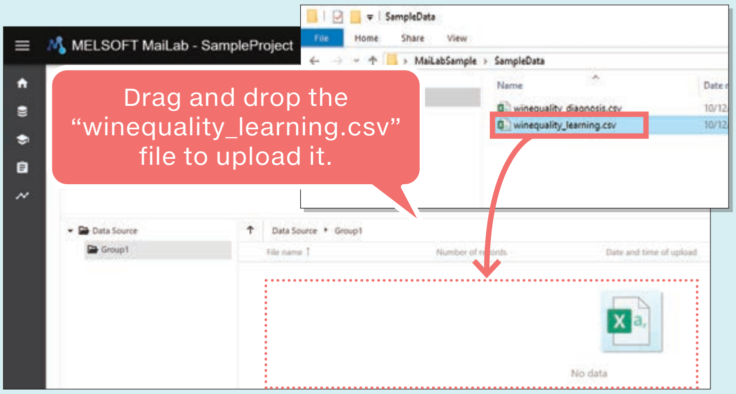
In the Data Set Management screen, click the “Create new” button.

3



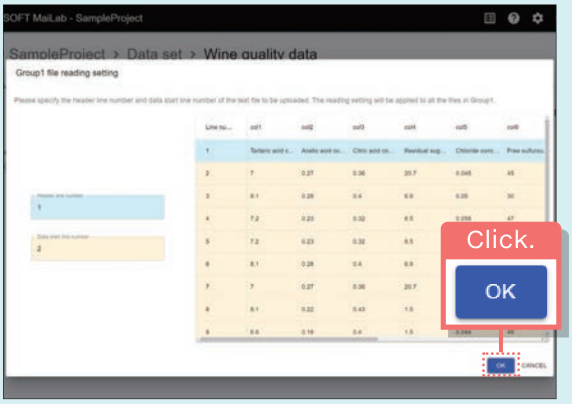
In the Create new data set dialog, input “Wine quality data” for Name and click the “OK” button.

4



Upload the previously prepared “winequality_learning.csv” file.

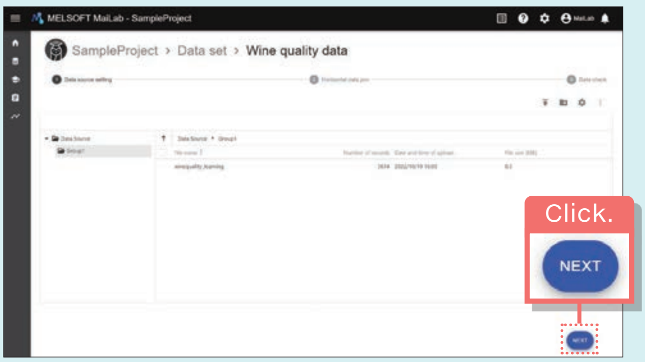
5



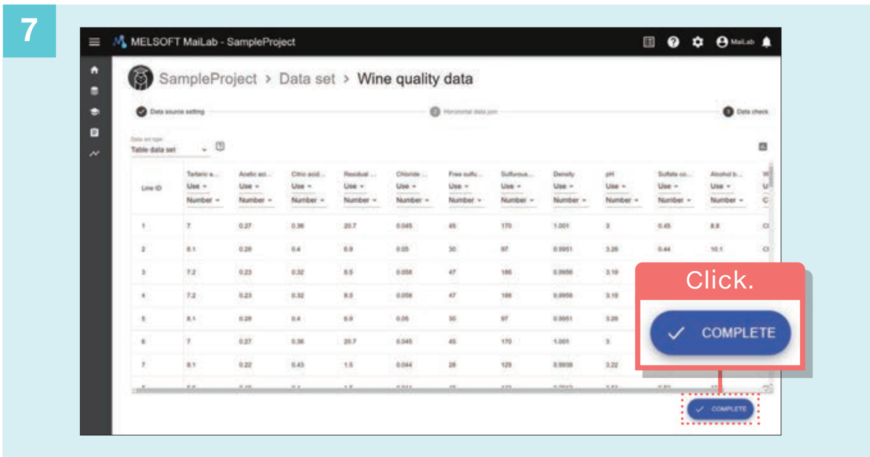
The file reading setting screen will appear. Click the “OK” button.

* The details of displayed items and setting content are explained in chapter 3. In this section, click the “OK” button without changing any settings.

6

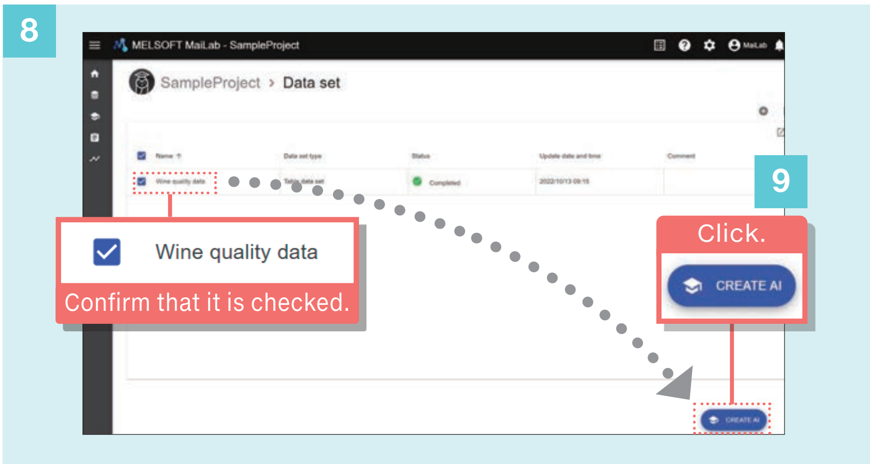


The uploaded CSV file will be displayed. Click the “NEXT” button.



The program will proceed to the Data Check screen. Click the “COMPLETE” button.

* The details of displayed items and setting content are explained in chapter 3. In this section, click the “COMPLETE” button without changing any settings.

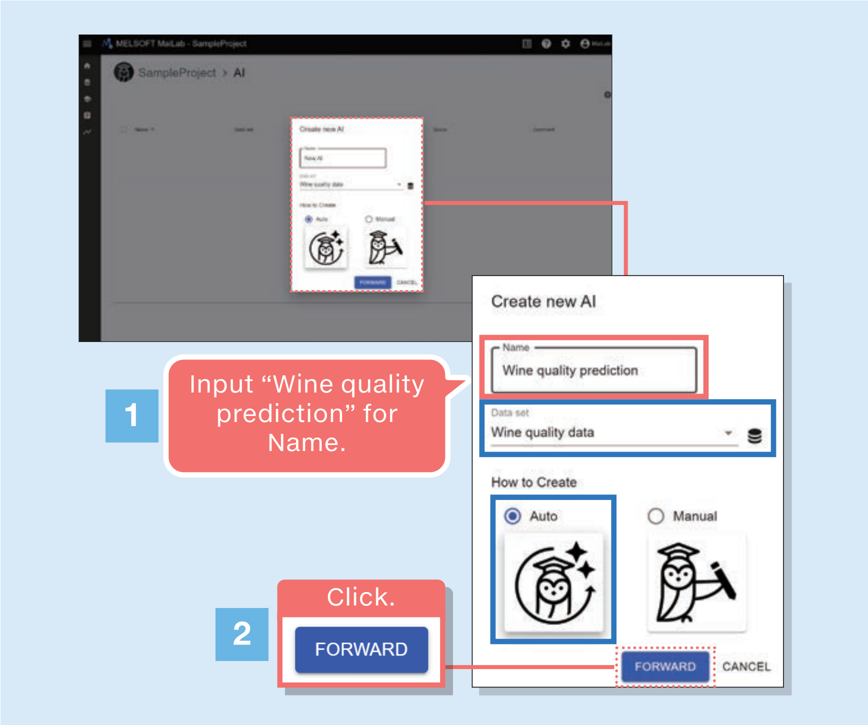


8 Data set creation has been completed.

9 Confirm that “Wine quality data” is checked and click the “CREATE AI” button.

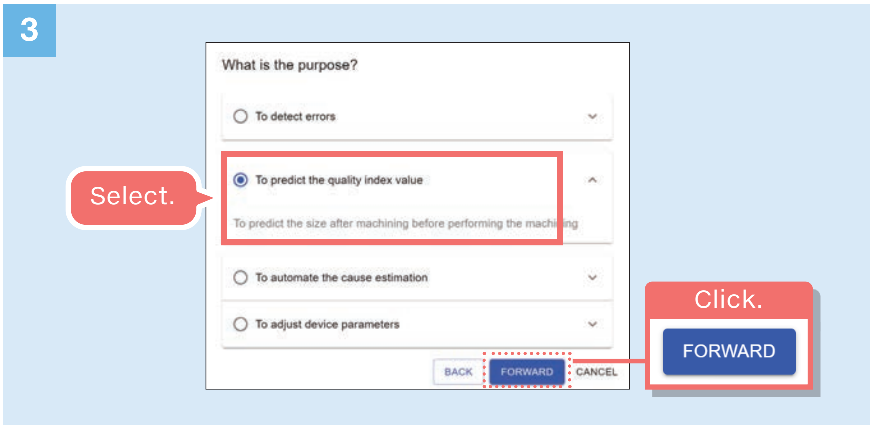
Create the AI

Step 2. AI creation

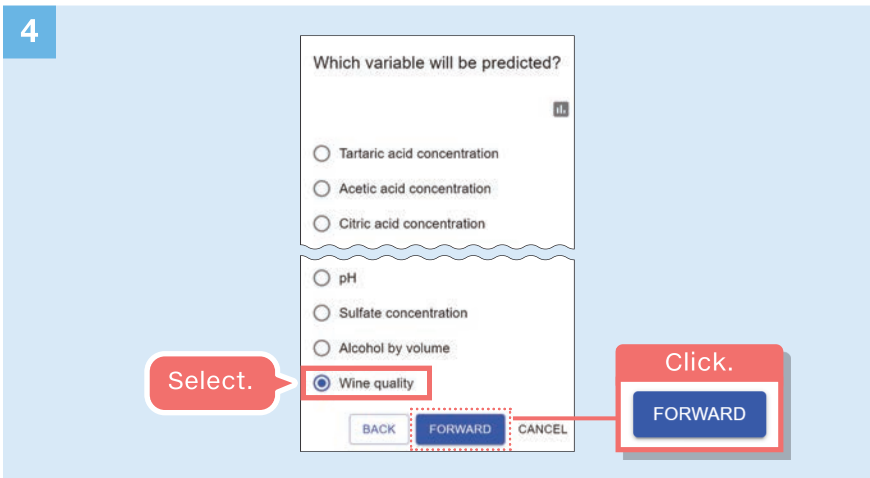


1 In the Create New AI dialog, input “Wine quality prediction” for Name.

2 Confirm that “Wine quality data” is selected for Data set and “Auto” is selected for How to Create, and click the “FORWARD” button.

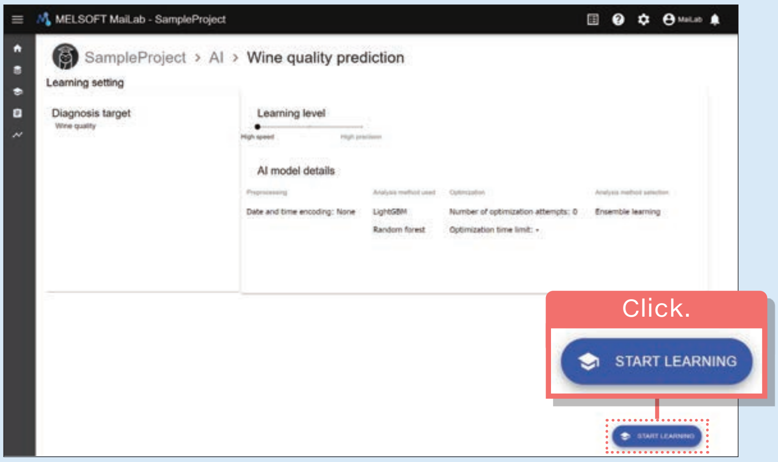


Select “To predict the quality index value” for Purpose and click the “FORWARD” button.



Select “Wine quality” for Variable to be predicted and click the “FORWARD” button.

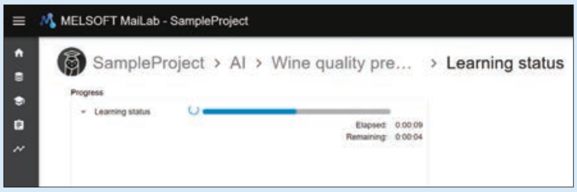
5



The program will proceed to the Learning screen. Click the "START LEARNING" button.

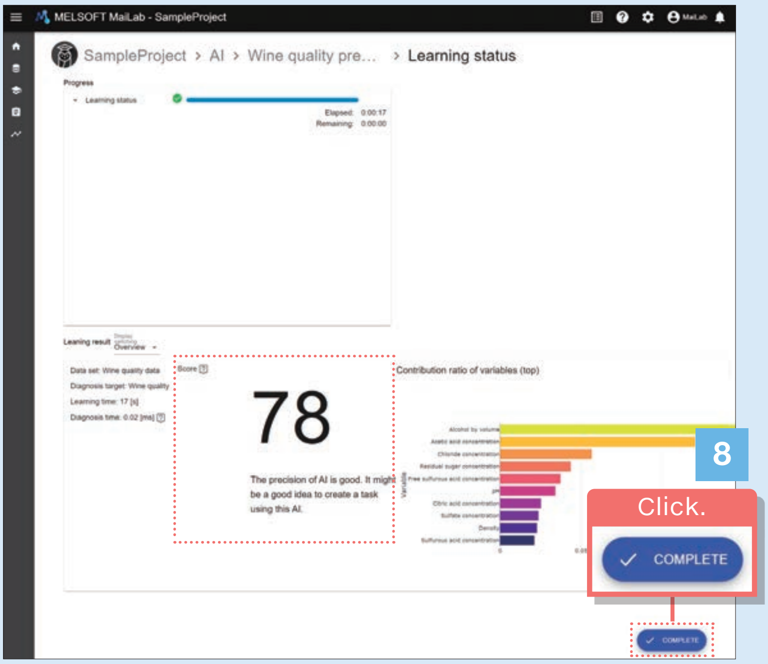
* The details of displayed items and setting content are explained in chapter 3. In this section, click the "START LEARNING" button without changing any settings.

6



Learning will start, and the learning progress will be displayed.

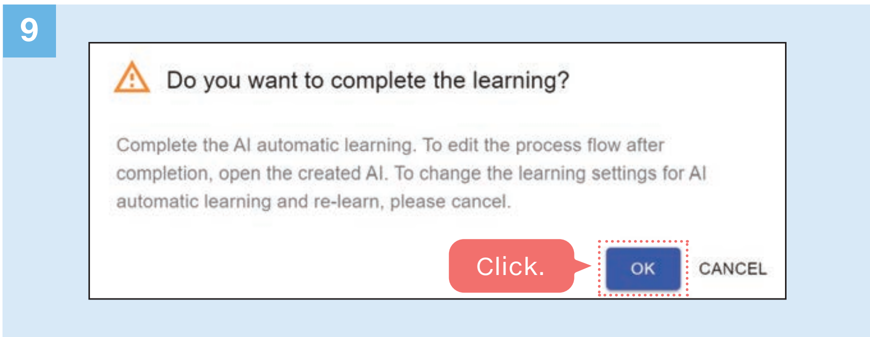
7



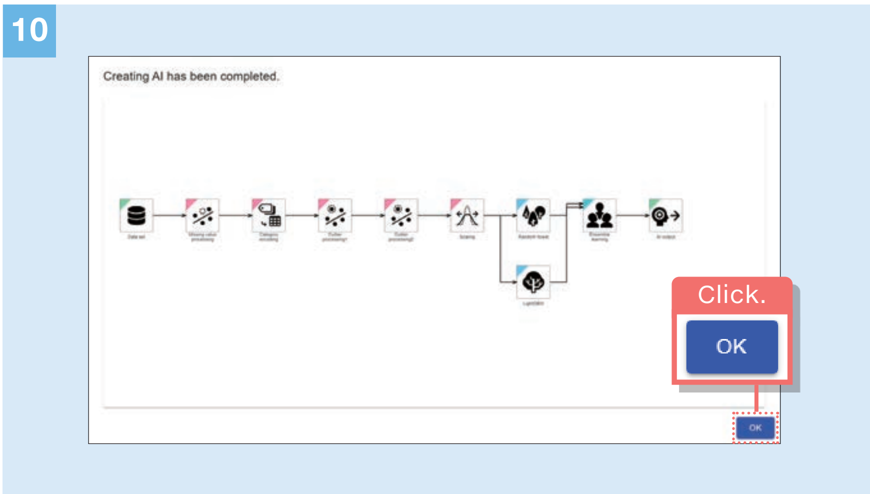
7 When learning has been completed, the learning results (scores) will be displayed.

The data set is divided into verification data and test data, and learning is performed. Since there is some randomness in the data division and parameter settings, the score will not always be 78.

8 Click the "COMPLETE" button.

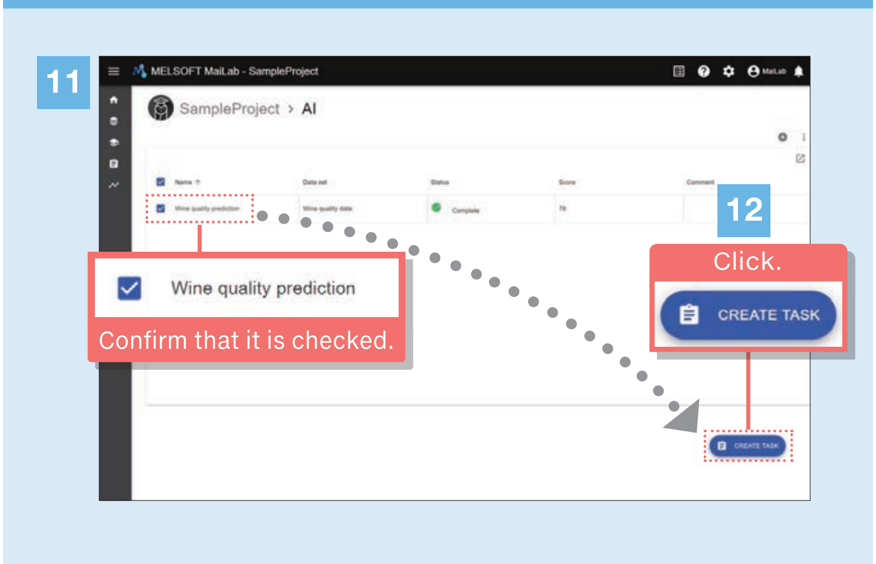


Click the “OK” button.



The created AI will be displayed. Click the “OK” button.

AI creation has been completed.

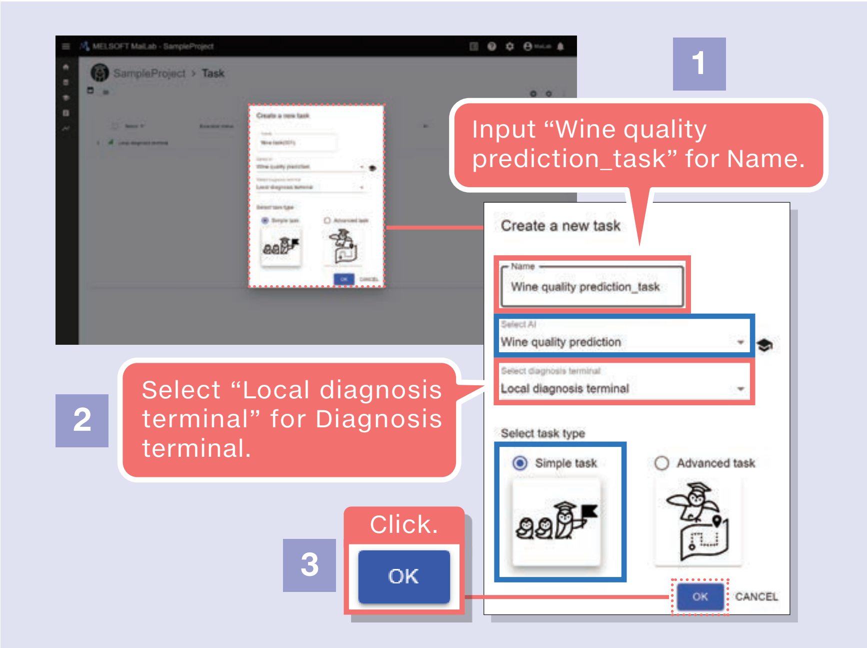


11 AI creation has been completed.

12 Confirm that “Wine quality prediction” is checked and click the “CREATE TASK” button.

Create task

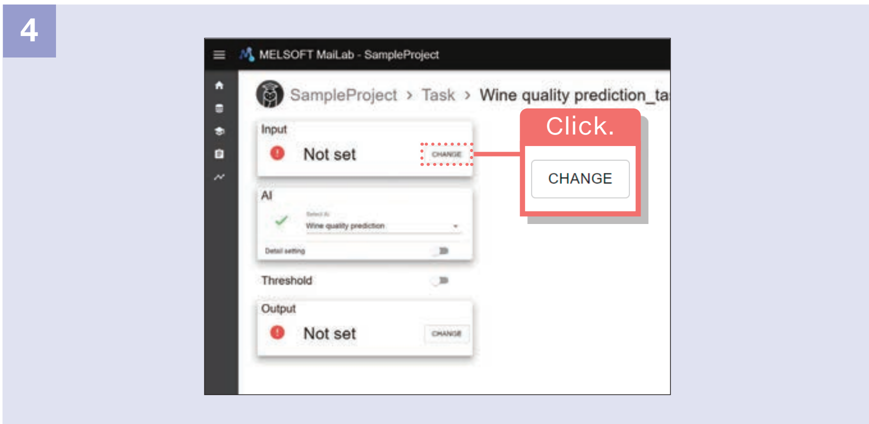
Step 3. Task creation



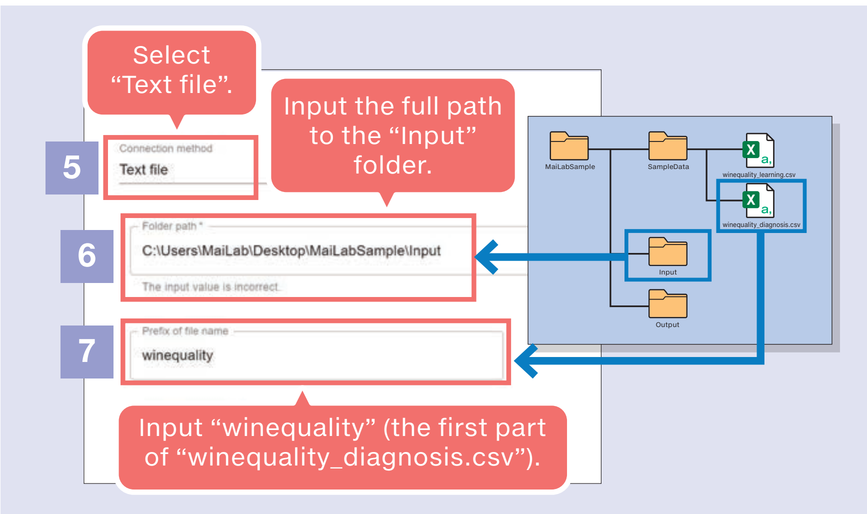
1 In the Create New Task dialog, input “Wine quality prediction_task” for Name.

2 Select “Local diagnosis terminal” for Diagnosis terminal.

3 Confirm that “Wine quality prediction” is selected for the AI and “Simple task” is selected for the Task type, and click the “OK” button.



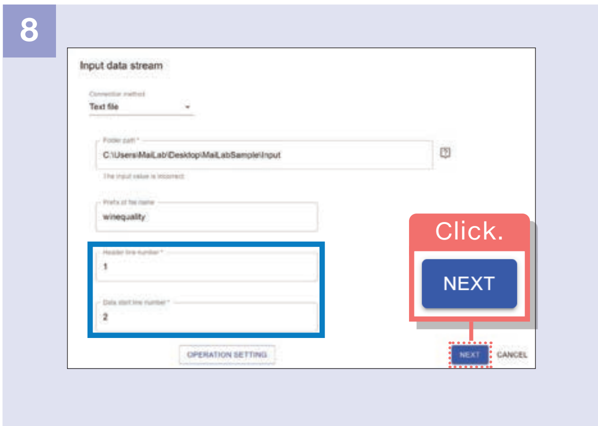
The Create Simple Task screen will appear. Click the “CHANGE” button in Input.



5 The input data stream dialog will appear. Select “Text file” for Connection method.

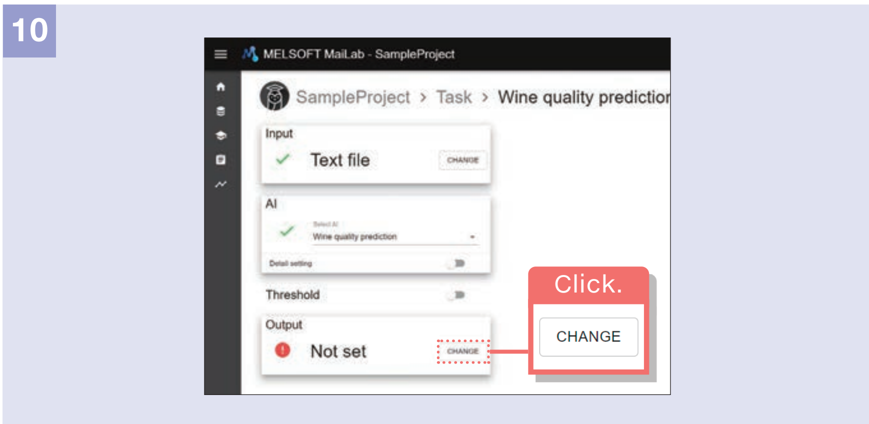
6 Input the full path to the previously prepared “Input” folder of the unzipped files for Folder path.

7 Input “winequality” (the first part of “winequality_diagnosis.csv”) for Prefix of file name.

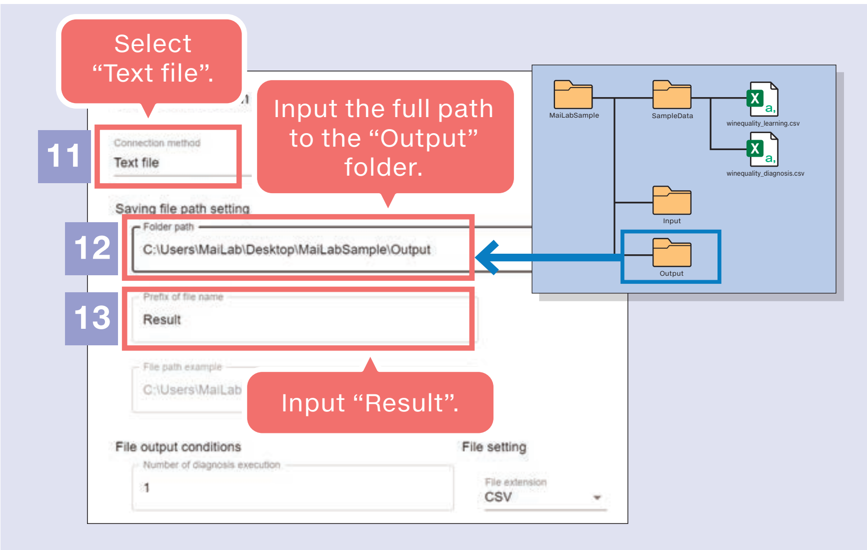


Confirm that the following have been set, and click the “NEXT” button.
• Header line number: 1
• Data start line number: 2

The program will switch to the Data assignment setting dialog. Click the “OK” button.



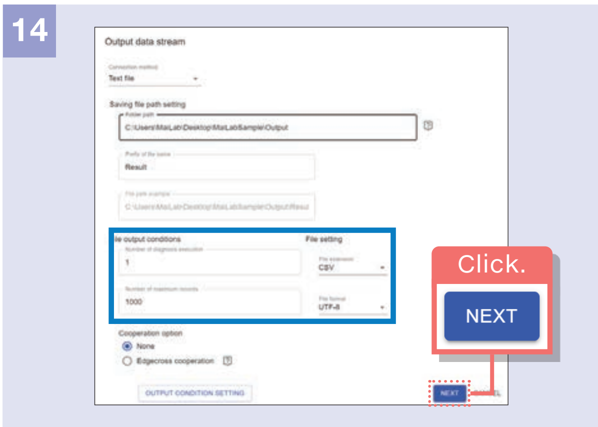
Click the Output “CHANGE” button.



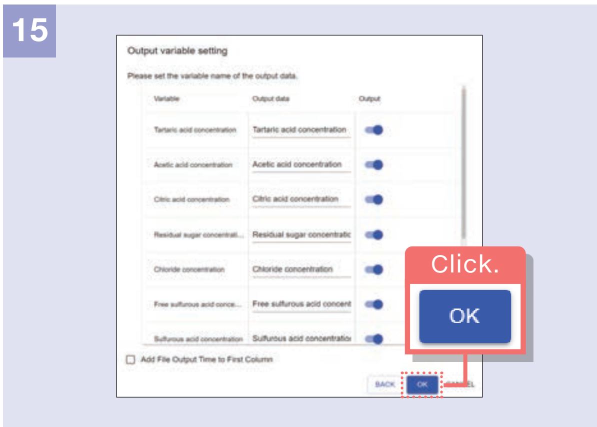
11 The Output data stream dialog will appear. Select “Text file” for Connection method.

12 Input the full path to the “Output” folder for the Saving file path setting.

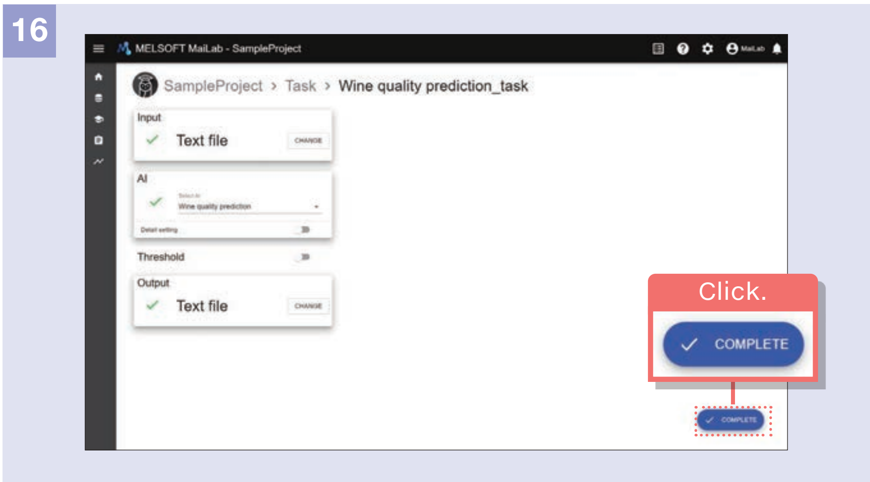
13 Input “Result” for Prefix of file name.



Confirm that the following have been set, and click the “NEXT” button.
- Number of diagnosis execution: 1
- Number of maximum records: 1000
- File extension: CSV
- File format: UTF-8

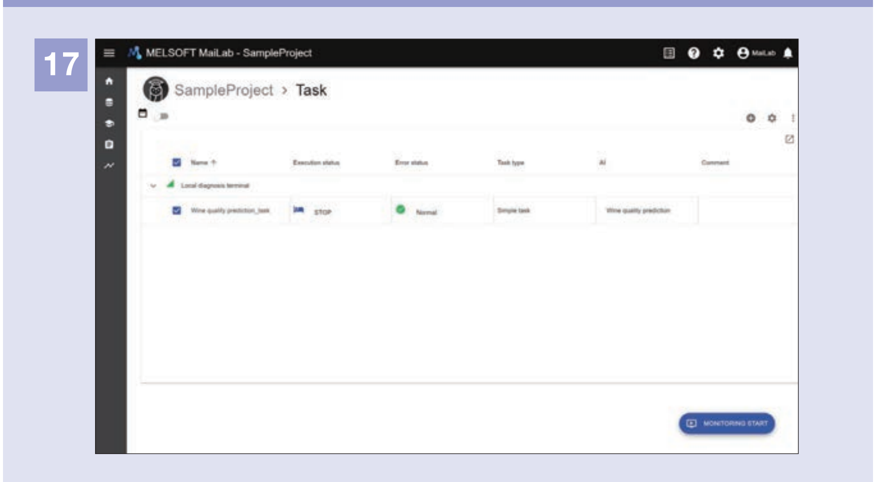


The program will switch to the Output variable setting dialog. Click the “OK” button.



Click the “COMPLETE” button.

Task creation has been completed.

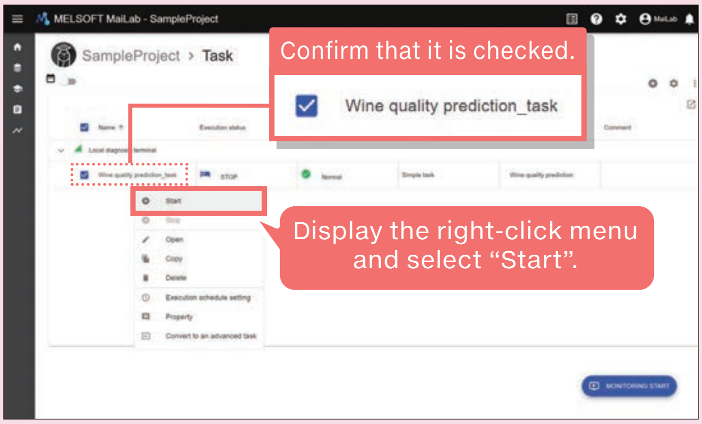


Task creation has been completed.

Execute the task

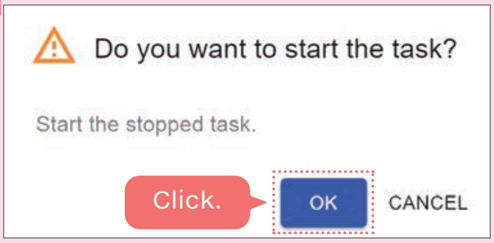
Step 4. Task execution

1



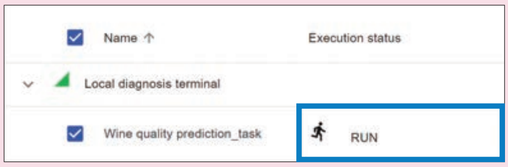
Confirm that "Wine quality prediction_task" is checked, and select "Start" from the right-click menu.

2



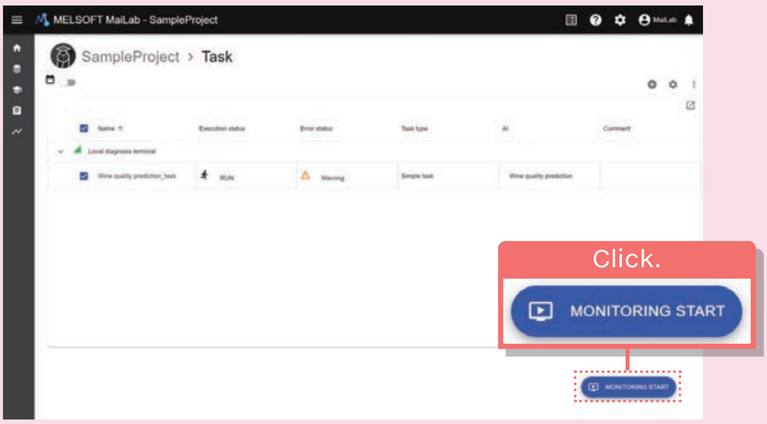
In the Task Start Confirmation dialog, click the "OK" button.

3



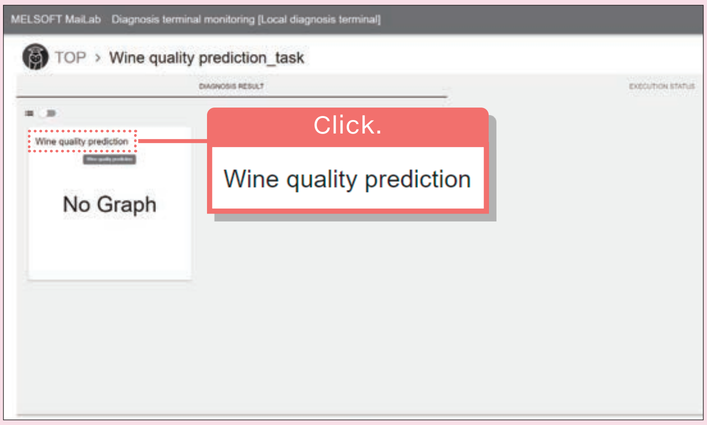
Confirm that the Task execution status changes to "RUN".

4



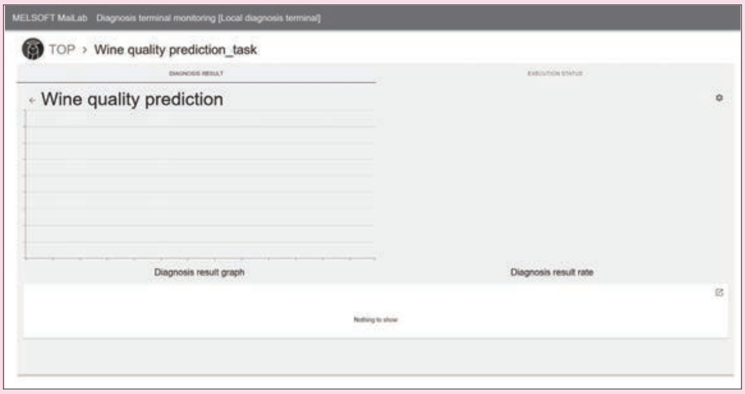
Click the "MONITORING START" button.

5



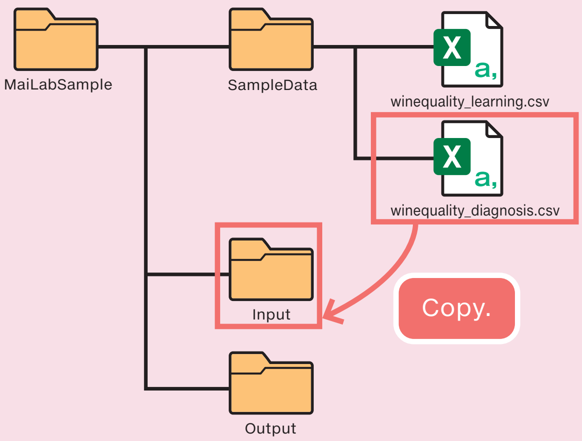
The monitor screen will be shown in a separate browser tab. Click the “Wine quality prediction” tab.

6



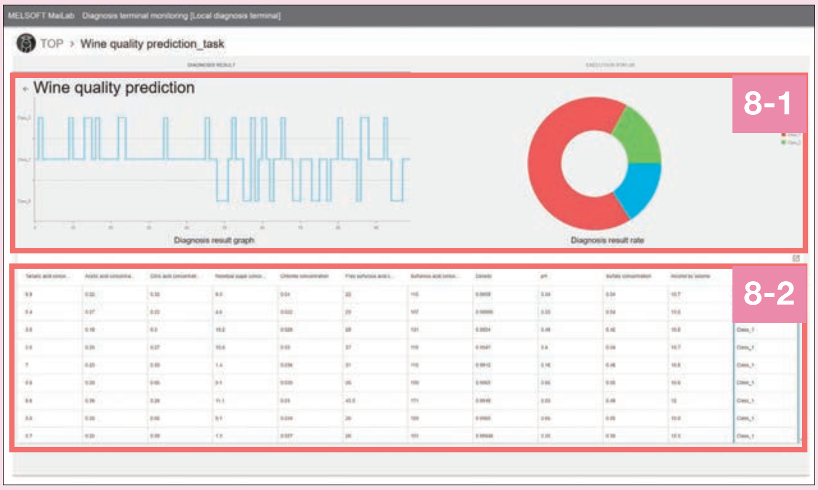
The Wine quality prediction task diagnosis results monitor screen will be shown.

7



Copy the previously prepared “winequality_diagnosis.csv” file to the “Input” folder for data input.

8



8 The diagnosis execution results will be shown in the Diagnosis terminal monitoring screen.

8-1 The results of category division of “winequality_ diagnosis.csv” by AI are shown in line graphs and pie charts.

8-2 Category results and data input to AI are shown in table format.

chapter 3

Using MaiLab in various ways

Various procedures for performing diagnosis using the AI created by the AutoML function will be explained.

3.1 Creating the data set → 3.2 Creating the AI → 3.3 Executing tasks using the created AI

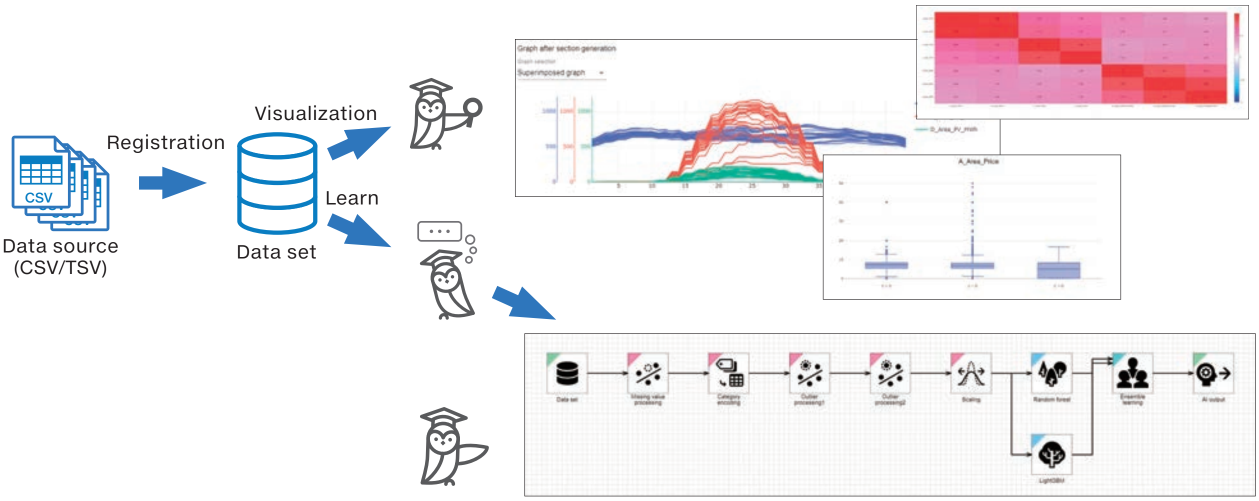
3.1 Creating the data set

What is a data set?

In order to analyze the data and create the diagnosis model, the data subject to analysis is registered in MaiLab. A group of registered data is called a “data set”. By registering the data set, the data can be visualized in tables or graphs, and diagnosis models (AI) can be created.

Data set specifications

Item	Explanation	Remarks
Maximum number of variables	256 variables	
Maximum number of records	864,000 rows	
Maximum number that can be created	128	Maximum number for 1 project
Maximum size that can be created	2 GB	Total size for 1 project



Data source

The original file of data to be registered as a data set is called the “data source”. Data sources which can be registered are CSV-format and TSV-format text files.

Data source specifications

Item	Explanation
File extension	.csv, .tsv
Supported character codes	UTF-8, Shift-JIS
Maximum number of characters for variable name	255 characters
Maximum number of characters for each data	255 characters
Maximum size for 1 file	1 GB

■ Data source structure

The data source structure consists of “header rows” containing the data names (variable names) of each column, and “data rows” containing the data.

Data source example

[LOGGING]	-	3	4	5
Comment 1				
DATETIME	STRING[8]	SHORT[DEC.0]	SHORT[DEC.0]	...
Time	Product ID	Current	Temperature	...
Comment 2				
2022/03/03 12:00:00	Prod1	5	40	...
2022/03/03 12:00:01	Prod1	3	38	...
2022/03/03 12:00:02	Prod1	3	45	...
2022/03/03 12:00:03	Prod1	4	50	...
:	:	:	:	:

Header rows:
Rows containing the data names (variable names) Must be within the range of rows 1 to 19.

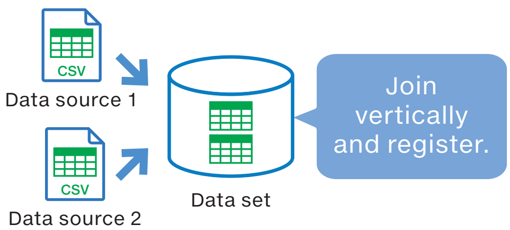
Data rows:
Consists of 2 or more rows.
Since the first row is the row right after the header rows, it must be within the range of rows 2 to 20.

■ Joining data sets

When the data sources are multiple files, data sources can be joined with each other and registered as a single data set.

• Vertical join

The data rows of multiple data sources with the same structure are joined vertically and registered as a single data source.



Join example

Data source 1

[LOGGING]	-	3	4	5
Comment 1				
DATETIME	STRING[8]	SHORT[DEC.0]	SHORT[DEC.0]	...
Time	Product ID	Current	Temperature	...
Comment 2				
2022/03/03 12:00:00	Prod1	5	40	...
2022/03/03 12:00:01	Prod1	3	38	...
2022/03/03 12:00:02	Prod1	3	45	...
2022/03/03 12:00:03	Prod1	4	50	...
:	:	:	:	:

Data source 2

[LOGGING]	-	3	4	5
Comment 1				
DATETIME	STRING[8]	SHORT[DEC.0]	SHORT[DEC.0]	...
Time	Product ID	Current	Temperature	...
Comment 2				
2022/03/04 12:00:00	Prod2	5	40	...
2022/03/04 12:00:01	Prod2	3	38	...
2022/03/04 12:00:02	Prod2	3	45	...
2022/03/04 12:00:03	Prod2	4	50	...
:	:	:	:	:

Join

The data rows of data sources with matching header rows are joined vertically.

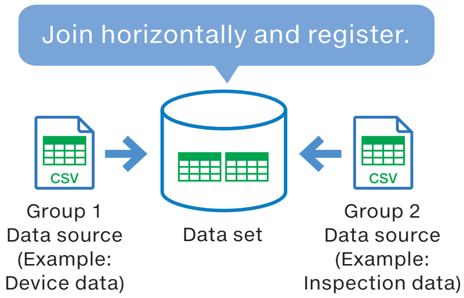
Data set

Time	Product ID	Current	Temperature	...
2022/03/03 12:00:00	Prod1	5	40	...
2022/03/03 12:00:01	Prod1	3	38	...
2022/03/03 12:00:02	Prod1	3	45	...
2022/03/03 12:00:03	Prod1	4	50	...
:	:	:	:	:
2022/03/04 12:00:00	Prod2	5	40	...
2022/03/04 12:00:01	Prod2	3	38	...
2022/03/04 12:00:02	Prod2	3	45	...
2022/03/04 12:00:03	Prod2	4	50	...
:	:	:	:	:

• Horizontal join

Two kinds of data sources are connected and joined, and registered as a single data set. This is used in cases such as connecting both “device data” measured by sensors at the time of manufacture and “inspection data” recorded from inspections after manufacturing, and performing learning.

Based on a specified join key (in the join example, “Product ID”), the data in rows with matching keys are joined horizontally.
Variables with the same name are specified as the join key. When there are groups of data sources with multiple files, join each group vertically and then join the groups horizontally.



Join example

Group 1 data sources

[LOGGING]			
DATETIME	STRING[8]	SHORT[DEC.0]	SHORT[DEC.0]
Time	Product ID	Current	Temperature
2022/03/03 12:00:00	Prod1	5	40
2022/03/03 12:00:01	Prod1	3	38
2022/03/03 12:00:02	Prod1	3	45
:	:	:	:
:	Prod2	4	39
:	Prod2	5	40
:	:	:	:
:	Prod3	5	38
:	Prod3	3	40
:	:	:	:

Join key

Group 2 data sources

Product ID	Inspection results	Judgment
Prod1	150	Good
Prod2	155	Bad
Prod3	151	Good

Join key

Join

Data set

Time	Product ID	Current	Temperature	Inspection results	Judgment
2022/03/03 12:00:00	Prod1	5	40	150	Good
2022/03/03 12:00:01	Prod1	3	38	150	Good
2022/03/03 12:00:02	Prod1	3	45	150	Good
:	:	:	:	150	Good
:	Prod2	4	39	155	Bad
:	Prod2	5	40	155	Bad
:	:	:	:	155	Bad
:	Prod3	5	38	151	Good
:	Prod3	3	40	151	Good
:	:	:	:	151	Good

1:1 and N:1 (1:N) joining can be performed.

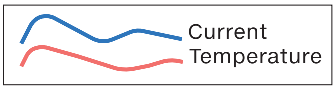
Data set types

There are 2 types of data sets: Waveform data sets and Table data sets.

• Waveform data set

Data that has the meaning of sequentiality, such as measurement data that changes continuously with the passage of time.
Data for which the record (row) order cannot be changed.
It corresponds to data that are continuously measured using sensors mounted on devices, etc.

Time	Product ID	Current	Temperature	...
0:00:00	Prod1	5	40	...
0:00:01	Prod1	3	38	...
0:00:02	Prod1	3	45	...
0:00:03	Prod1	4	50	...
...	...	...	...	...

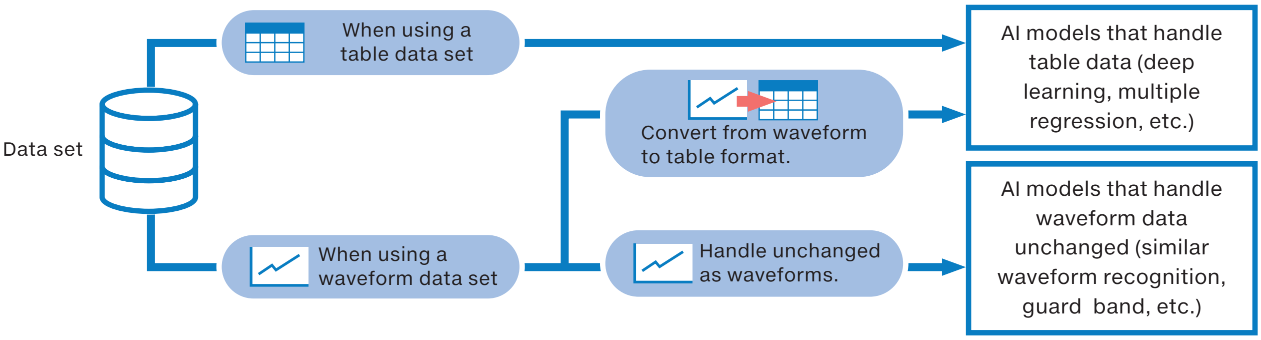


• Table data set

Data for which even if the record (row) order is changed, it does not change the overall meaning. It corresponds to per-factory or per-product production data, data in which the inspection results for individual products are recorded, etc.

Factory	Product name	Production quantity	...	Profits
Factory A	Servo	1000	...	5
Factory A	PLC	2000	...	18
Factory B	Servo	500	...	2
Factory B	PLC	1200	...	20
...	...	...	...	...

Details of what is specified when creating the AI will differ depending on the data set type.
In addition, the details of what is specified will also be different when manually creating the AI.
When registering a data set, set the data set type suitable for the user.



Variable type

MaiLab handles 3 kinds of variables.
The variable types are set at the time of data set registration. Visualization methods and handling methods during AI creation will be different according to the variable type.

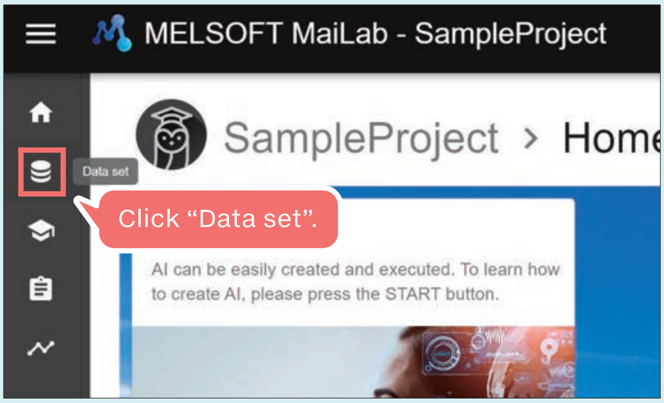
Variable type	Explanation	Example
Number	Data in which the numerical values have meaning as large or small and can be added, subtracted, or otherwise manipulated. When you want to predict numerical values using AI, set the objective variables as numerical value type.	• Temperature (-10°C, 15°C, 20°C, etc.) • Test points (20 points, 50 points, 95 points, etc.) • Current values (0.01 mA, 1.1 A, 100 A, etc.) • Pressure (1 mPa, 10 Pa, 1013 hPa, etc.)
Category	Data which represent a category or classification, and cannot be used as is for addition/subtraction. Category type is used mainly for values which are character strings. Set category type in cases such as those in which even if the value is a number, it expresses an ID or classification.	• Survey results (1: Unsatisfied, 2: Average, 3: Satisfied) • Blood type (Type A, Type B, Type O, Type AB) • Lot number (A0001, A0002, etc.) • Status (0: Normal, 1: Appearance defect, 2: Internal defect, etc.)
Timestamp	Data which expresses a time connected with data, such as the data collection time, etc. It can be used as information for expressing the sequentiality of data in easy-to-understand visualization, for performing data processing, etc. It cannot be used for objective variables.	• YYYY/MM/DD • YYYY-MM-DD • MM/DD/YYYY • hh:mm:ss.fff • hh:mm:ss • YYYY/MM/DD hh:mm:ss.fff • hh:mm:ss.fff YYYY/MM/DD, etc.

3.1.1 Creating the data set

Upload the data source and create a new data set. When creating a data set from a single kind of data source, execute only Step 1.
When horizontally joining 2 different kinds of data sources, execute Step 1 and then Step 2 in order.

Step 1. Create a data set from a single kind of data source

1

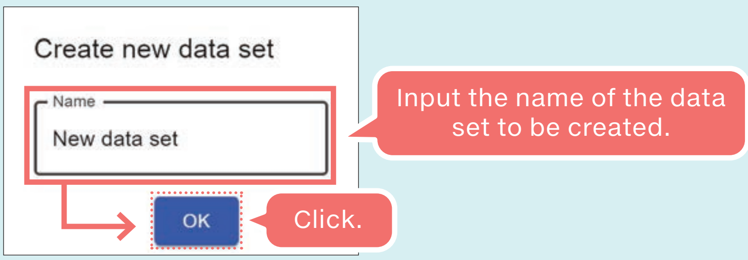


Click “Data set” in the side bar.

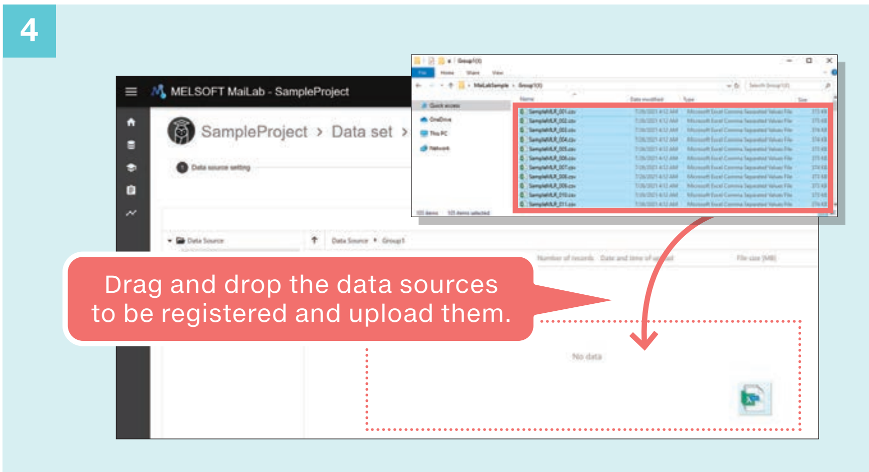
2

In the Data Set Management screen, click the “Create new” button.

3

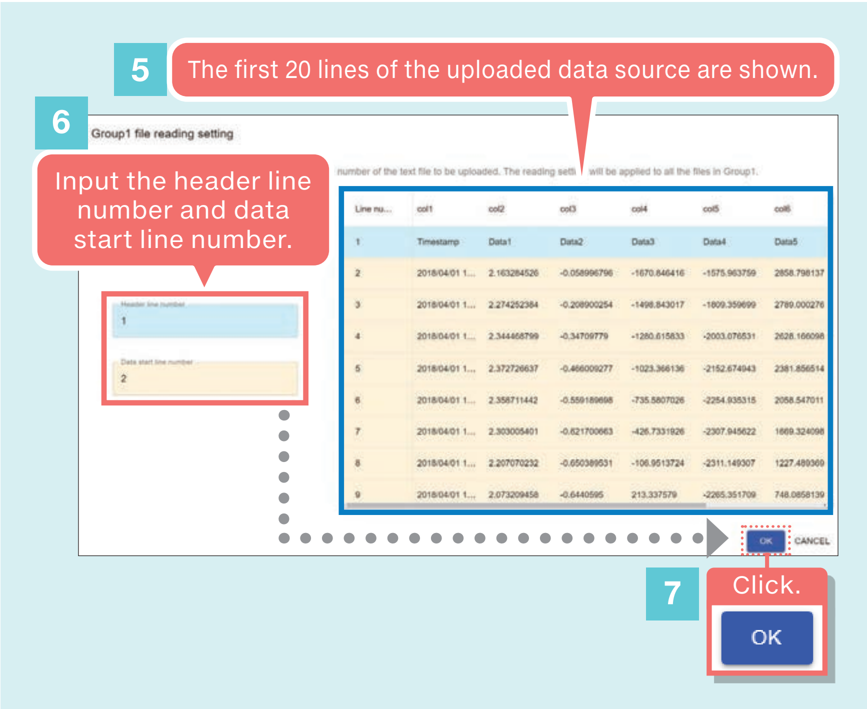


In the Create new data set dialog, input the name of the data set to be created and click the “OK” button.



Drag and drop the data sources to be registered onto the data source setting screen and upload them.

* A maximum of 1,000 files can be uploaded at 1 time.



5 In the data source setting screen, the first 20 lines of the uploaded data source are shown in a popup.

6 Input the "Header line number" and "Data start line number" of the data source.

7 Click the “OK” button.

Header line number and Data start line number

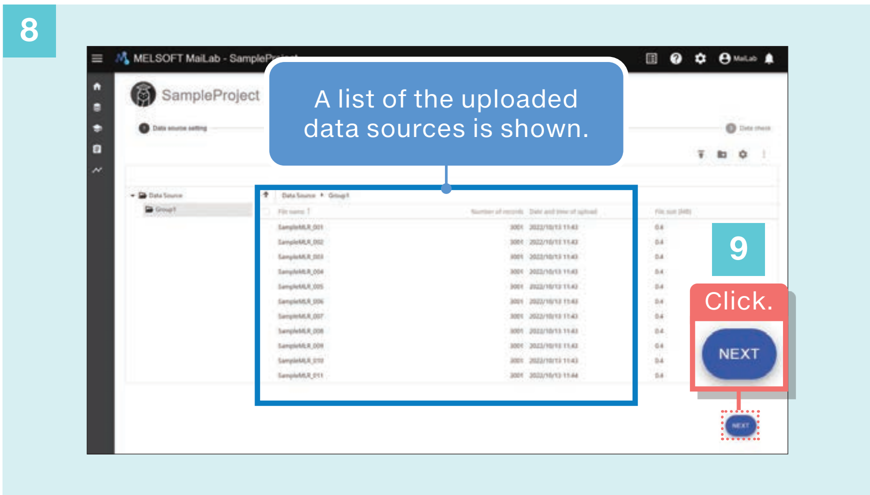
Example

	A	B	C
1	[LOGGING]	1	2
2	DATETIME[YYYY/MM/DD hh:mm:ss.sss]	DOUBELE[DEC.14]	DOUBELE[DEC.14]
3	Timestamp	Data1	Data2
4	2018/05/01 00:00:00.000	2.185884657	-0.060188205
5	2018/05/01 00:00:01.000	2.298011811	-0.213118883
6	2018/05/01 00:00:02.000	2.368961787	-0.354107244
7	2018/05/01 00:00:03.000	2.397514839	-0.475420085
8	2018/05/01 00:00:04.000	2.383353225	-0.570482235

Header row

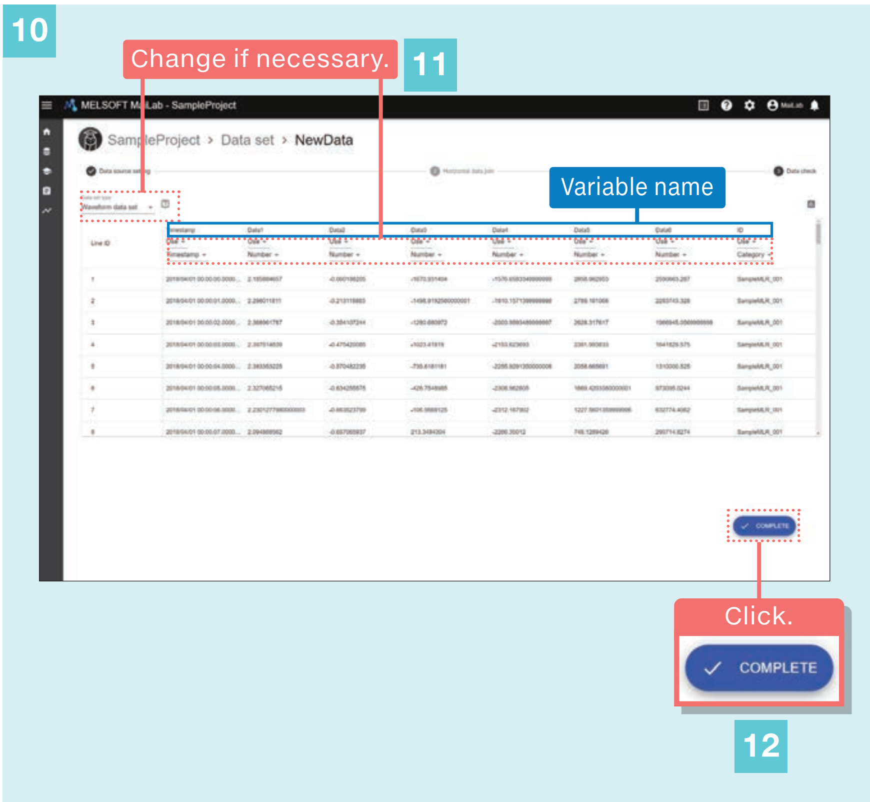
Data rows

Header line number: 3
Data start line number: 4



8 A list of the uploaded data sources is shown.

9 If there are no mistakes in the data sources, click the “NEXT” button.

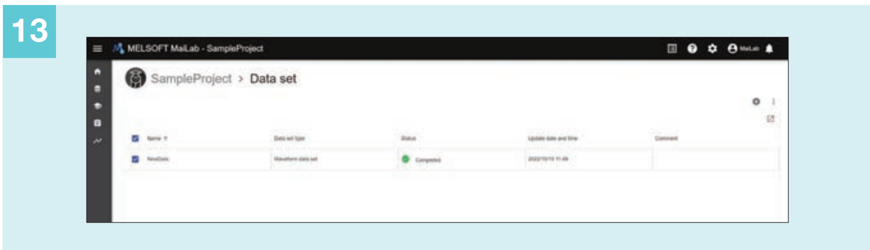


10 The results of joining data sources vertically are shown.

11 If necessary, change the following items:

- Data set type: Specify “Table data set” or “Waveform data set”.
- Variable use/not use: If they are not used for visualization or AI creation, specify “Not use”.
- Variable type: Specify “Number”, “Category”, or “Timestamp” (The selectable variables will be different depending on data contents.)

12 Click the “COMPLETE” button.



13 Data set creation has been completed.

When joining data sets horizontally, go to Step 2.

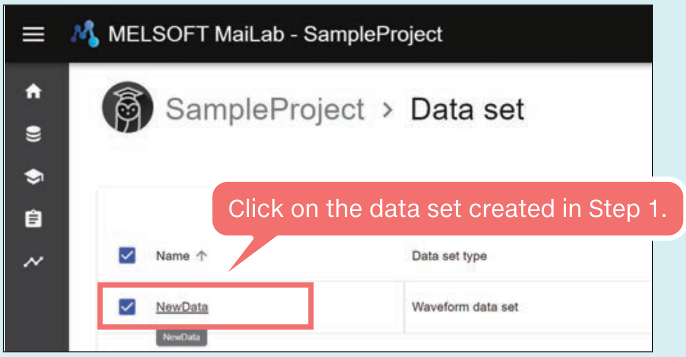
For visualization (graph display) of the created data set, go to 3.1.2.

For creating the AI, go to 3.2.

Step 2. Horizontally join a second kind of data source with the data set created in Step 1

This is the procedure for creating a data set by horizontally joining 2 different kinds of data sources.
Here, the method for adding the second kind of data source to the data set created in Step 1 and joining them horizontally will be explained.
The variable name of the join key is “ID”, and it is included in both the data set created in Step 1 and the second kind of data source.

1

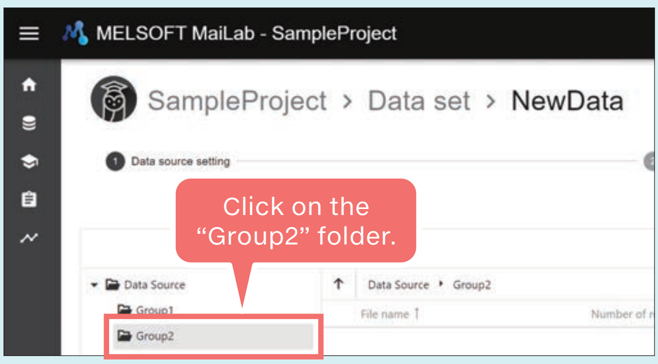


In the Data Set Management screen, click on the data set created in Step 1.

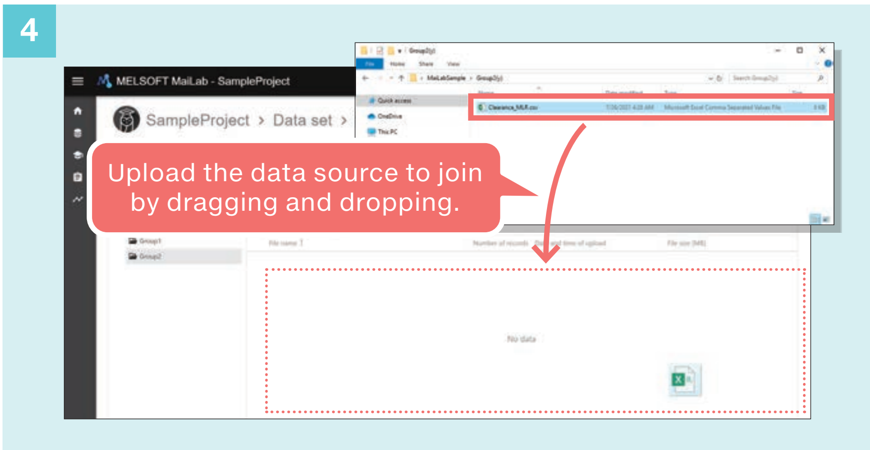
2

Click the “Create new folder” button.

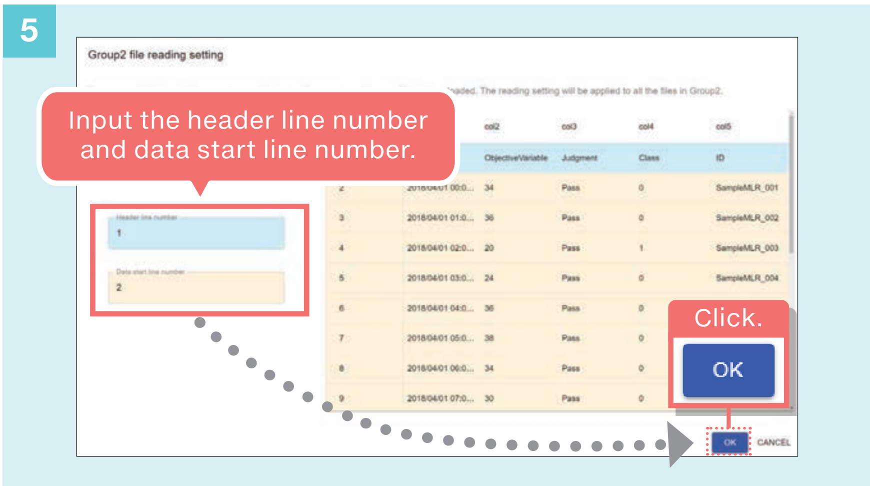
3



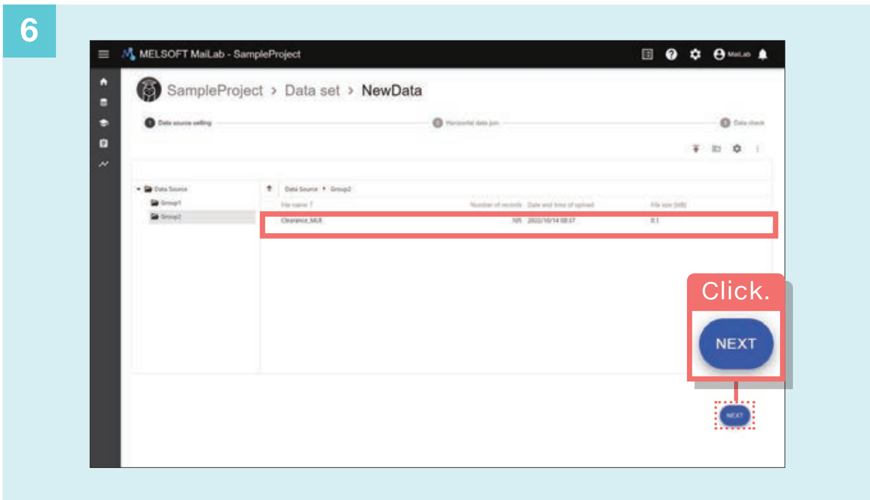
Click on the newly created “Group2” folder.



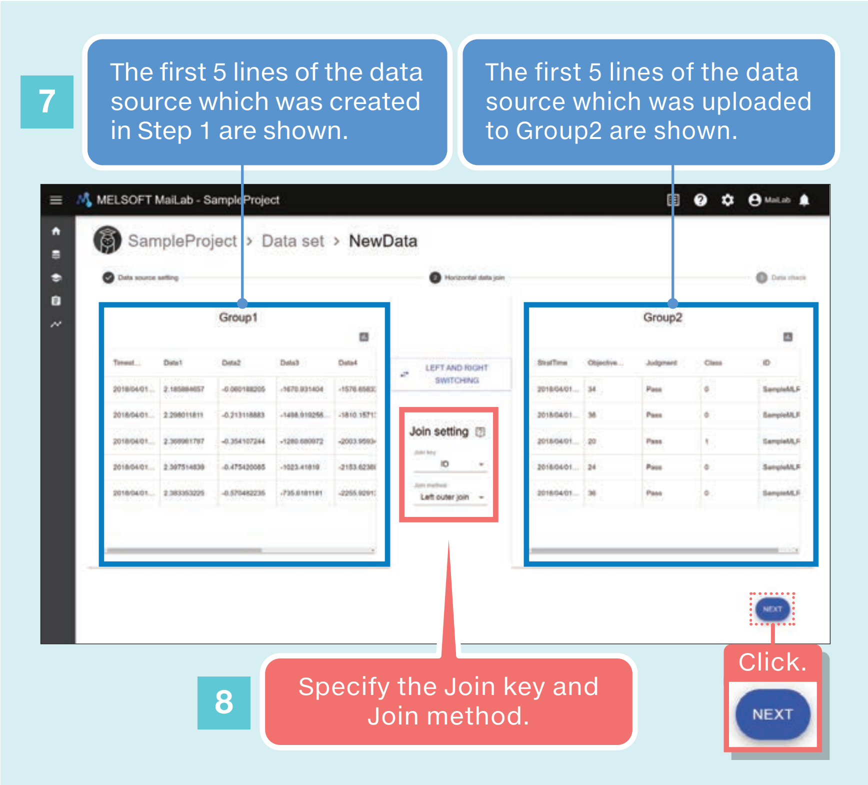
Upload the data source to join to the Group2 folder.



Input the "Header line number" and "Data start line number" of the data source and click the “OK” button.



The uploaded data source will be shown in the list. If there are no mistakes, click the “NEXT” button.



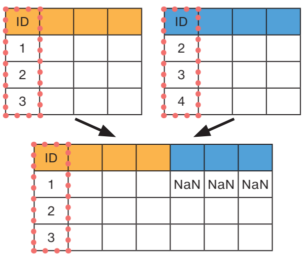
7 The first 5 lines of Group1 (the data set which was created in Step 1) and Group2 (the data set which was uploaded this time) will be shown in the horizontal join screen.

8 Specify the "Join key" and "Join method", and click the “NEXT” button.

Join method

Left outer join

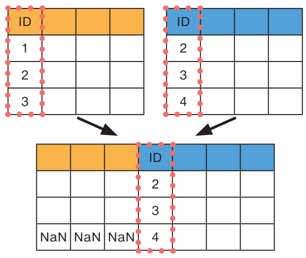
Method in which the data on the right are joined based on the join key on the left.



If there is no join key on the right that matches, the value will be null.

Right outer join

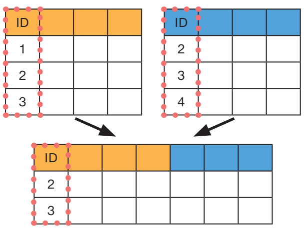
Method in which the data on the left are joined based on the join key on the right.



If there is no join key on the left that matches, the value will be null.

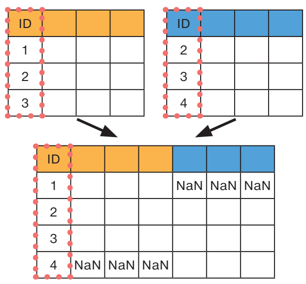
Inner join

Method in which only data which match the join key on both left and right are joined.

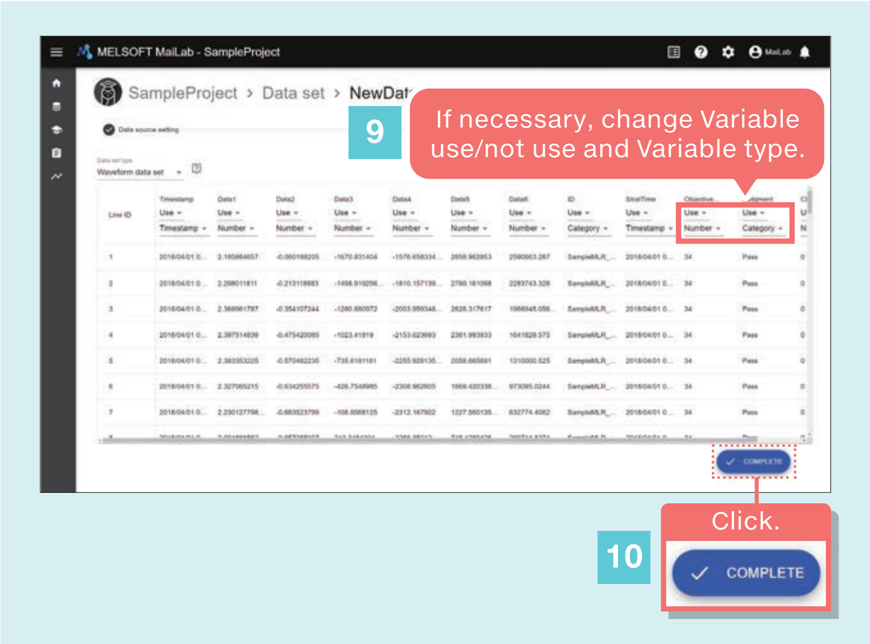


Full outer join

Method in which all data whether it matches each other or not are joined.

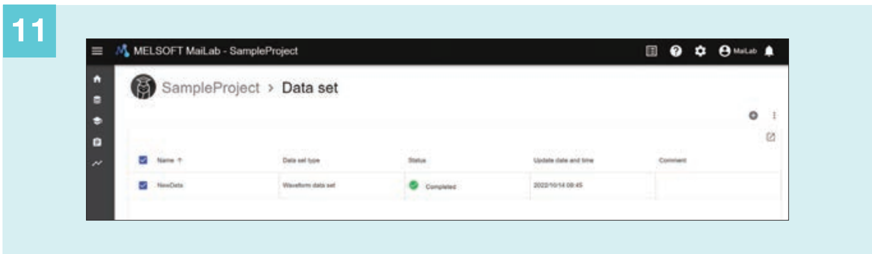


If there is no join key that matches, the value will be null.



9 The horizontal join results will be shown. If necessary, change "Variable use/not use" and "Variable type".

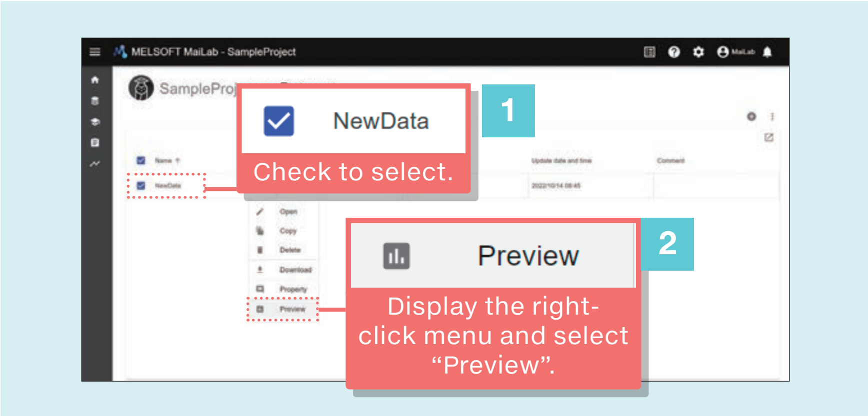
10 Click the “COMPLETE” button.



Creation of the horizontally joined data set has been completed.

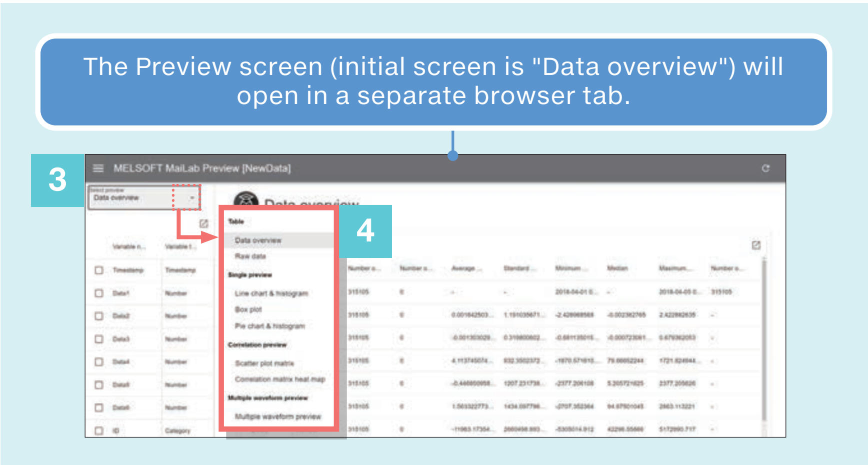
3.1.2 Visualizing the created data set

Visually check the created data set using the preview function to visualize it.
Visualization can be performed in various formats in MaiLab.



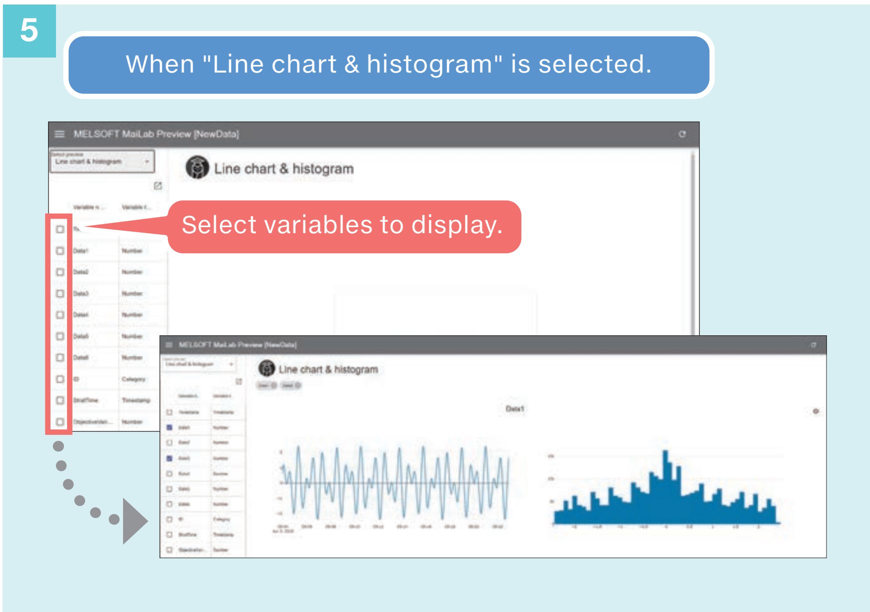
1 In the Data Set Management screen, select the data set to preview.

2 Select "Preview" from the right-click menu.



3 The Preview screen will be shown in a separate browser tab.

4 From the Preview selection pulldown menu, select the graph type.



The preview screen for the selected graph will be shown. Select the variables to graph.

Applications and viewing methods for each graph to answer questions such as "When is performing visualization effective?", "What can be understood from which graph?", "What actions can be taken from what is understood?", etc. will be explained in chapter 5: "Improving the accuracy of the diagnosis model".

3.2 Create the AI

By performing learning using the data set, the regularity and rules of the data will be derived and unknown data can be diagnosed. A model that enables diagnosis of unknown data is called “AI” in MaiLab.

AI creation

The 2 methods for creating AI are as follows:

Auto (Select from objectives.)

Create the AI interactively.

Select this when “What you want to do (objective)” is clear, but you don’t know what analysis method to use. MaiLab selects the optimum pre-processing and analysis methods based on the objectives and data set contents, and automatically creates the AI.

Manual (Select from methods.)

Select the analysis method and create the AI.

Select this when “What you want to do (objective)” and the method appropriate for the objectives is clear.

In this section, the procedure for creating the AI by “Auto” will be explained.

In “Auto”, select “What you want to do (objective)” from the following 4 categories, and proceed interactively.

Objective 1 To detect errors

The aim is predictive maintenance of device/equipment by detecting signs of device failure due to wearing of parts. It can be applied even when data sets consist of only normal values.

Use case	Issue
Preventive maintenance of thick-plate hydraulic press equipment	For hydraulic presses, recovery after abnormal stoppage requires a lot of time. Detect symptoms of abnormalities in advance and perform predictive maintenance.
Predicting occurrence of defects in molded plastic products	Prevent molding defects by foreseeing the occurrence of defects and performing maintenance instead of the conventional way of performing maintenance after a problem occurs.

Objective 2 To predict quality index value

Select to predict the future, such as the predicting the post-processing dimensions before processing. The aim is to improve production efficiency (improve yield), etc. The data set must have a “prediction target (objective variable)”.

Use case	Issue
Predictive detection of deposition defects in electronic components	Reduce the flow of defective products to downstream processes by improving film quality judgment accuracy based on the average vacuum level during deposition.

Objective 3 To automate the cause estimation

Select to identify main causes, such as by automating the failure cause estimation that was previously performed by experienced personnel. The data set must have a “prediction target (objective variable)”.

Use case	Issue
Estimation of failure location when an abnormality occurs on the device.	Reduce down time when device abnormality occurs by quickly identifying the failure location without relying on the experience of personnel.

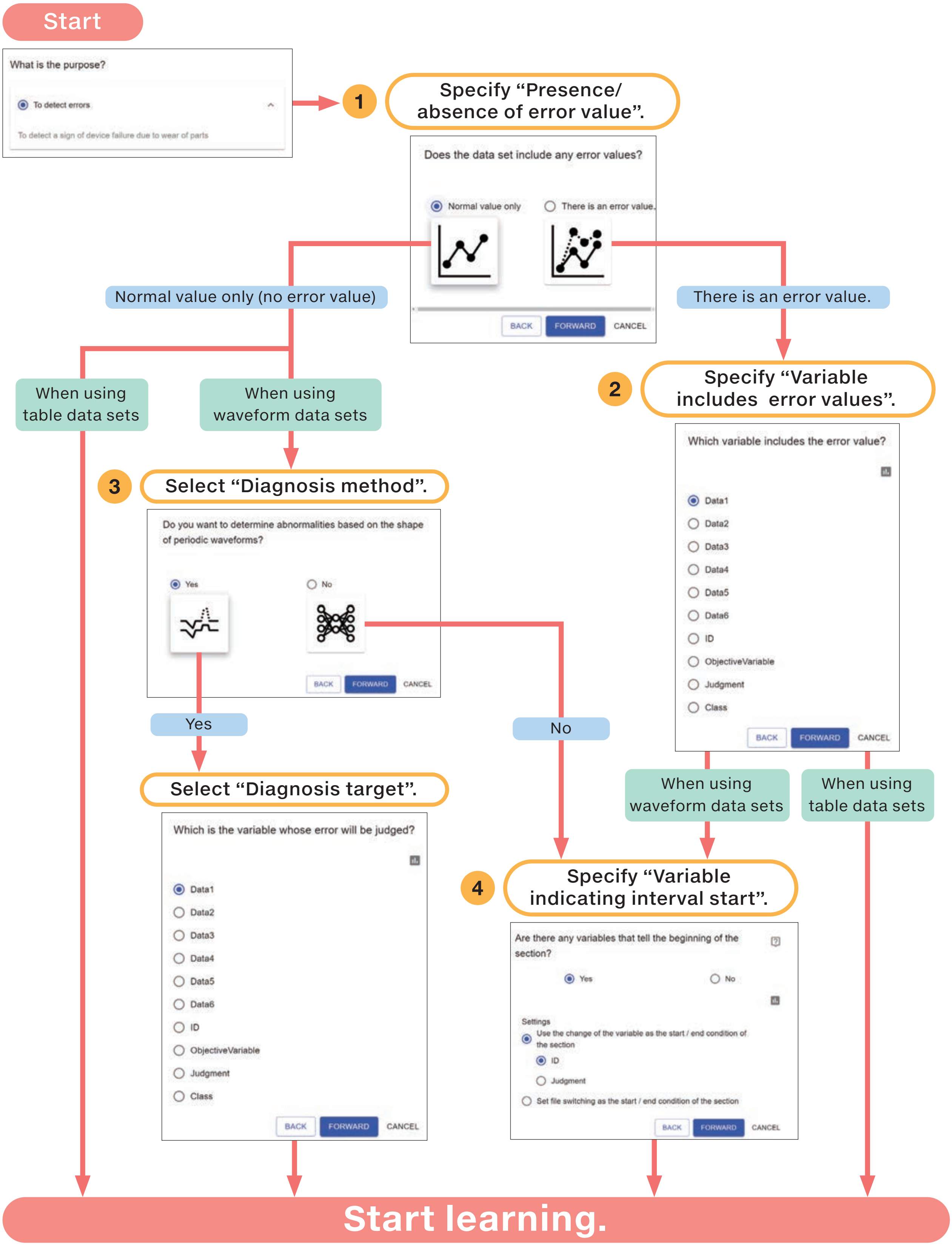
Objective 4 The adjust device parameters

The aim is stable production that does not rely on personnel by inheriting the knowhow of experienced personnel through automation of parameter tuning, etc. that previously relied on intuition and experience.The data set must have an “adjustment target (objective variable)”.

Use case	Issue
Automatic adjustment of welding condition setting values	Eliminate the trial-and-error setting work based on operator experience and intuition by automatically extracting welding conditions according to the condition of the welding object.

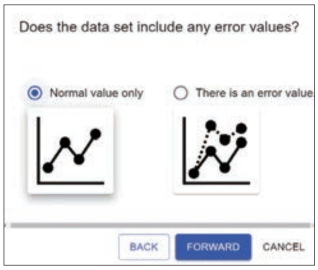
3.2.1 For the case of “To detect errors”

An AI that will detect symptoms of abnormality occurrence based on equipment/devices operating data, etc. will be interactively created. It can be applied for unsupervised learning even when data at time of abnormality could not be collected. An outline of the interactive flow for “To detect errors” is shown in the figure below. Specify “Presence/absence of abnormal values” that will serve as training data, “Variables include abnormal data” if abnormal values are present, etc.



1 Specify "Presence/absence of error value".

Select whether or not the abnormal data that would serve as training data are included in the data set.
When normal data only is specified, an AI that detects "Different than usual" conditions that are different from normal conditions will be created.



2 Specify "Variable includes error values".

The variable that serves as the label classifying normal data and abnormal data.
A category type with binary format data (OK/NG, true/false, etc.) is required.

* If the binary values are numerical values such as 0/1, change the variable type to "Category" when creating the data set.

Example

Product ID	Processing condition 1	Processing condition 2	…	Product good/bad
Prod1	90	100		Good
Prod2	88	100		Good
Prod3	85	95		Bad
Prod4	90	98		Good
Prod5	92	110		Bad
:	:	:		:

Variable indicating abnormal

Time	Product ID	Current	Temperature	…	Good/bad judgment
2022/03/03 12:00:00	Prod1	5	40		Good
2022/03/03 12:00:01	Prod1	3	38		Good
2022/03/03 12:00:02	Prod1	3	45		Good
:	:	:	:		Good
:	Prod2	4	39		Bad
:	Prod2	5	40		Bad
:	:	:	:		Bad

Normal data

Abnormal data

Variable indicating abnormal

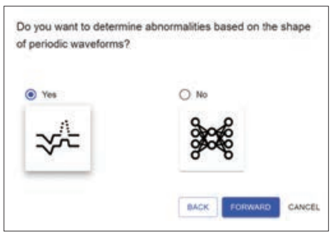
If the variable type is numerical value, as in the data set shown at right, when estimating the value create the AI from "To predict quality index value".

Product ID	Processing condition 1	Processing condition 2	…	Inspection results
Prod1	90	100		150
Prod2	88	100		151
Prod3	85	95		155
Prod4	90	98		152
Prod5	92	110		156
:	:	:		:

Variable indicating abnormal

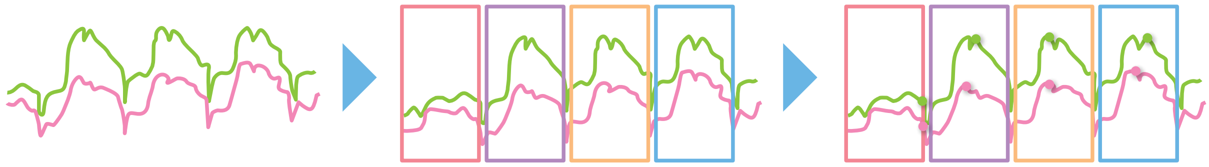
3 Select "Diagnosis method": When using waveform data sets

When using a waveform data set and the data set is "Normal value only", the select dialog will be shown. If "Yes (Judge abnormality from periodic waveform shape)" is selected, an AI that judges the similarity between waveform data and the normal waveform shape will be created.



4 Specify “Variable indicating interval start”: When using waveform data sets

If learning using waveform shape is not performed, the waveform data will be processed and converted to table data.
In the flow below, the data are automatically converted to table data by the AutoML function.



5	1	4	5.17	5.17	6.49	10.2	1861.812	8071.75	7638.574	615.614	0	0	0	0
16	1	15	5.213	5.147	6.383	10.253	1939.347	8274.703	8043.984	636.516	99.87755	491.472	556.893	69.61162
27	1	26	5.527	5.527	7.357	10.217	1994.084	8845.064	8414.571	581.93	253.0564	1250.396	992.1795	148.4597
38	1	37	8.234	8.234	9.744	12.374	2072.769	9052.303	8558.291	618.986	45.32037	33.35266	4.142003	3.335266

Convert quantified feature data into table data.

Learn.

In “Variable indicating interval start”, specify the marker that AutoML function will use to divide the intervals.
This is the variable in category type whose value will change at the interval change timing.
If “No” is set for the variable, the conditions for dividing manually should be set.
In manual settings, dividing the interval using detailed conditions can be performed by specifying the value of numerical type variables such as current, temperature, etc. as the conditions.
Refer to “5.2.3 Feature quantity engineering: (2) Taking a specific part of the data and extracting features” regarding dividing intervals using detailed conditions.

Example

Time	Product ID	Current	Temperature	...
2022/03/03 12:00:00	Prod_1	5	40	
2022/03/03 12:00:01	Prod_1	3	38	
2022/03/03 12:00:02	Prod_1	3	39	
:	:	:	:	
2022/03/03 12:01:10	Prod_1	5	39	
2022/03/03 12:01:11	Prod_2	4	39	
2022/03/03 12:01:12	Prod_2	5	40	
:	:	:	:	
2022/03/03 12:02:23	Prod_2	4	38	
2022/03/03 12:02:24	Prod_3	5	38	
2022/03/03 12:02:25	Prod_3	3	40	
:	:	:	:	

Interval 1

Interval 2

Interval 3

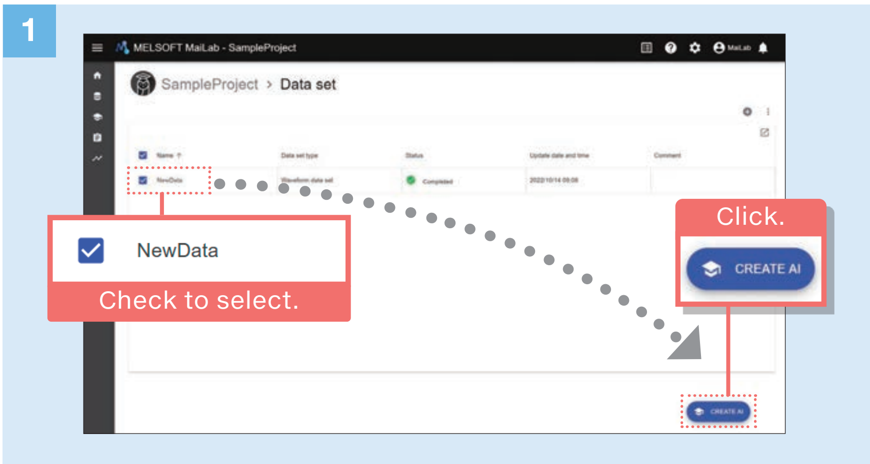
Variable indicating interval start

In the example, the interval changes as it changes from Prod_1 → Prod_2 → Prod_3.

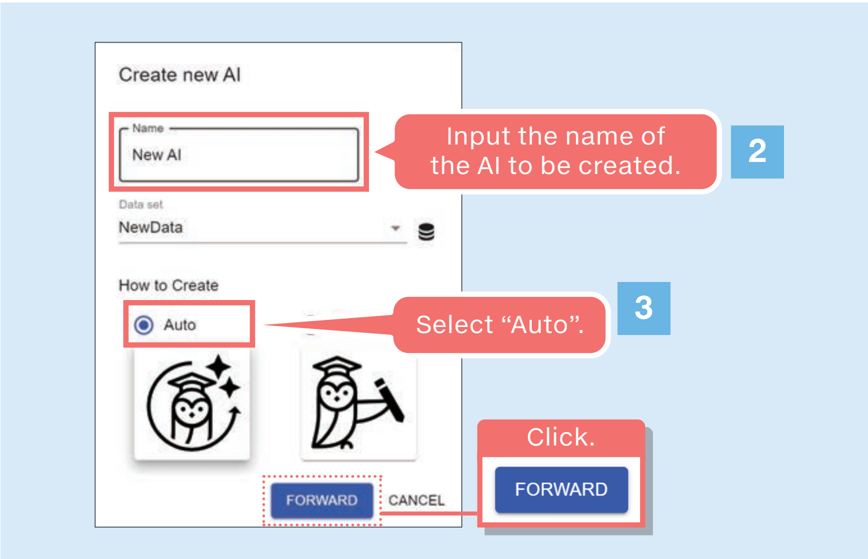
If “Yes” is selected, the category variables in the data set will be shown. Further, if “Set file switching as the start / end ondition of the section” is selected, the interval will be changed in data source file units.

Step 1. Interactively specify the learning method

The specific operating procedures for “To detect errors” will be explained.

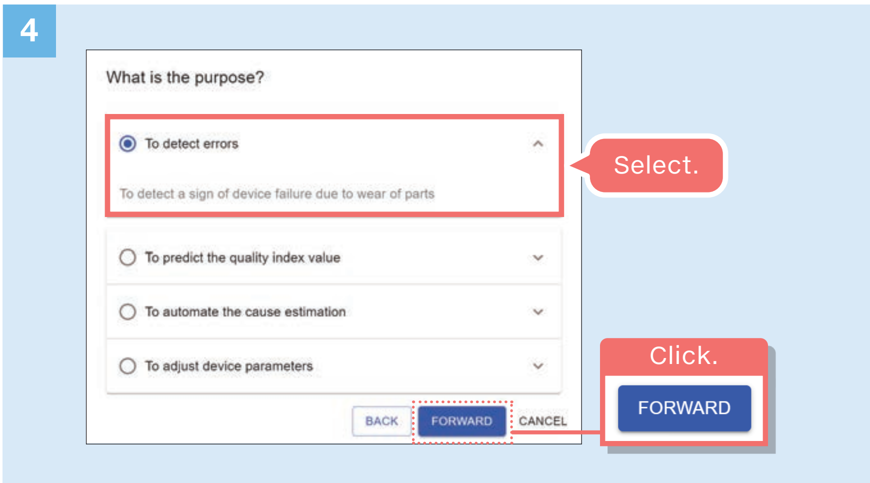


In the Data Set Management screen, select the data set to use for AI creation and click the “CREATE AI” button.



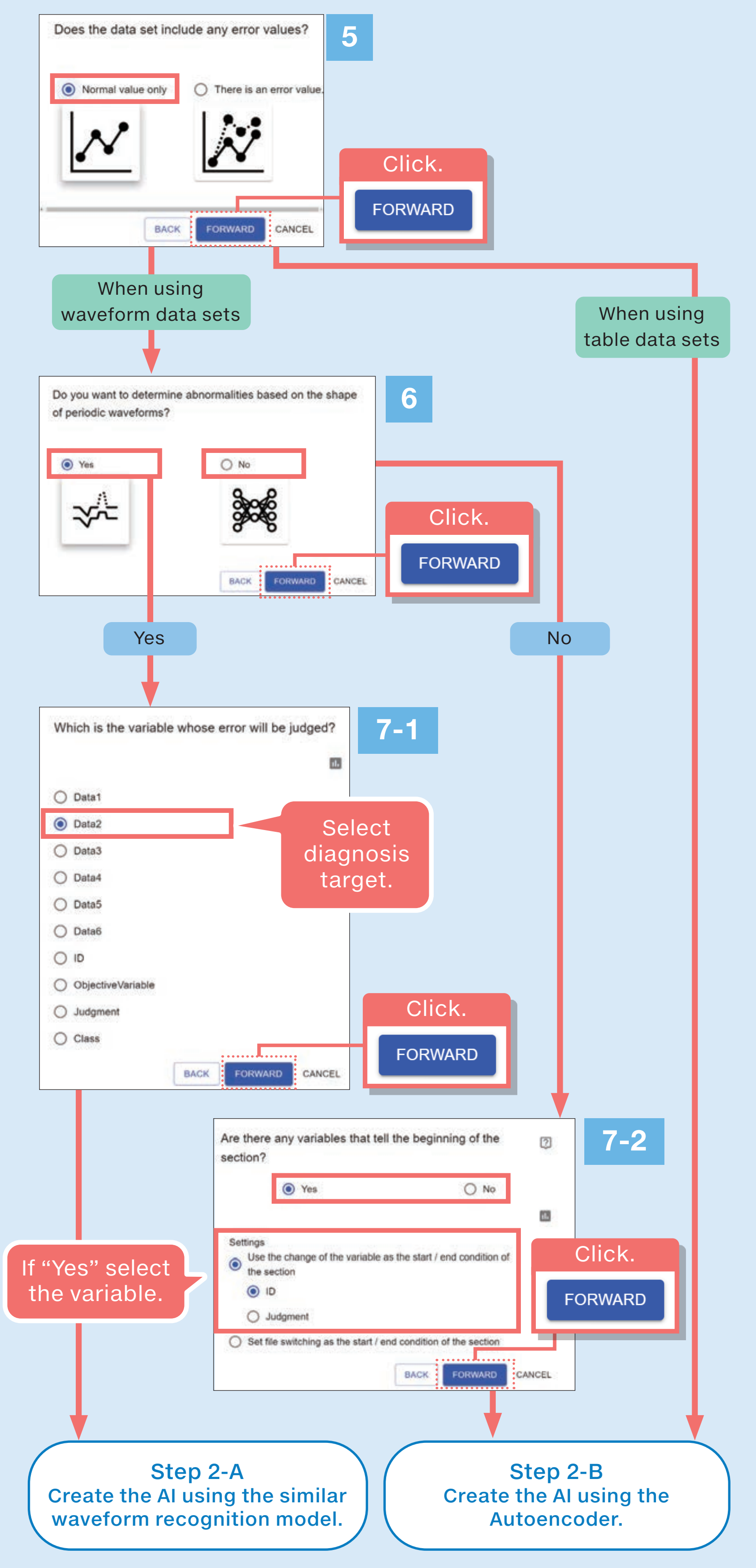
2 Input the name of the AI to be created in Name.

3 Select “Auto” for How to create and click the “FORWARD“ button.



Select “To detect errors” for Objective and click the “FORWARD“ button.

In the case of “Normal value only” (no abnormal values)



5 Select “Normal value only” and click the “FORWARD” button.
When using table data sets, interactive ends here.
Proceed to Step 2-B.

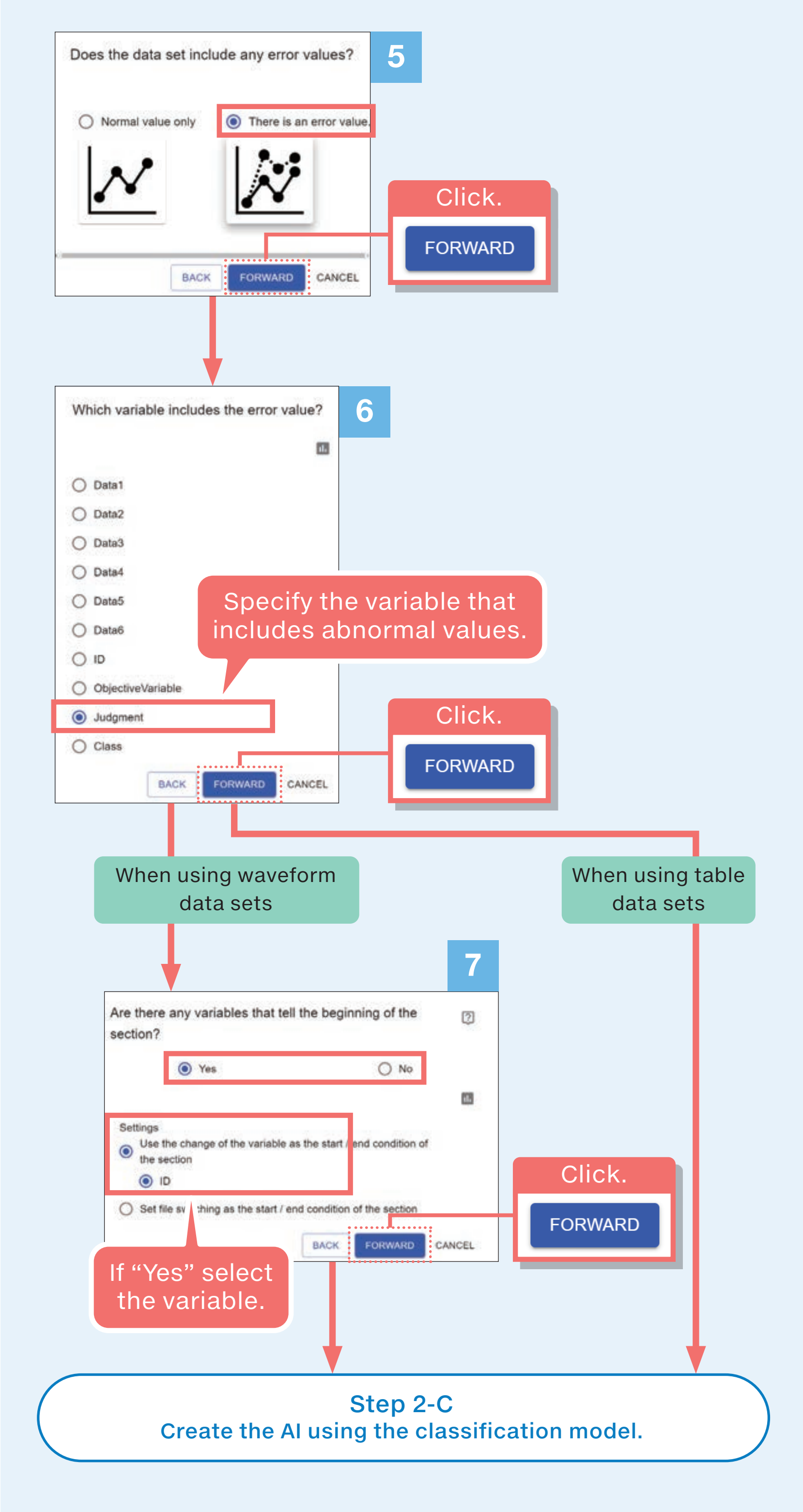
6 Select the waveform diagnosis method and click the “FORWARD” button.

7-1 [When “Yes” is selected]
Specify the diagnosis target variables and click the “FORWARD” button.

7-2 [When “No” is selected]
Select whether or not there is a variable indicating interval start and if “Yes”, select the variable indicating interval start, and then click the “FORWARD” button.

Interactive ends.
Proceed to Step 2-B.

In the case of “There is an error value.”



5 Select “There is an error value.” and click the “FORWARD” button.

6 Select the variable that includes abnormal values and click the “FORWARD” button.

When using table data sets, interactive ends here.
Proceed to Step 2-C.

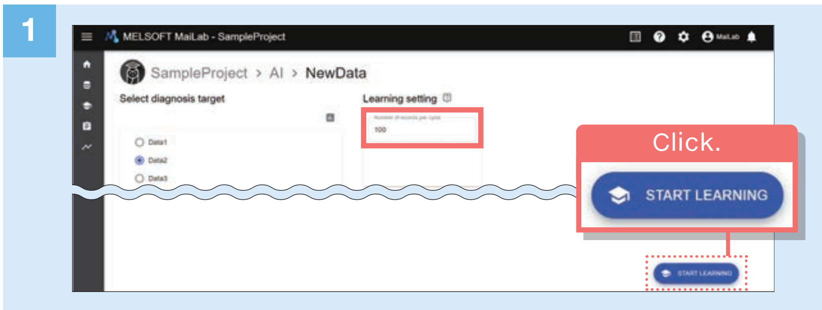
7 Select whether or not there is a variable indicating interval start and if “Yes”, select the variable indicating interval start, and then click the “FORWARD” button.

Interactive ends.
Proceed to Step 2-C.

Step 2-A. Create the AI using the similar waveform recognition model

Process from “Normal value only” → “Judge abnormality from waveform pattern”.

The AI will be created by learning the waveform shape of normal data.
During diagnosis using the task, an index (similarity score) indicating the level of similarity between the diagnosis target input waveform and the learned waveforms. By performing threshold value judgment on the similarity score, “Different than usual” conditions can be detected.

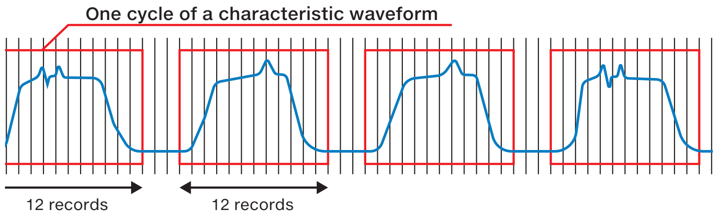


Set the number of records per cycle and click the “START LEARNING” button.

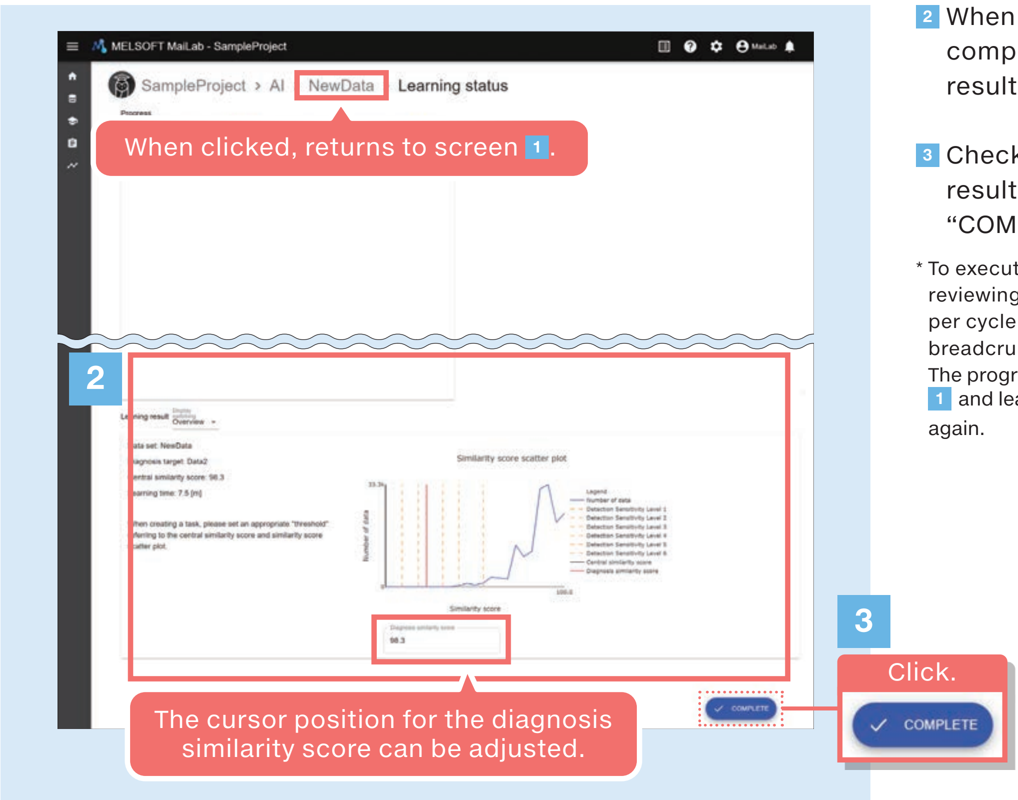
* For details on the learning behavior and diagnosis behavior of the similar waveform recognition model, refer to the MELSOFT MaiLab User’s Manual.

Number of records per cycle

The number of records per cycle of periodically occurring characteristic waveforms. Set in multiples of 4 in the range of 8 to 1000. In the example diagram at right, characteristic waveforms are enclosed in red frames, and for a single cycle there are 12 records.



Caution: The similar waveform recognition model (Judge abnormality from waveform pattern) cannot be applied to waveforms that are not periodic. For waveforms that are not periodic, investigate other learning/diagnosis methods such as an Autoencoder (select “No” for “Judge abnormality from waveform patterns”).



2 When learning has been completed, the learning results will be displayed.

3 Check the learning results and click the “COMPLETE” button.

* To execute learning again after reviewing the number of records per cycle, click “AI name” in the breadcrumb list.
The program will return to screen 1 and learning can be executed again.

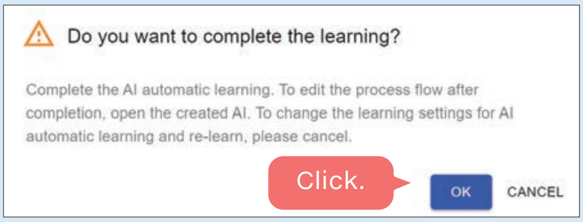
Similarity score scatter plot

Distribution of similarity scores resulting from training and validating with the data set used. The distribution and median similarity score should be considered when determining the diagnosis threshold value and specifying it during task creation.

Central similarity score

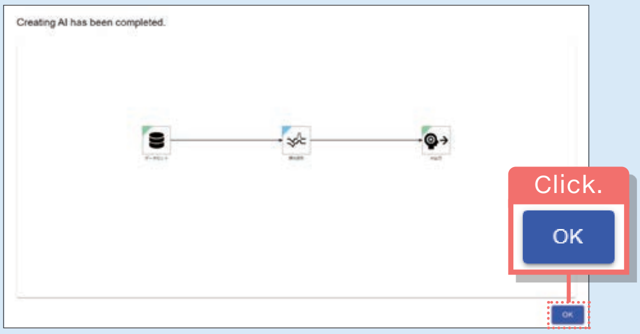
The threshold value recommended by MaiLab from the verification results.
There is a tendency that when the variation of the waveform shapes included in the data set is low, the similarity score will increase, and when the variations are high, the score will decrease.

4



The Learning Completed Confirmation dialog will appear. Click the “OK” button.

5



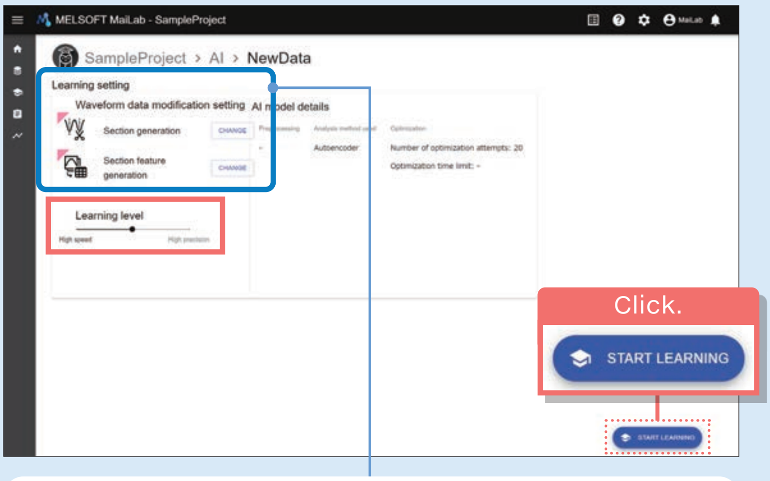
The created AI will be displayed. Click the “OK” button.

Step 2-B. Create the AI using the Autoencoder

Process from “No” for “Normal value only” → “Judge abnormality from waveform pattern”.

The AI will be created with the Autoencoder and by learning using normal data only.
The Autoencoder is a neural network that encodes (encodes, compresses) input data and converts it into separate data, and recovers and outputs the original data. When normal data are input, recovery succeeds and a form close to the input data is output. When abnormal data are input, they cannot be recovered correctly, and the error from the input data becomes great.
By performing threshold value judgment on the recovery error (abnormality) when performing diagnosis in the task, “Different than usual” conditions can be detected.

1



The waveform data set will be displayed, and the detailed settings for interval dividing can be set. When interval division is performed using the contents of the simple settings in Step 1 7-2, setting (changing) is not necessary.
When performing detailed settings, refer to “5.2.3 Feature quantity engineering: (2) Taking a specific part of the data and extracting features”.

Set the learning level using the slider and click the “START LEARNING” button.

Learning level

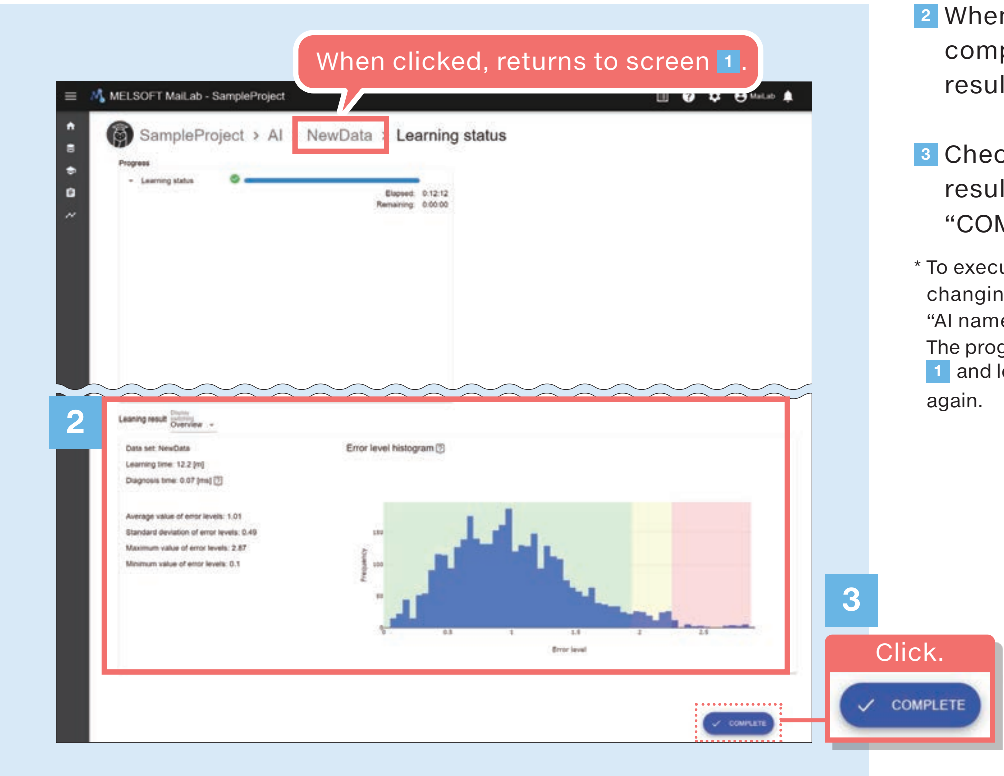
Specify the number of hyperparameter optimization trials.
If the high level (Level 3) is specified, although higher levels of learning accuracy can be expected, the time required for learning will become long.

High speed ←→ High accuracy

	Learning level 1	Learning level 2	Learning level 3
Number of optimization trials	0 times	20 times	100 times

Hyperparameter

Settings to control the behavior of analysis methods. The set values will affect prediction results and processing functions.
In MaiLab, the number of trials shown above according to the learning level will be performed, and the parameter with the highest performance will be selected as the parameter.



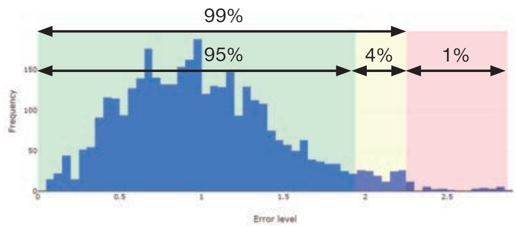
2 When learning has been completed, the learning results will be displayed.

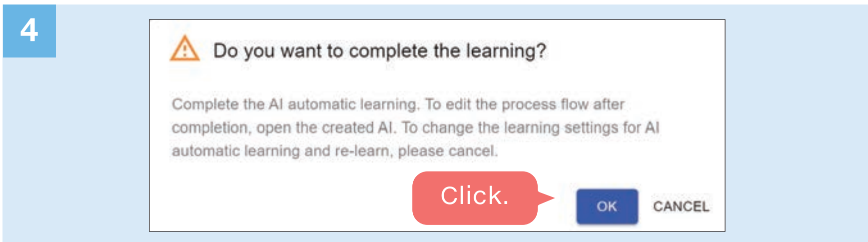
3 Check the learning results and click the "COMPLETE" button.

* To execute learning again after changing the learning level, click "AI name" in the breadcrumb list. The program will return to screen 1 and learning can be executed again.

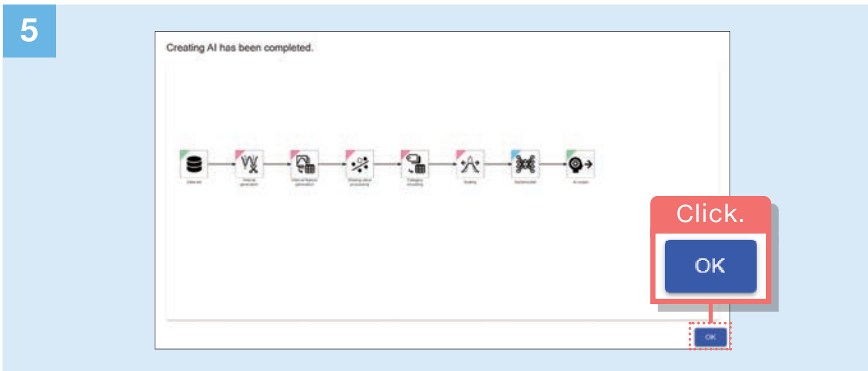
Error level histogram

Distribution of abnormality levels resulting from training and validating with the data set used.
The green area indicates 95% of the distribution, and the green and yellow areas indicate 99% of the distribution.
The distribution and abnormality statistics should be considered when determining the diagnosis threshold value and specifying it during task creation.





The Learning Completed Confirmation dialog will appear. Click the "OK" button.



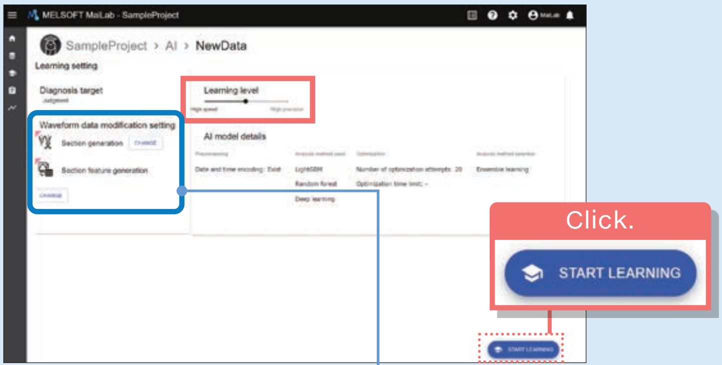
The created AI will be displayed. Click the "OK" button.

Step 2-C. Create the AI using the classification model

Process from “There is an error value.”

An AI in which the input data infers the affiliation of the binary category specified by the objective variable (variable including abnormal data) will be created.
During task diagnosis, the inference results will be output as binary format data (OK/NG, true/false, etc.)

1



The waveform data set will be displayed, and the detailed settings for interval dividing can be set. When interval division is performed using the contents of the simple settings in Step 1 7, setting (changing) is not necessary. When performing detailed settings, refer to “5.2.3 Feature quantity engineering: (2) Taking a specific part of the data and extracting features”.

Set the learning level using the slider and click the “START LEARNING” button.

Learning level

Depending on the learning level, the number of analysis methods used during learning and the number of hyperparameter optimization trials for each analysis method will be different. If the high level (Level 3) is specified, although higher levels of learning accuracy can be expected, the time required for learning will become long.

High speed ⟷ High accuracy

	Learning level 1	Learning level 2	Learning level 3
Analysis method used	LightGBM Random forest	LightGBM Random forest Deep learning	LightGBM XGBoost Random forest Deep learning k-nearest neighbors algorithm
Number of hyperparameter optimization trials	0 times	20 times	20 times

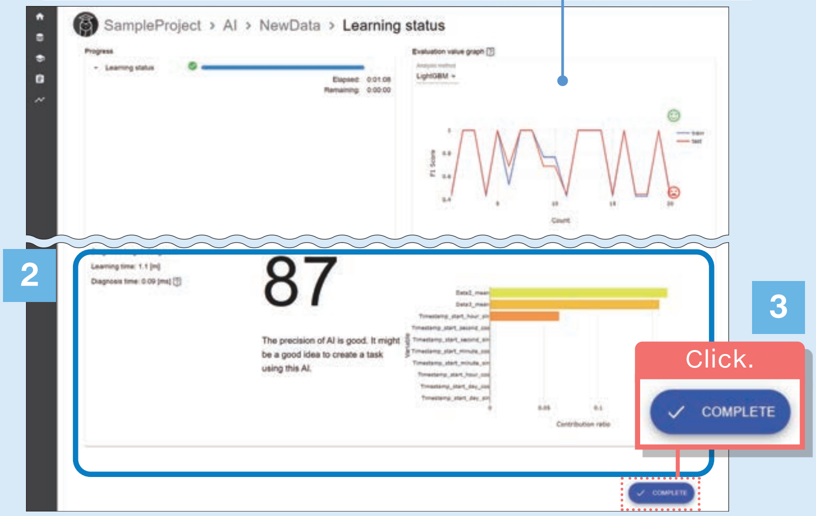
Analysis method used

To improve learning results, combine the inference results of several analysis methods and perform ensemble learning. Depending on the learning level, the number of analysis methods that can be combined will be different.

Hyperparameter

Settings to control the behavior of analysis methods. The set values will affect prediction results and processing functions. In MaiLab, the number of trials shown above according to the learning level will be performed for each analysis method, and the parameter with the highest performance will be selected.

The progress of the hyperparameter trial is displayed.
Not shown for learning level 1.



2 When learning has been completed, the learning results will be displayed.

3 Check the learning results and click the “COMPLETE” button.

* To execute learning again after changing the learning level, click “AI name” in the breadcrumb list. The program will return to screen 1 and learning can be executed again.

Learning results (scores)

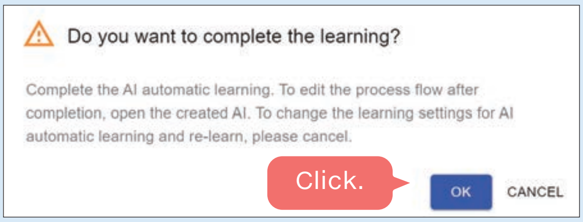
The learning results for the data set used are displayed from 0 to 100 points. For the classification model, the F1 Score will be an integer value 100 times the score. For the regression model, the R2 will be an integer value 100 times the value. A task to perform real-time diagnosis while referring to the displayed scores and comments will be created. If the score is insufficient, it can be improved by changing the learning level and performing learning again, etc. Methods for improving scores are explained in chapter 5 “Improving the accuracy of the diagnosis model”.

Contribution ratio of variables

A numerical value indicating the degree of influence of each explanatory variable.
The explanatory variables with the top 10 contribution rates are shown.

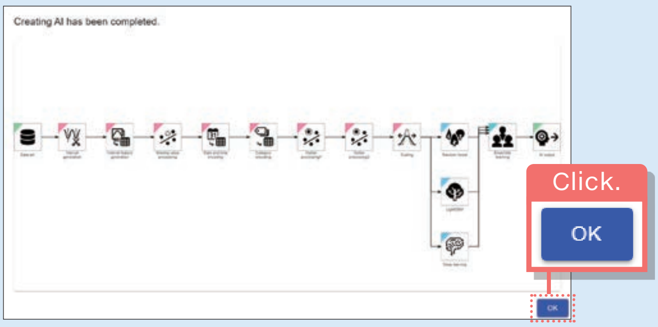
* The displayed explanatory variables also include some created automatically by MaiLab.

4



The Learning Completed Confirmation dialog will appear. Click the “OK” button.

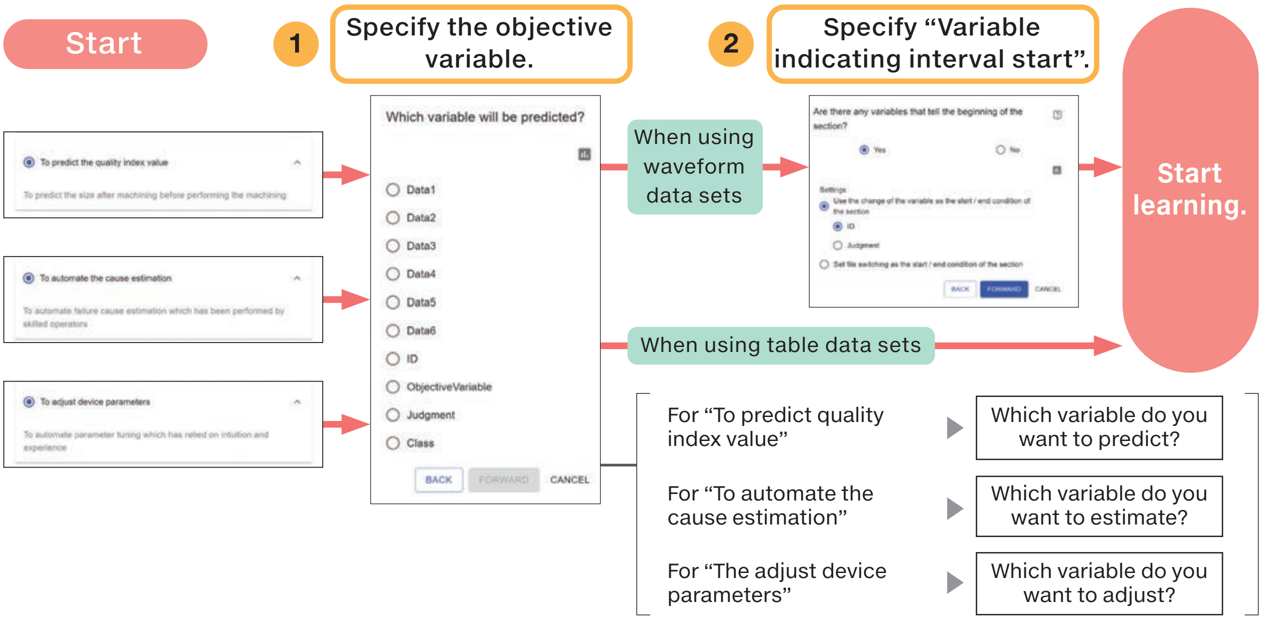
5



The created AI will be displayed. Click the “OK” button.

3.2.2 For other than “To detect errors”

The interactive flow for “To predict quality index value”, “To automate the cause estimation”, and “The adjust device parameters” is shown below.
In any of the cases, specify the objective variables for prediction, estimation, or adjustment, and create the AI.

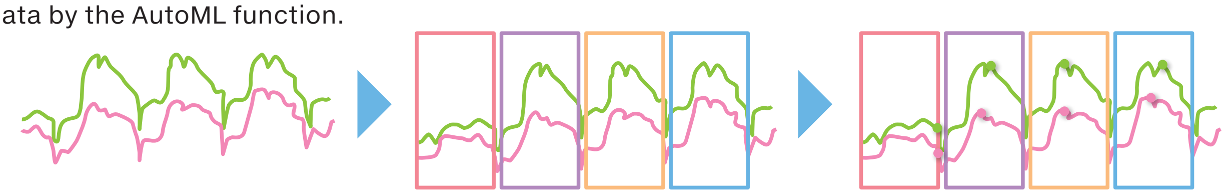


1 Specify the objective variable.

The target variable to predict, estimate, or adjust.
For “To automate the cause estimation”, category type variables can be specified. For “To predict quality index value” or “The adjust device parameters”, category type or numerical value type variables can be specified.

2 Specify “Variable indicating interval start”: When using waveform data sets

When using waveform data sets, the waveform data will be processed and converted to table data, and then learning will be performed.. In the flow below, the data are automatically converted to table data by the AutoML function.



Sequential waveform data

Divide waveform data into intervals.

Quantify features for each interval.

5	1	4	5.17	5.17	6.49	10.2	1861.812	8071.75	7638.574	615.614	0	0	0	0
16	1	15	5.213	5.147	6.383	10.253	1939.347	8274.703	8043.984	636.516	99.87755	491.472	556.893	69.61162
27	1	26	5.527	5.527	7.357	10.217	1994.084	8845.064	8414.571	581.93	253.0564	1250.396	992.1795	148.4597
38	1	37	8.234	8.234	9.744	12.374	2072.769	9052.303	8558.291	618.986	45.32037	33.35266	4.142003	3.335266

Convert quantified feature data into table data.

Learn.

In “Variable indicating interval start”, select the marker that AutoML function will use to divide the intervals. This is the variable in category type whose value will change at the interval change timing. If “No” is set for the variable, the conditions for dividing manually should be set.
In manual settings, dividing the interval using detailed conditions can be performed by specifying the value of numerical type variables such as current, temperature, etc. as the conditions.
Refer to “5.2.3 Feature quantity engineering: (2) Taking a specific part of the data and extracting features” regarding dividing intervals using detailed conditions.

Example

Time	Product ID	Current	Temperature	...
2022/03/03 12:00:00	Prod_1	5	40	
2022/03/03 12:00:01	Prod_1	3	38	
2022/03/03 12:00:02	Prod_1	3	39	
:	:	:	:	
2022/03/03 12:01:10	Prod_1	5	39	
2022/03/03 12:01:11	Prod_2	4	39	
2022/03/03 12:01:12	Prod_2	5	40	
:	:	:	:	
2022/03/03 12:02:23	Prod_2	4	38	
2022/03/03 12:02:24	Prod_3	5	38	
2022/03/03 12:02:25	Prod_3	3	40	
:	:	:	:	

Interval 1

Interval 2

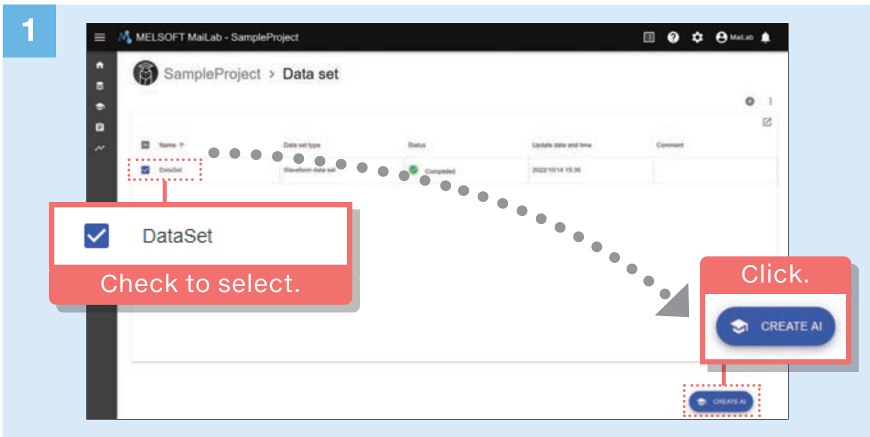
Interval 3

Variable indicating interval start

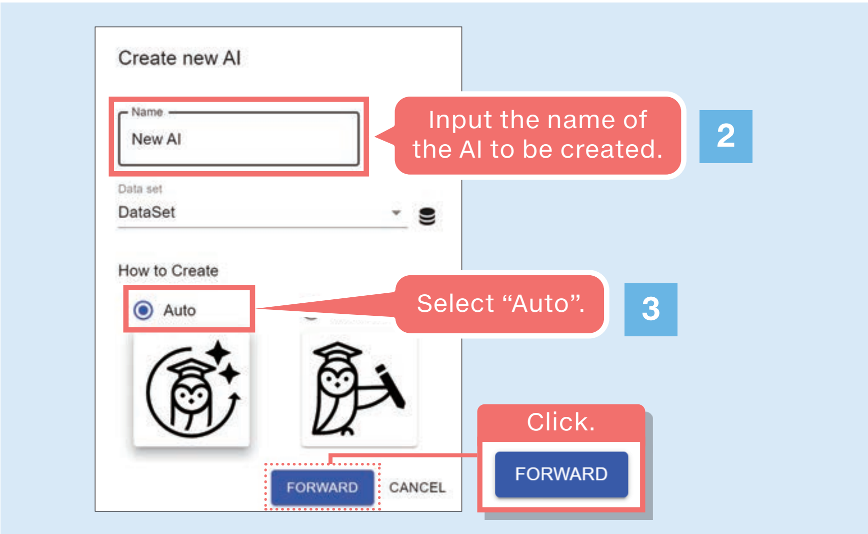
In the example, the interval changes as it changes from Prod_1 to Prod_2 to Prod_3.

If “Yes” is selected, the category variables in the data set will be shown. Further, if “Use file change as interval start/end condition” is selected, the interval will be changed in data source file units.

The specific operating procedures for “To predict quality index value”, “To automate the cause estimation”, and “The adjust device parameters” will be explained. The AI will be generated using the classification model if the objective variable is category type, and using the regression model if the objective variable is numerical type. In the classification model, the affiliation of the specified category is inferred by the objective variable of the input data. In the regression model, the value (numerical value) of the objective variable is inferred from the input data.

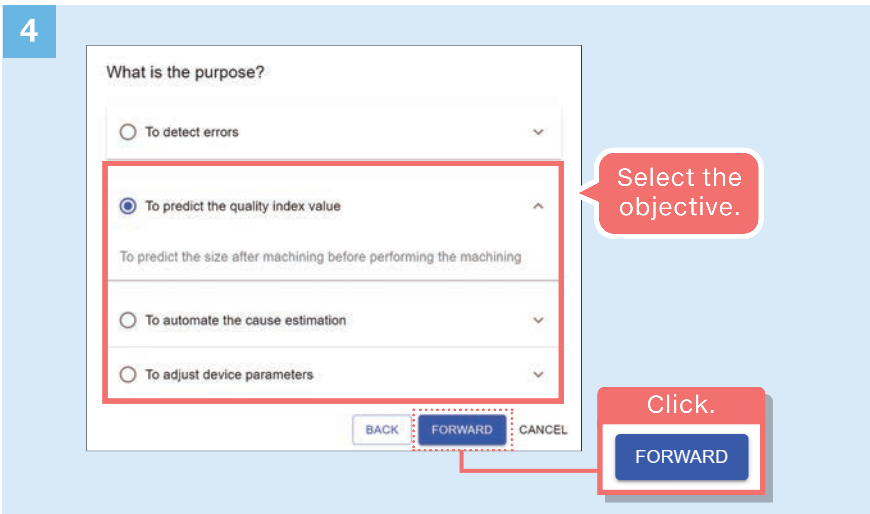


In the Data Set Management screen, select the data set to use for AI creation and click the “CREATE AI” button.

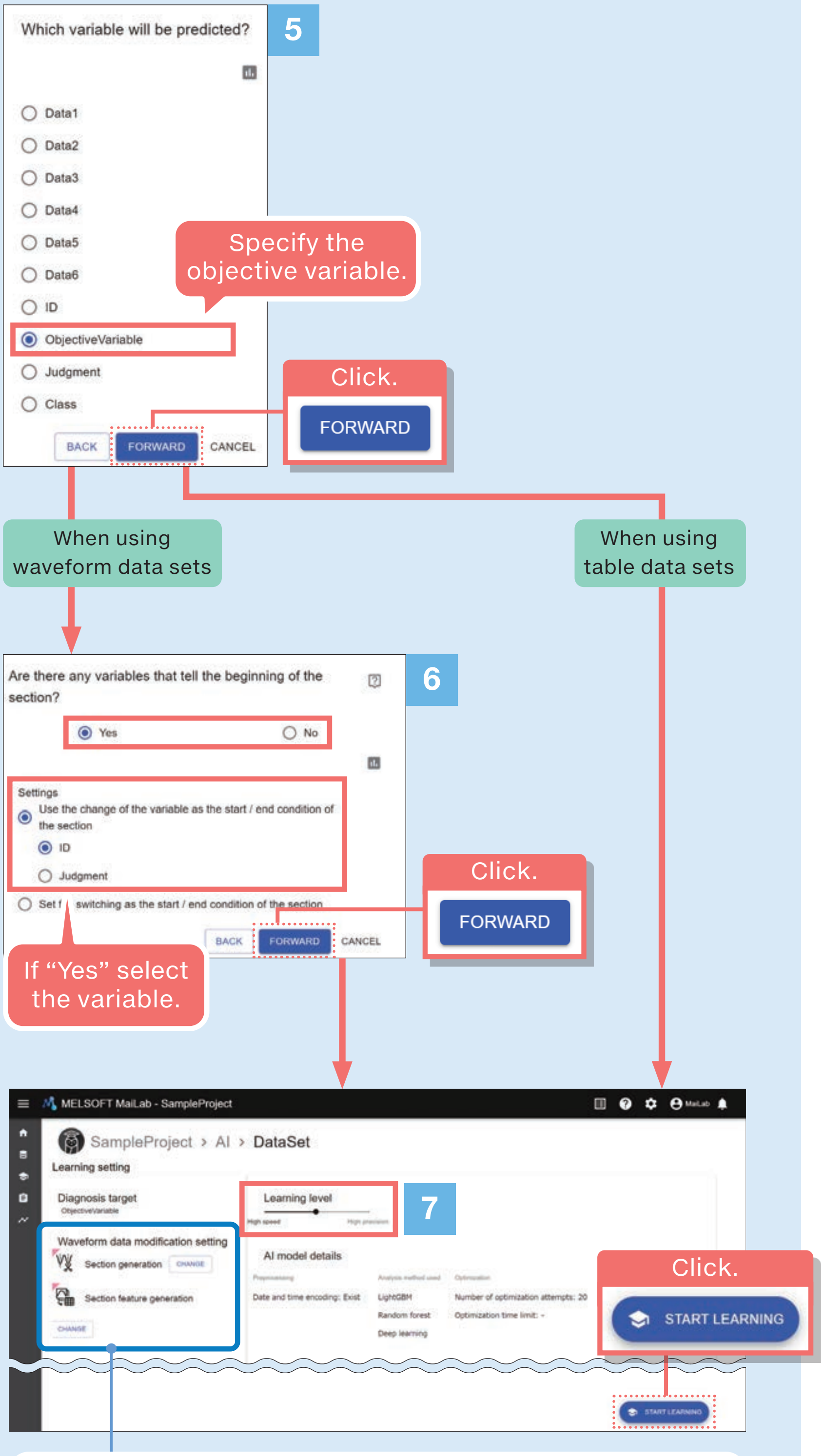


2 Input the name of the AI to be created in Name.

3 Select “Auto” for the creation method and click the “FORWARD” button.



Select the objective and click the “FORWARD“ button.



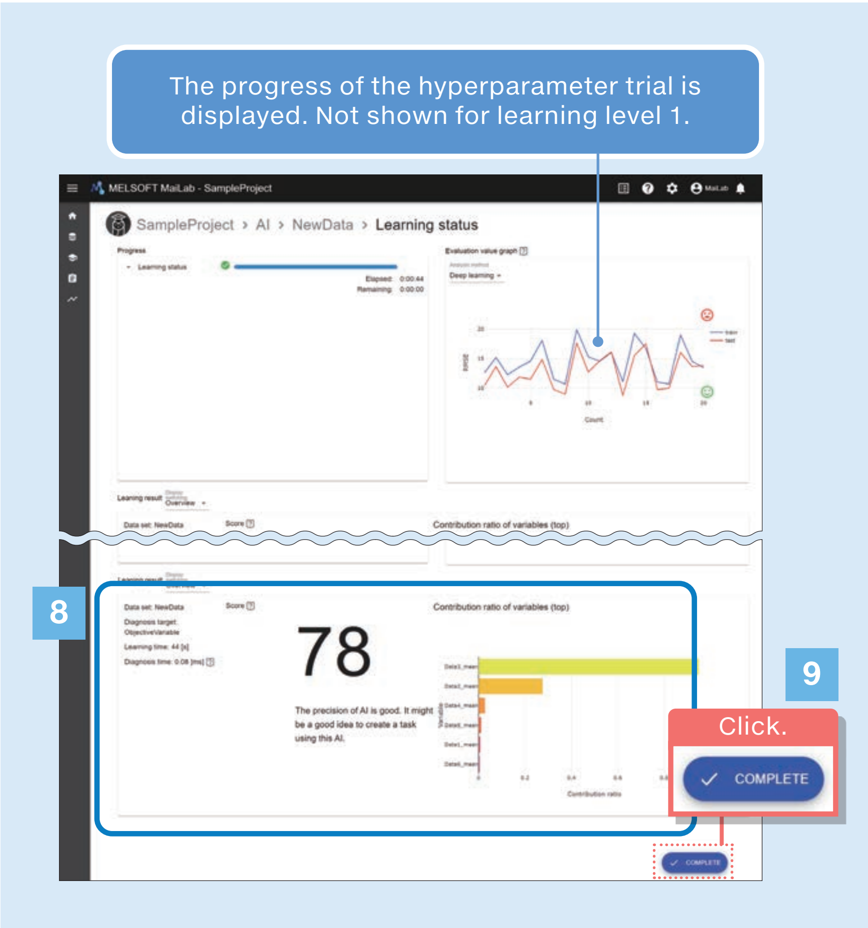
The waveform data set will be displayed, and the detailed settings for interval dividing can be set. When interval division is performed using the contents of the simple settings in 6, setting (changing) is not necessary. When performing detailed settings, refer to "5.2.3 Feature quantity engineering: (2) Taking a specific part of the data and extracting features".

5 Specify the objective variable and click the "FORWARD" button.

6 When using waveform data sets, select whether or not there is a variable indicating interval start and if "Yes" the variable indicating interval start, and then click "FORWARD" button.

7 Set the learning level using the slider and click the "START LEARNING" button.

For more information regarding learning level, analysis methods used, and hyperparameters, refer to Step 2-C of 3.2.1 For the case of “To detect errors”.



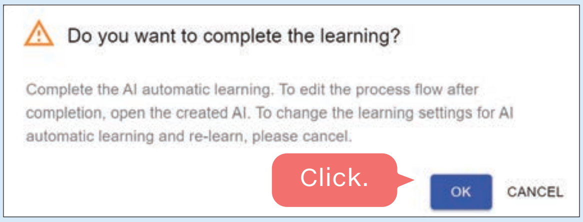
8 When learning has been completed, the learning results will be displayed.

9 Check the learning results and click the “COMPLETE” button.

* To execute learning again after changing the learning level, click “AI name” in the breadcrumb list. The program will return to screen 7 and learning can be executed again.

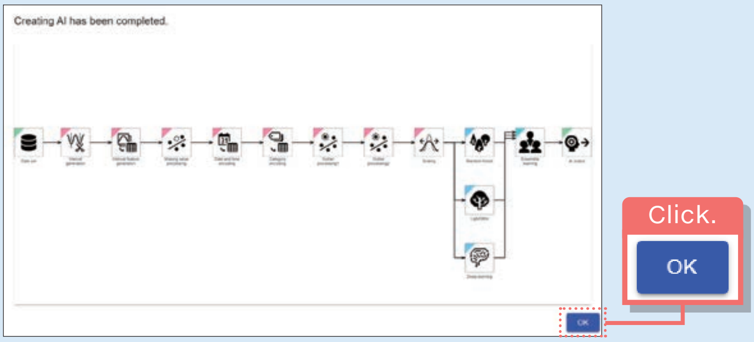
For more information regarding learning results (scores) and contribution ratio of variables, refer to Step 2-C of 3.2.1 For the case of “To detect errors”.

10



The Learning Completed Confirmation dialog will appear. Click the “OK” button.

11



The created AI will be displayed. Click the “OK” button.

3.3 Executing tasks using the created AI

A group of processes (process flow) using the created AI to perform diagnosis on unknown input data and output the diagnosis results is called a "task" in MaiLab.
There are 2 types of tasks, and the creation method is different depending on the type.

Task creation

The 2 methods for creating tasks are as follows:

Simple task

Creating a simple task

Select to easily create a task. The process flow is automatically created by just setting processing parameters.

Advanced task

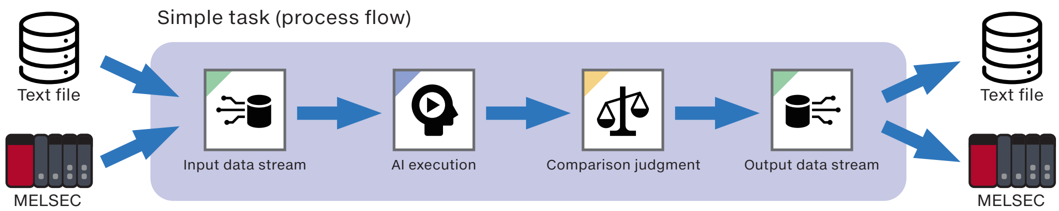
Creating a task by making detailed settings

Users can freely combine processes to create the process flow. Select to make detailed settings of processing contents or procedures in the task.

In this section, the procedure for creating a simple task type task and executing it will be explained.

3.3.1 Creating a simple task

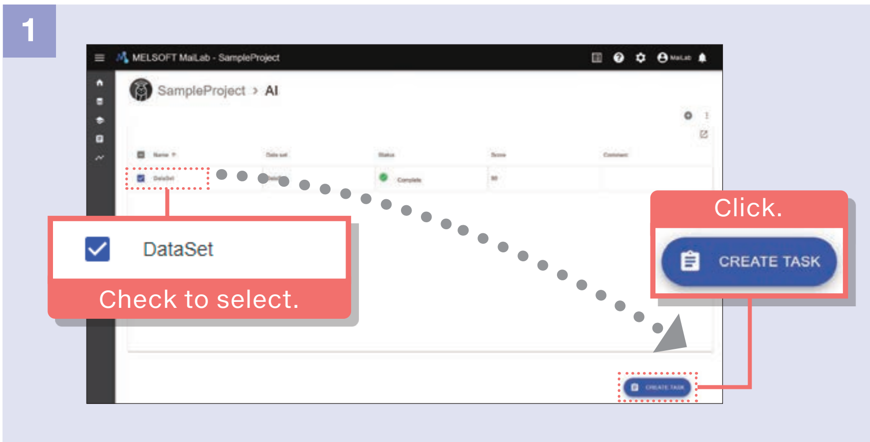
The processes executed by a simple task and their flow are shown below.
The simple task is automatically created by setting the necessary parameters for the operation of each process.



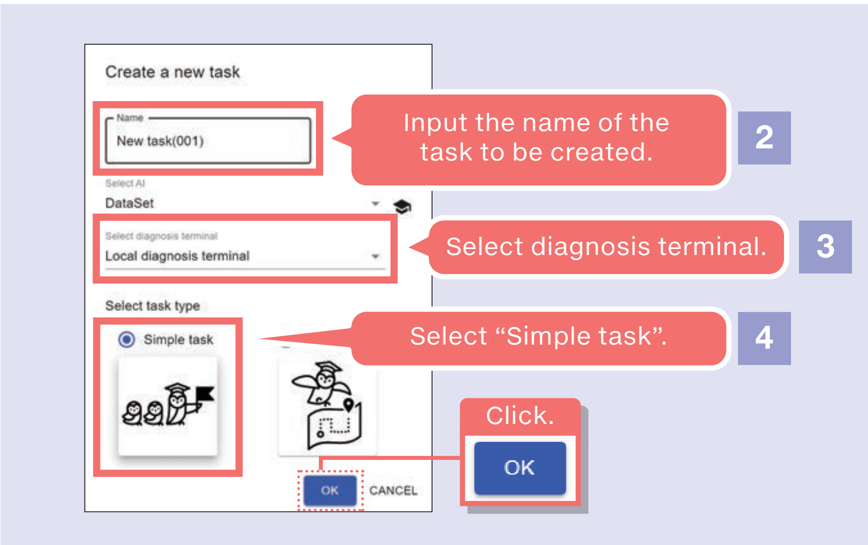
Process	Explanation
Input data stream	Process that collects data for input to the AI execution. There are 2 methods for collecting input data: importing from a text file and collecting directly from MELSEC devices. * Linking with Edgecross is performed using text files.
AI execution	Process that performs diagnosis using the AI created in the previous section.
Comparison judgment	Process that performs threshold value judgment on the results output by the AI execution.
Output data stream	Process that outputs diagnosis results. There are 2 methods for outputting results: outputting to a text file or writing directly to MELSEC devices. * Linking with Edgecross is performed using text files.

Step 1. Creating a new task of simple task type

The specific operating procedures for simple task creation will be explained.



In the AI Management screen, select the AI that will be used by the task and click the “CREATE TASK” button.

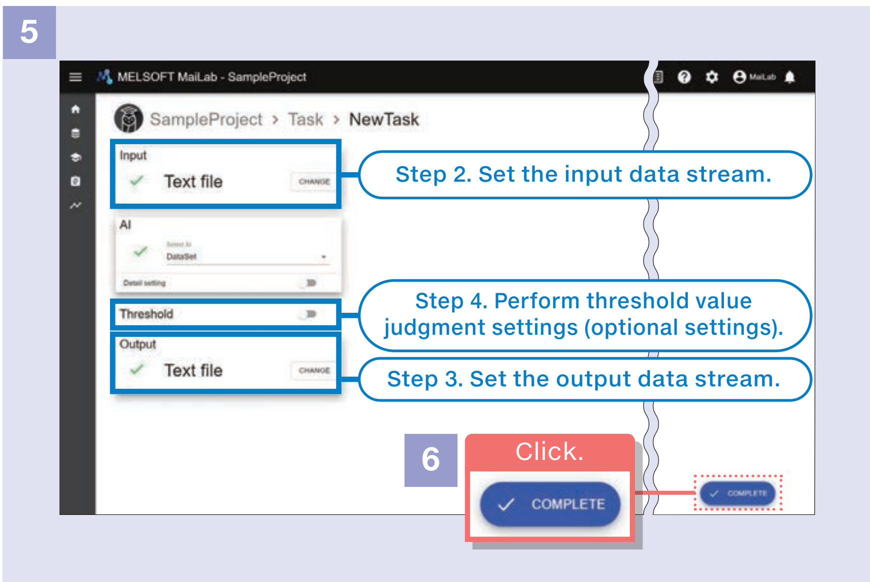


2 Input the name of the task to be created in Name.

3 Select the diagnosis terminal that will execute the task.

* When the diagnosis terminal and learning server are on the same PC, select “Local diagnosis terminal”.

4 Select “Simple task” for Task type and click the “OK” button.

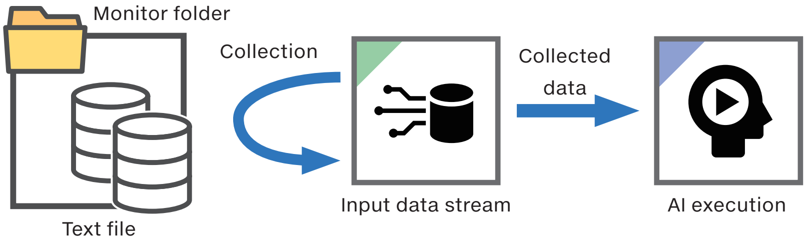


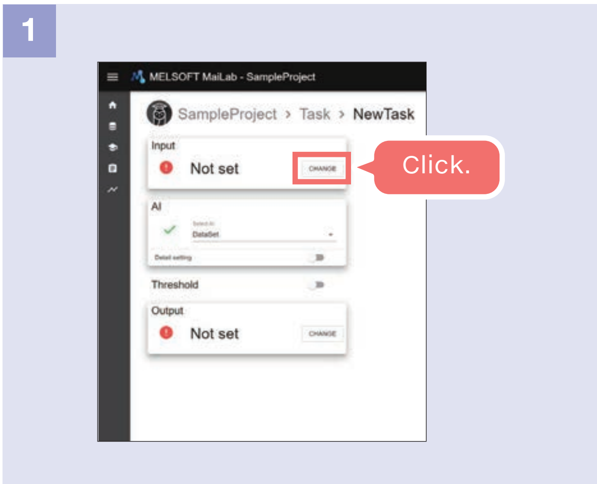
5 Proceed to the parameter setting dialog for each process of the simple task and set the parameters for each process. (Details are explained in Step 2 and later.)

6 Click the “COMPLETE” button.

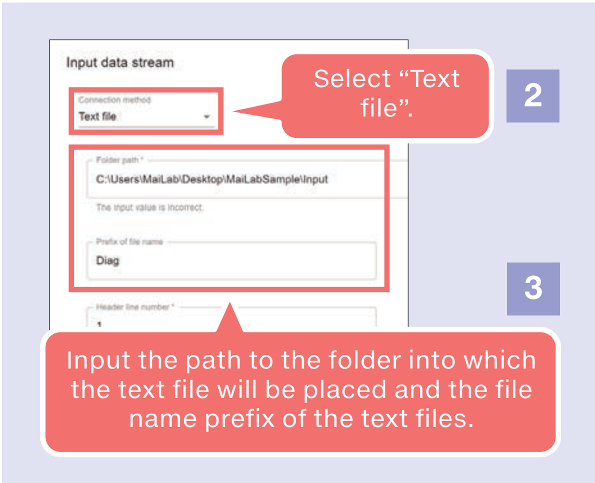
Step 2-A. Perform input data stream settings (when inputting a text file)

The procedure for collecting data from a text file will be explained.
In collecting from text files, a text file in the monitor folder is read at regular intervals and the read data are output to the AI execution.
For collecting directly from MELSEC, proceed to Step 2-B.





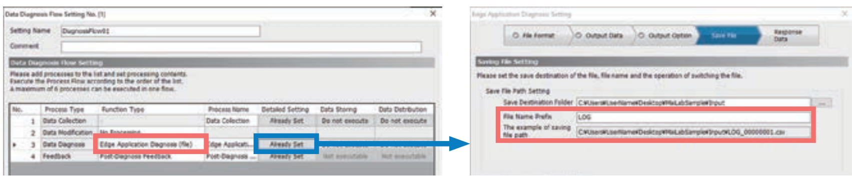
Click the Input “CHANGE” button.



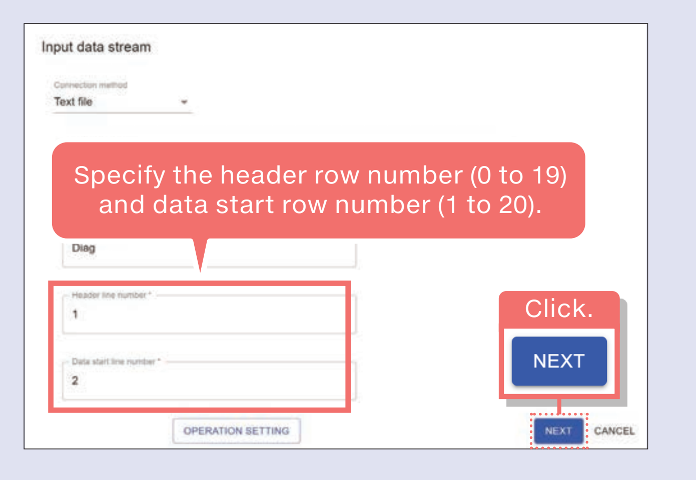
2 The input data stream dialog will appear. Select “Text file” for Connection method.
3 Set the path to the monitor folder and the file name prefix of the text files.

Edgecross linkage

The Edgecross data flow can be linked with and operated by setting the Edgecross data diagnostic process function type to “Edge Application Diagnostics (File)” and the save folder and prefix in the save file settings to the same values as the input data stream (Figure 3 above).



4

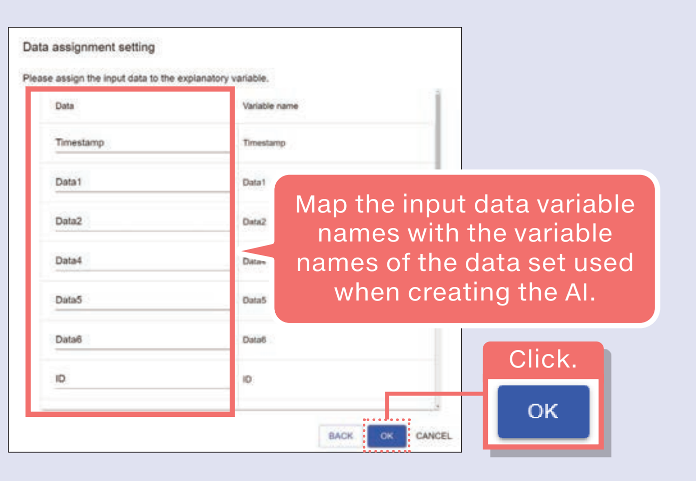


Input the header line number and data start line number of the input text file and click the “NEXT” button.

Header line number, Data start line number

For the meanings of header line number and data start line number, refer to “3.1 Creating the data set”.

5



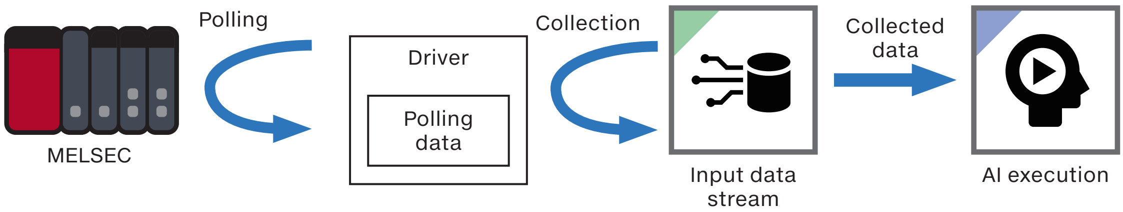
Assign input data to the explanatory variables and click the “OK” button.

Assigning input data to explanatory variables.

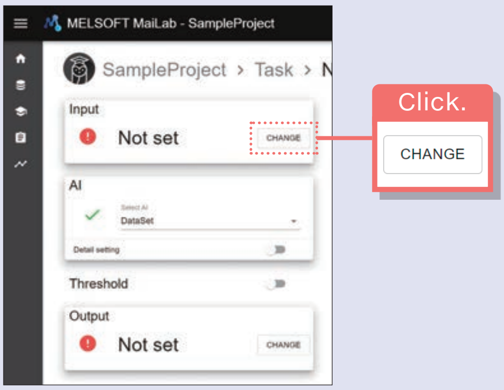
Assign input data to the explanatory variables (explanatory variables column on the right side of the Assignment screen) which will be input to the AI execution. The variable names of the data set used when creating the AI will be shown in the data column. If the variable names of the input text file are the same as the data set variable names, click the “OK” button without editing.
If the variable names are different, correct the variable names in the data column and map the explanatory variables. If the input text file has no header row (when header row number has been set to 0), the row number will be shown in the data column. Map by row number.

Step 2-B. Perform input data stream settings (when connecting to MELSEC)

The procedure for collecting directly data from MELSEC will be explained.
In collecting from MELSEC, the driver for MELSEC access reads data collected by polling at regular intervals, and outputs the data read out by the AI execution.

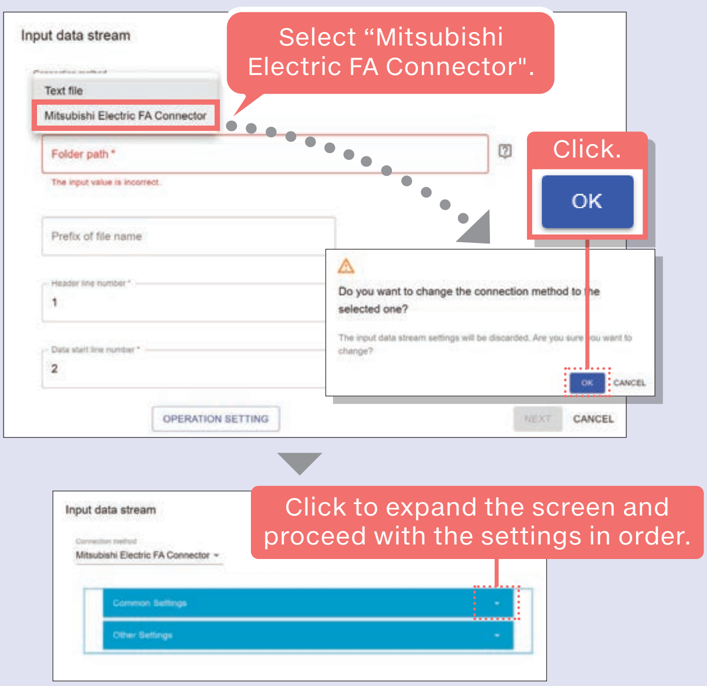


1

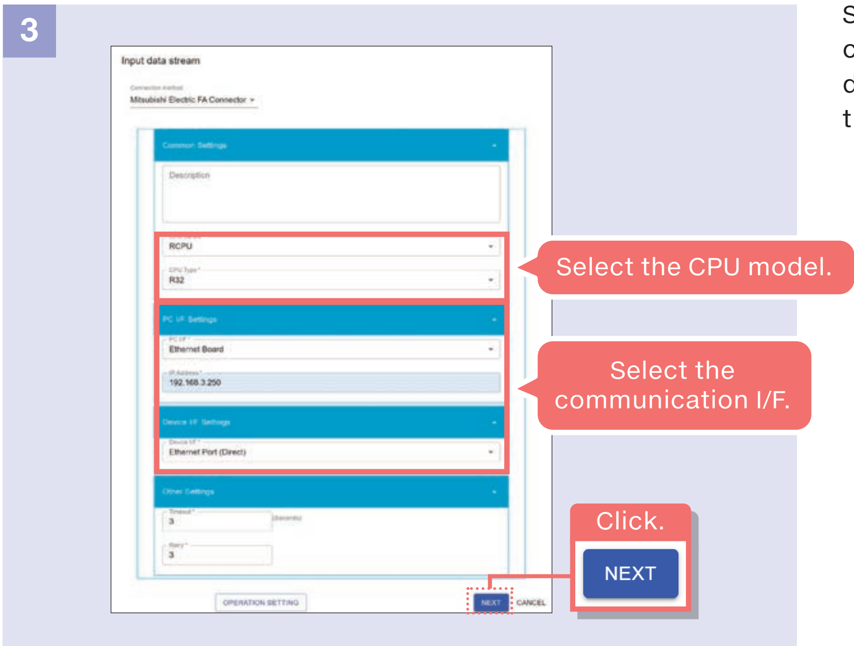


Click the Input “CHANGE” button.

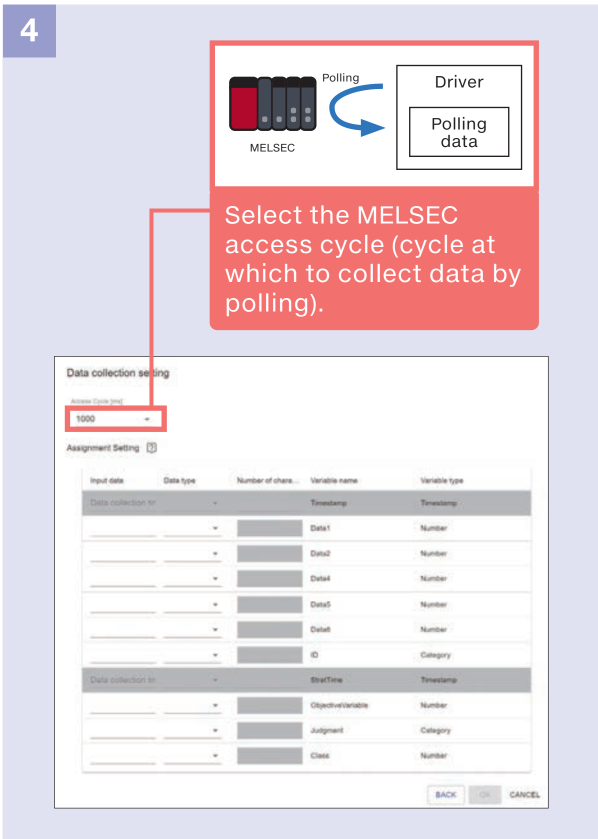
2



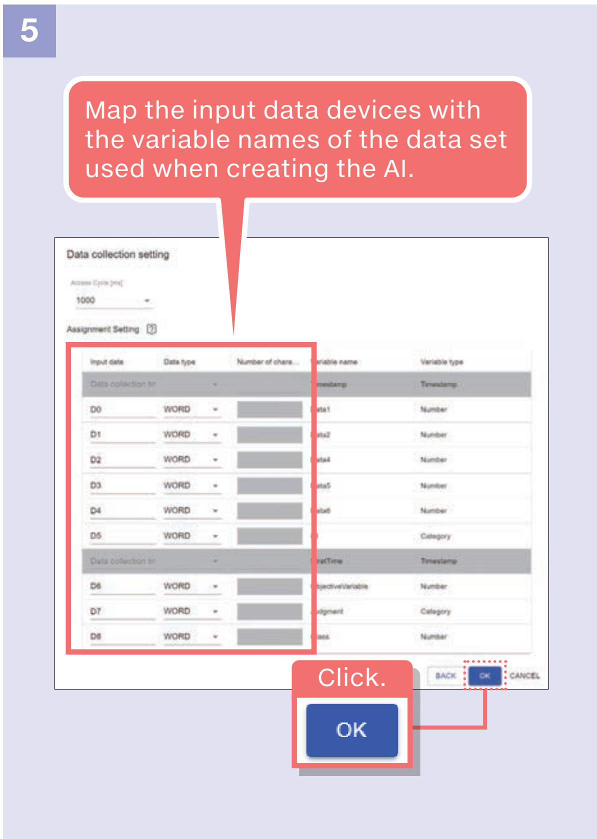
The input data stream dialog will appear. Select “Mitsubishi Electric FA Connector” for Connection method and click the “OK” button in the change confirmation dialog.



Set the model information and communication method for the data collection target, and click the “NEXT” button.



Select the MELSEC access cycle (cycle at which to collect data by polling).



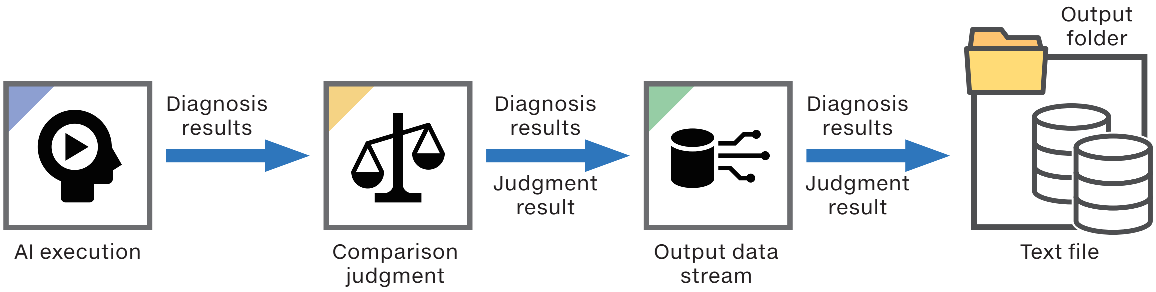
Assign input data (devices) to the explanatory variables and click the “OK” button.

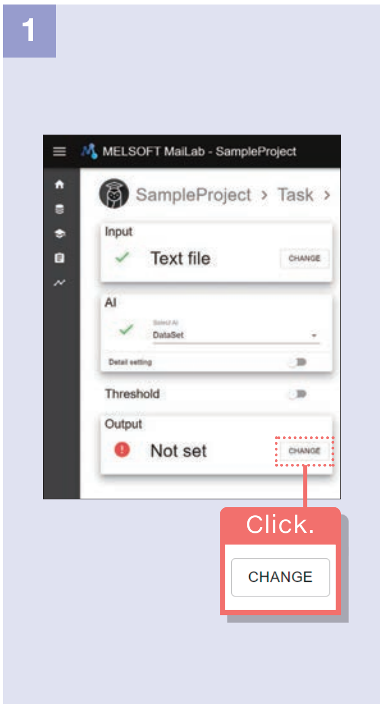
Assigning input data to explanatory variables.

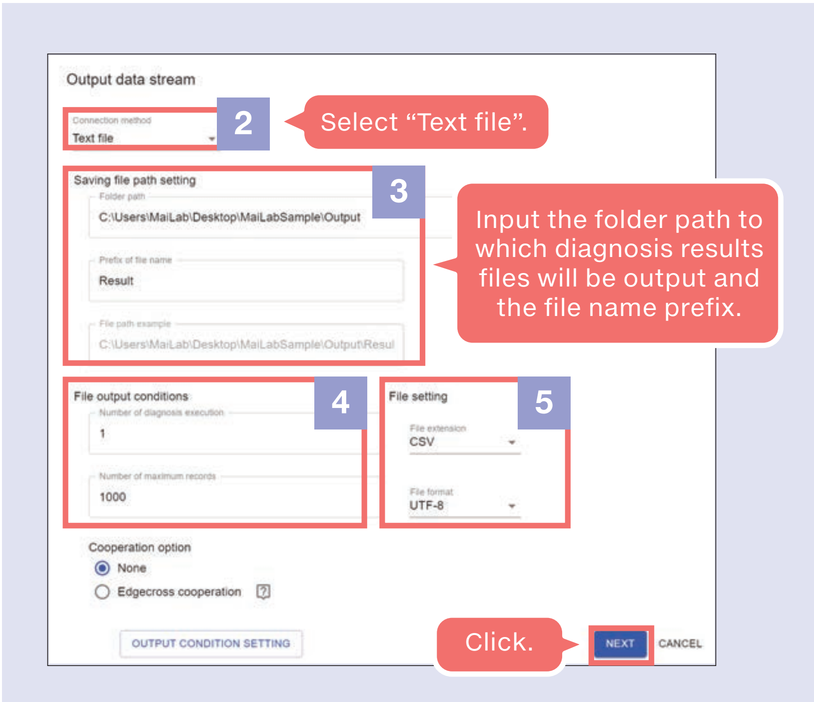
Assign input data (devices) to the explanatory variables (explanatory variables column on the right side of the Assignment screen) which will be input to the AI execution.

Step 3-A. Perform Output data stream settings (when outputting a text file)

The procedure for outputting diagnosis results to a text file will be explained.







Click the Output “CHANGE” button.

2 The Output data stream dialog will appear. Select “Text file” for Connection method.

3 Set the path to the Output folder and the file name prefix of the output files.

4 Set the file output conditions.

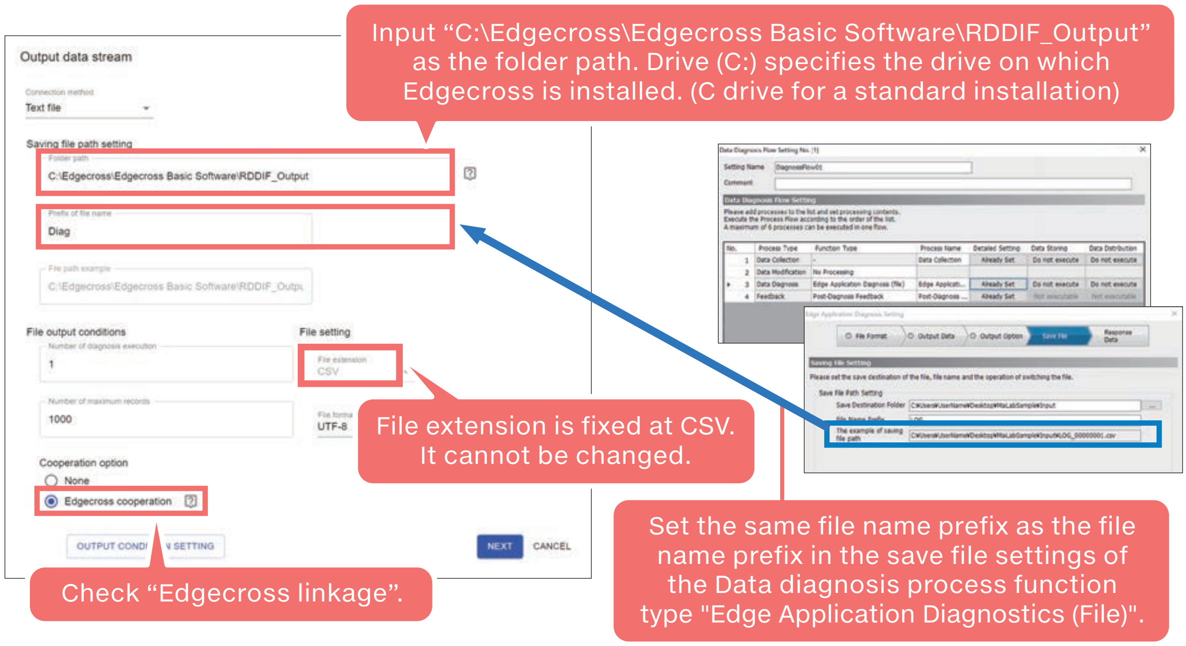
5 Select the output file extension and the file format, and click the “NEXT” button.

Diagnosis execution times

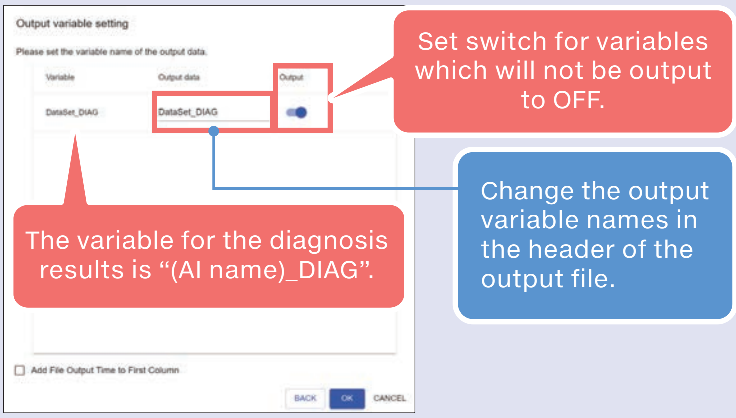
Specify the results file output timing by the diagnosis execution times.
To output every time diagnosis is executed, set 1 (time).

Edgecross linkage

Edgecross data flow can be linked with and operated by performing the settings shown in the following diagram, and the diagnosis results can be fed back to the devices.



6



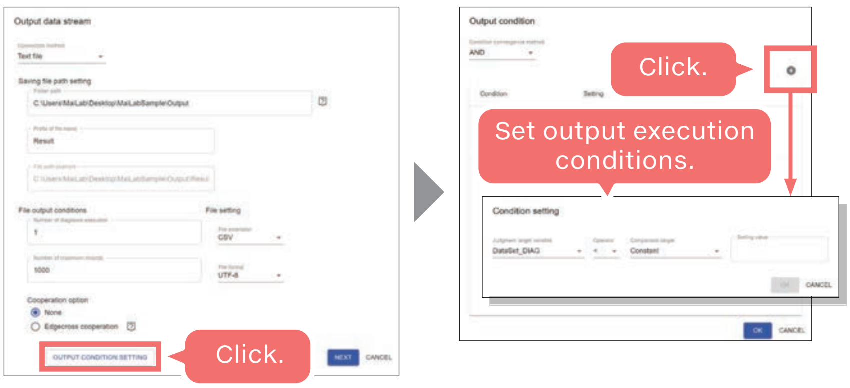
Select the variables to output and click the “OK“ button.

* Variables: The explanatory variables used for diagnosis results, threshold value judgment, and diagnosis.

Set output conditions.

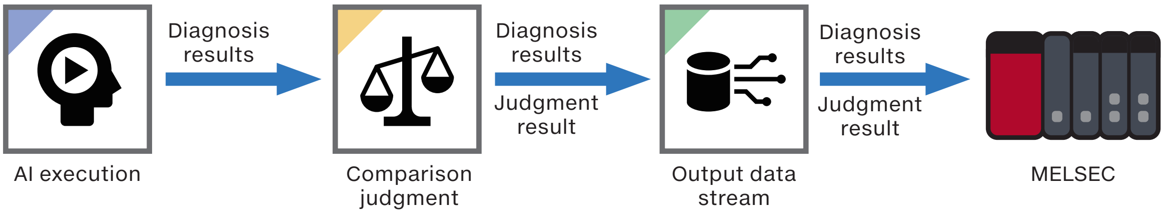
You can set conditions for executing data output, such as performing output only when the diagnosis results are NG or when the threshold value is exceeded, etc.

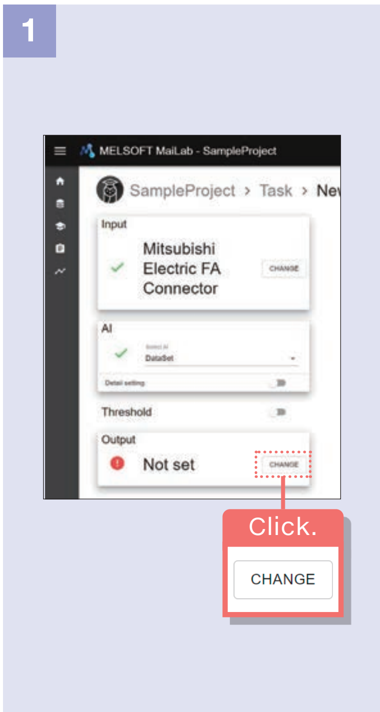
- Click the “OUTPUT CONDITION SETTING” button in the Output data stream settings screen.
- In the Output condition settings screen. click the “Add condition (+)” button.
- Set output execution conditions.



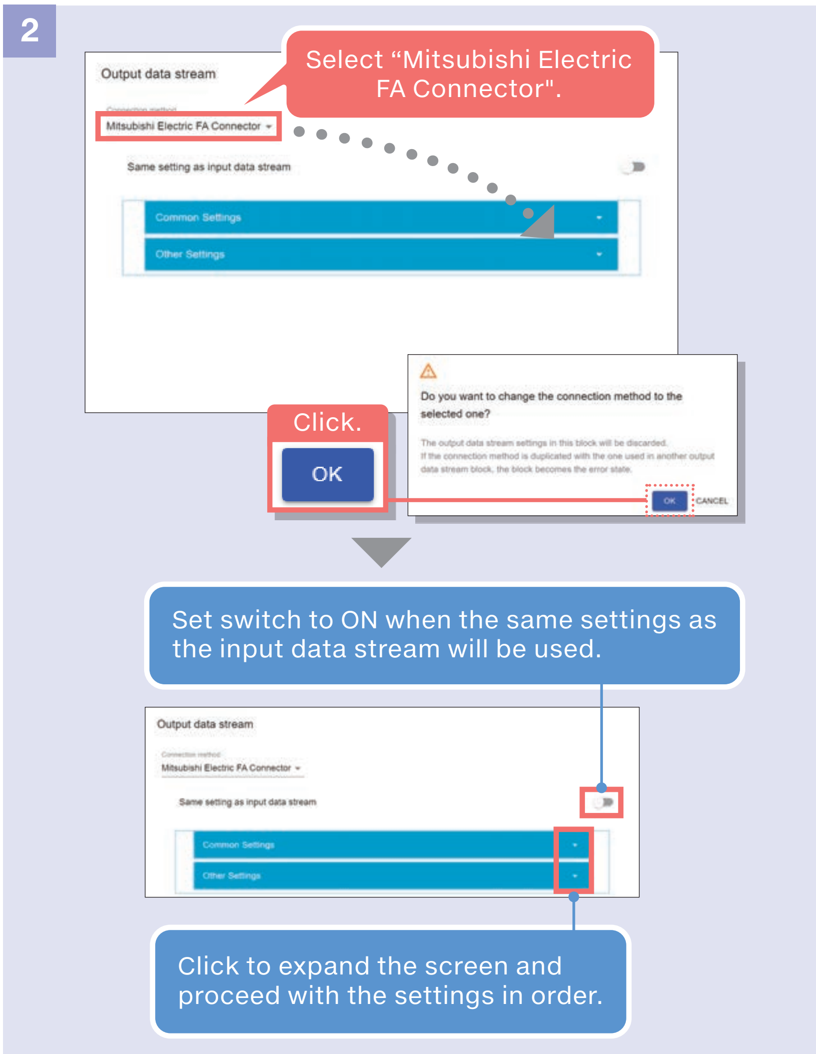
Step 3-B. Perform output data stream settings (when connecting to MELSEC)

The procedure for writing diagnosis results directly to MELSEC will be explained.



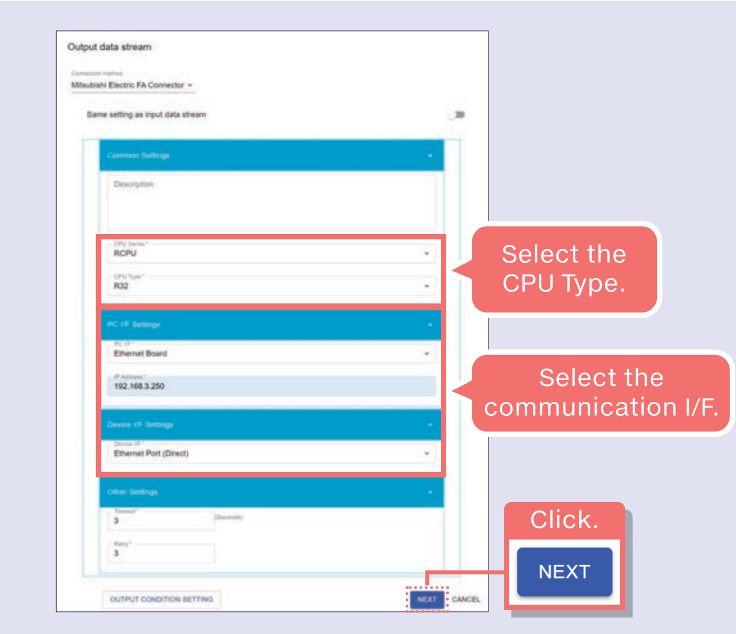


Click the Output “CHANGE” button.



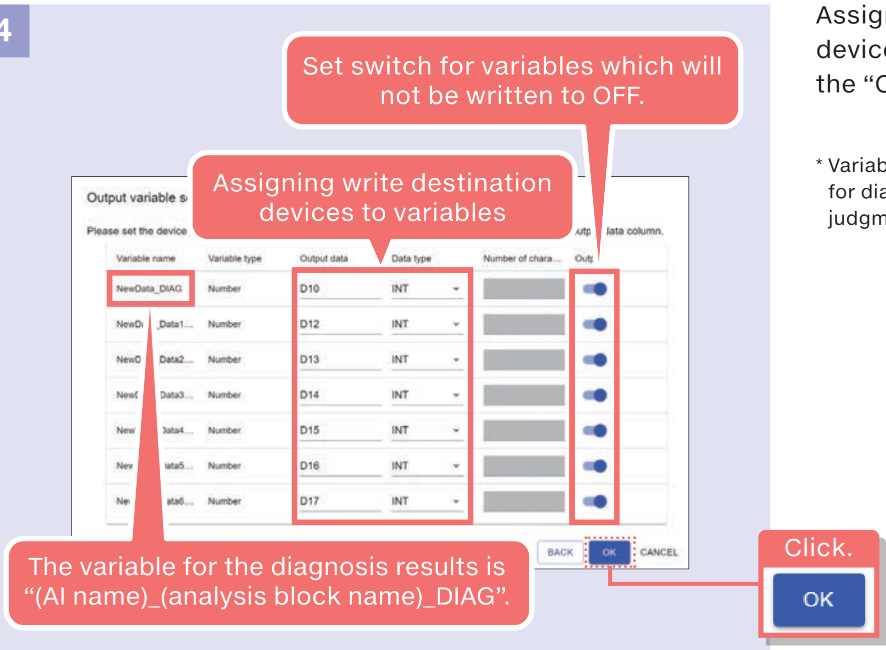
2 The Output Data Stream screen will appear. Select “Mitsubishi Electric FA Connector” for Connection method and click the “OK” button in the change confirmation dialog.

3



Set the model information and communication method for the write target, and click the “NEXT” button.

4



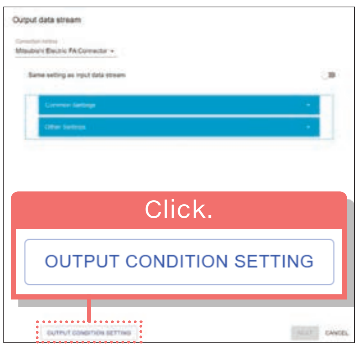
Assign the write destination devices to the variables and click the “OK” button.

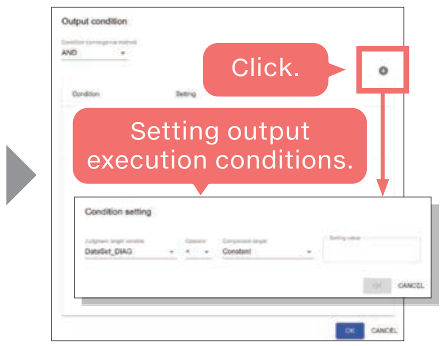
* Variables: The explanatory variables used for diagnosis results, threshold value judgment, and diagnosis.

Set output conditions.

You can set conditions for executing data output such as performing output (writing to MELSEC) only when the diagnosis results are NG or when the threshold value is exceeded, etc.

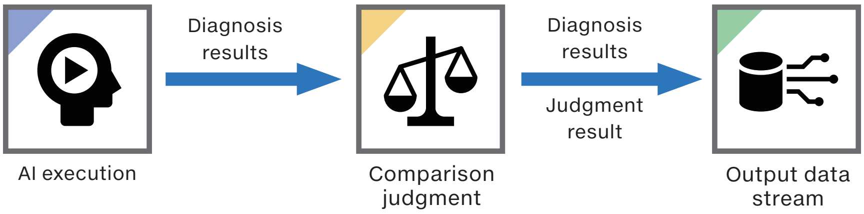
- Click the “OUTPUT CONDITION SETTING” button in the Output Data Stream Settings screen.
- In the Output condition settings screen. click the “Add Condition (+)” button.
- Set output execution conditions.

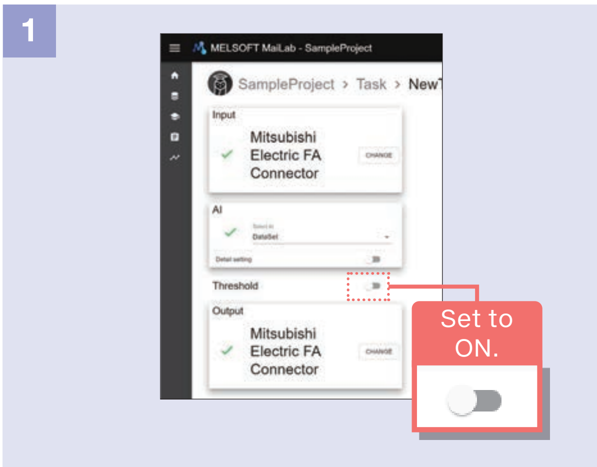




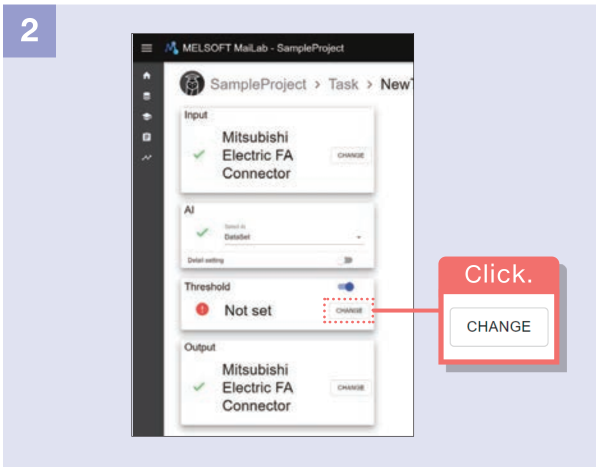
Step 4. Perform threshold value judgment settings

By performing threshold value judgment settings, the threshold value judgment results for the diagnosis results can be output.

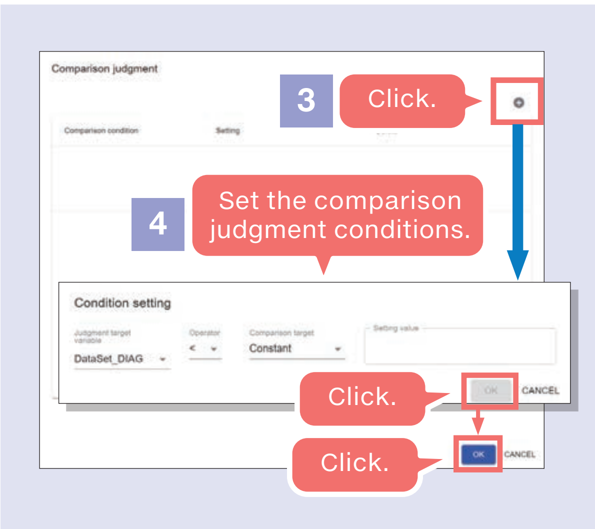




Set the Threshold switch to ON.



Click the Threshold “CHANGE” button.



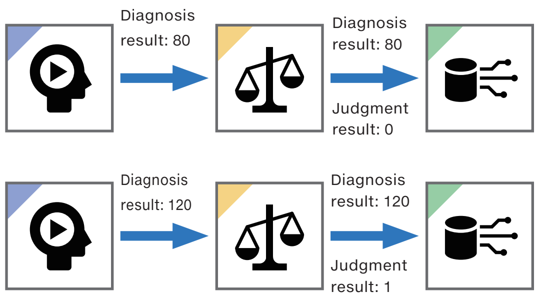
3 The Comparison Judgment Settings screen will appear. Click the “Add Comparison Condition Setting (+)” button.

4 Set the comparison judgment conditions.

Operation of threshold value judgment process

The threshold value judgment results will be additionally output.
“1” will be added when the diagnosis results meet the judgment conditions or “0” will be added when they do not meet the judgment conditions.

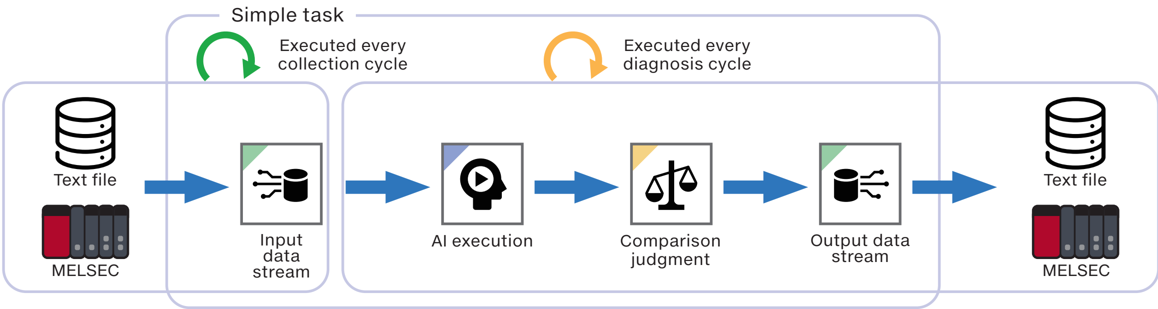
Example) Judgment condition setting: Diagnosis result > 100



It can be utilized for performing threshold value judgment on diagnosis results (estimation value) and performing OK/NG (0/1) feedback to the PLC, etc.

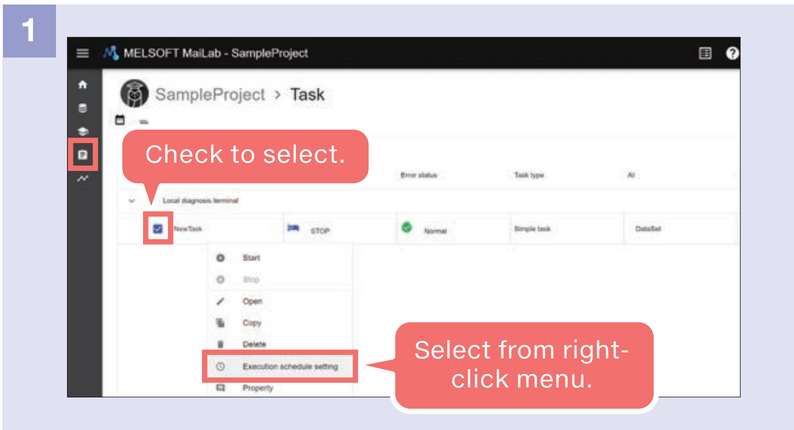
3.3.2 Executing the task

The task operation is started by performing the start operation.
When operation is started, data collection is performed at the set collection cycle, and diagnosis and results output based on the collected data is performed at the set diagnosis cycle.

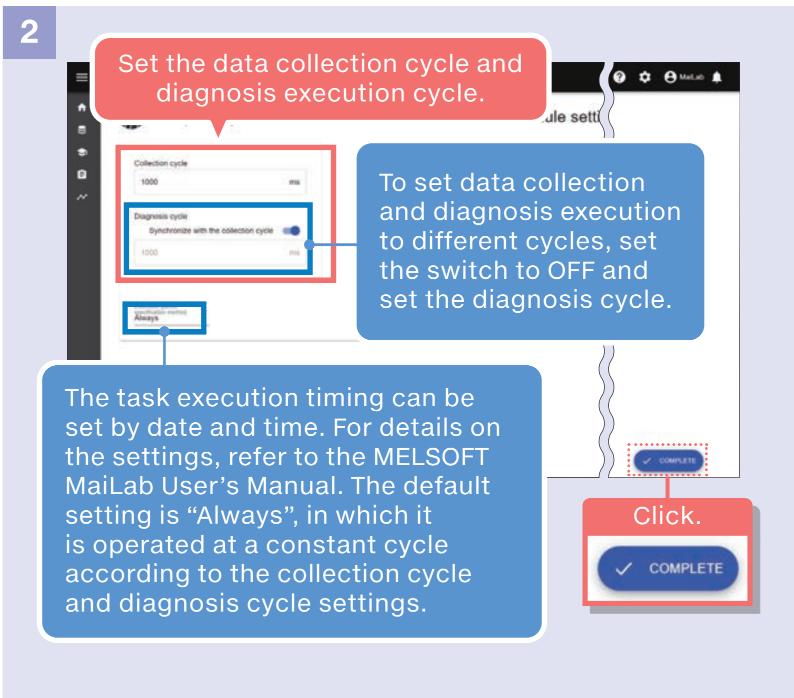


Step 1. Perform setting of the execution schedule

Set the data collection cycle and diagnosis execution cycle.



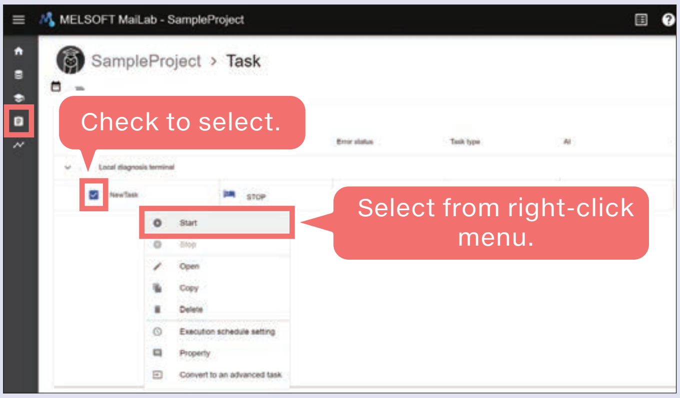
In the Task Management screen, select the target task and select “Execution schedule setting” from the right-click menu.



Set the data collection cycle and diagnosis execution cycle, and click the “COMPLETE” button.

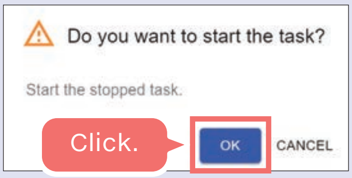
Step 2. Start task execution

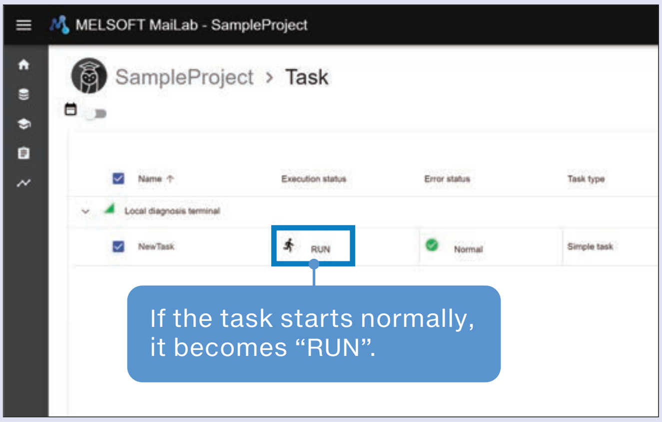
1



In the Task Management screen, select the target task and select "Start" from the right-click menu.

2





The Execution Confirmation dialog will appear. Click the "OK" button.

chapter 4

Creating an original diagnosis model

An AI can be easily and automatically created using the AutoML function.
You can also customize the AI or create an original AI.

4.1 Customizing the AI

In MaiLab, arrange the blocks representing each AI process and connect the blocks to prepare the AI processing flow.
You can use AutoML to edit the flow of the prepared AI, freely customize it, or create an original AI from scratch.

4.2 Using the various function blocks

Various types of blocks are available for performing ideal processing using AI.

4.3 Executing original processing using Python blocks

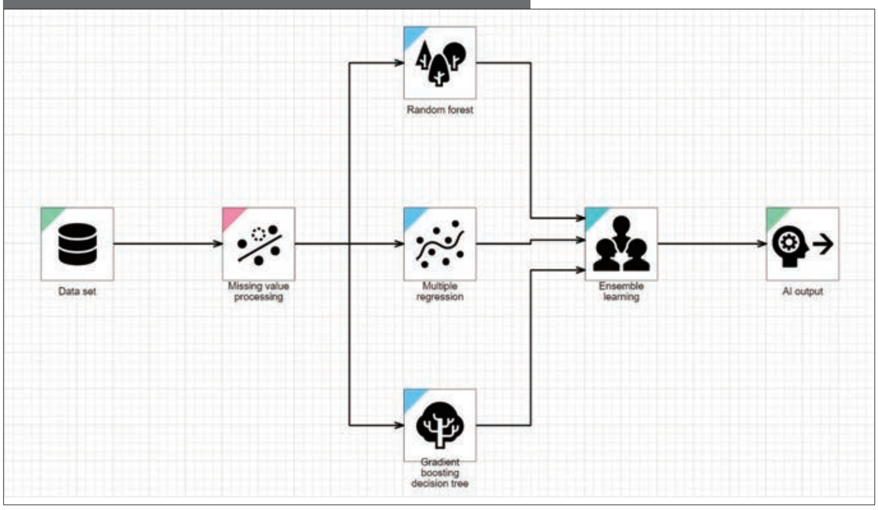
A Python block which can execute Python code as is is also available as a special function block.
User programs can be executed directly within the AI flow.

4.1 Customizing the AI

4.1.1 Customizing the AI

MaiLab can perform analysis easily and automatically using the AutoML function. In addition, since MaiLab can customize AI created using AutoML, AI models with higher accuracy can be constructed.

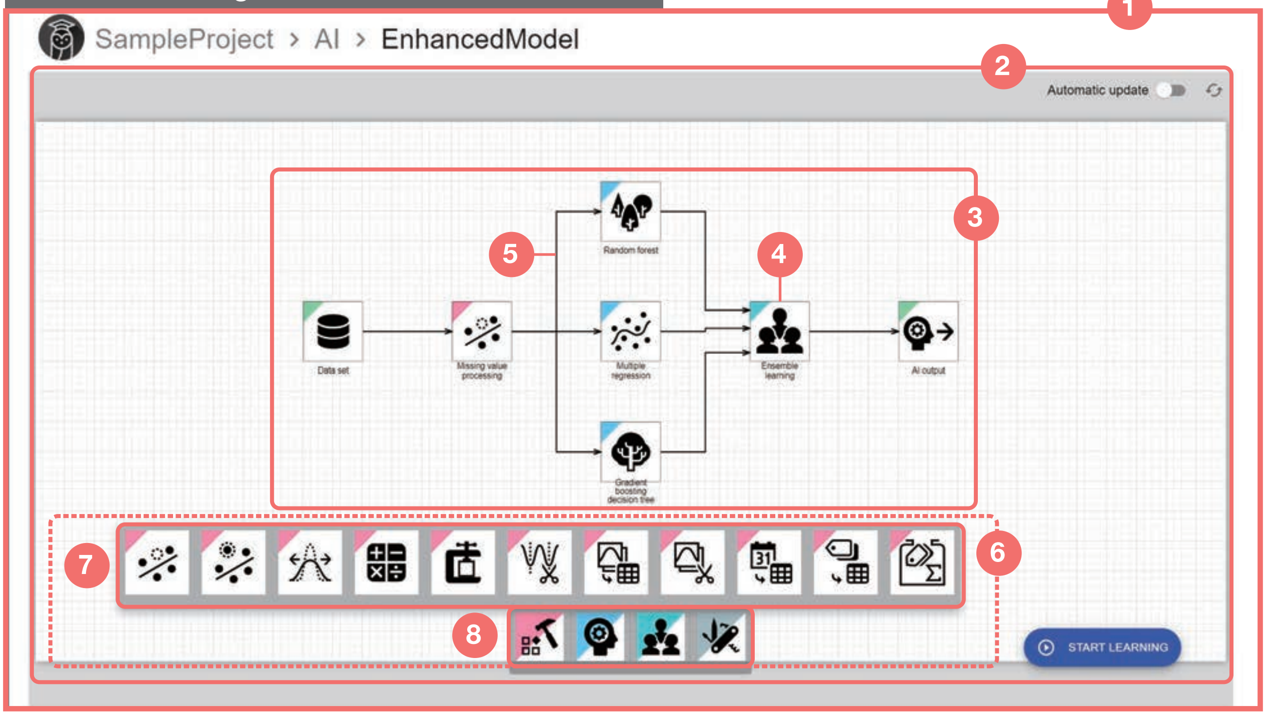
AI model used in this document



4.1.2 AI customization screen configuration and its content

The AI can be edited and freely customized using the dedicated editor. The editor screen configuration will be introduced.

Screen configuration of AI editor screen



1 AI editor screen

AI flow can be customized using the blocks and connectors described later.

2 Canvas

The place where blocks and connectors are arranged. It is shown like graph paper.

3 Process flow*

Consists of function expansion blocks and connectors. The data processing method is defined by connecting the output and input of blocks on the canvas with connectors.

*: When referring to only the process flow created and edited using the AI customize function, it is called the AI flow.

4 Block

Performs processes such as processing, analysis methods, etc. to the input data. The details of the processes performed are different for different blocks.

5 Connector

Connects one block to another. The arrow indicates the direction of data flow.

6 Dock

Stores the blocks that can be added to the process flow.

7 Block dock

Shows the blocks associated with the selected category. (In the diagram above, blocks in the preprocessing category.)

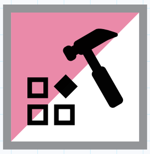
8 Category dock

Indicates the block category for the blocks shown in the block dock.

4.1.3 AI customization category dock and its content

Blocks are available for performing ideal processing using AI. In this paragraph, the types of blocks will be introduced.

Category dock and its content



▶ Preprocessing category

Refer to this paragraph for details.

Includes blocks that do pre-processing of input data in order to improve the accuracy of analysis performed downstream.

▶ Analysis method category

Refer to chapter 5 for details.

Includes blocks that execute each type of analysis method algorithm and output diagnosis rules.

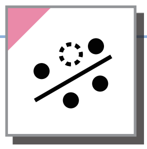
▶ Ensemble learning category

Includes ensemble learning blocks that combine multiple analysis methods and output a single diagnosis rule.

▶ Utility category

Includes blocks for various purposes such as changing processing content or input/output, etc. They can be installed at various places in the process flow.

■ Explanation of preprocessing category blocks



Missing value processing block

- Used to improve the accuracy of AI models.
- When there is a value missing from the data being used, it completes the data with an appropriate value so that the AI can learn properly.

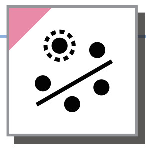
Input

Variable
10
11
9
8
Missing value
12

Completes data using average value for missing block.

Output

Variable
10
11
9
8
10
12



Outlier processing block

- Used to improve the accuracy of AI models.
- Since using an outlier value as a correct value reduces AI model accuracy, this block processes outlier values using appropriate methods.

Input

Variable
10
11
9
8
1000
12
7
6
5

Outlier value record is replaced with a median value.

Output

Variable
10
11
9
8
9
12
7
6
5

Scaling block

- Used when using methods such as deep learning that cannot be executed when variables with different numbers of digits are mixed together.
- Performs conversions such as standardization, etc. on numerical values to align the number of digits.

Input

Variable
1
2
3
4
5
6
7
8
9

Standardization example

Output

Variable
-1.460593
-1.095445
-0.7302967
-0.3651484
0
0.3651484
0.7302967
1.095445
1.460593

Numerical value operation block

- Used to improve the accuracy of AI models.
- Adds the result of arithmetic operation as a new variable. For example, the difference between 2 different variables is created and added as a new variable. The result of numerical arithmetic operation on 2 variables is added as a new variable.

Input

Variable 1	Variable 2
500	490
510	500
505	510
502	490
495	505
480	499

Output

Variable 1	Variable 2	(Variable 1) - (Variable 2)
500	490	10
510	500	10
505	510	-5
502	490	12
495	505	-10
480	499	-19

Variable 1 - Variable 2

Dimensionality compression block

- Used to improve the accuracy of AI models.
- Even if there are numerous variables, not all of them are necessarily useful, and conversely some may have a bad effect. The dimensionality compression block reduces the number of variables without losing important information of the numerical value data.

Input: There are 4 variables (not including objective variables) with numerical value as the variable type.

Variable 1	Variable 2	Variable 3	Variable 4	Category	Variable 5 (Objective variable)
14.23	1.71	2.43	15.6	A	5.64
13.2	1.78	2.14	11.2	B	4.38
13.16	2.36	2.67	18.6	C	5.68
14.37	1.95	2.5	16.8	A	7.8
13.24	2.59	2.87	21	B	4.32
14.2	1376	2.45	15.2	C	6.75
14.39	1.87	2.45	14.6	A	5.25
14.06	2.15	2.61	17.6	B	5.05

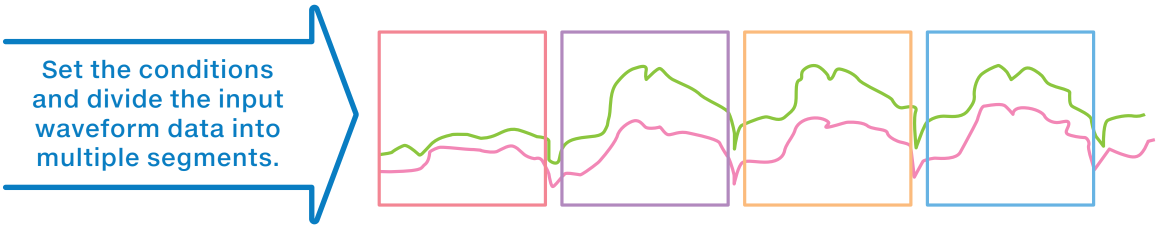
After compression there are 2 dimensions.

Output: Data after the variables with numerical value variable type undergo dimension compression.

Category	Variable 5 (Objective variable)	PCA_001	PCA_002
A	5.64	0.769	-0.394
B	4.38	5.110	0.884
C	5.68	-2.327	0.620
A	7.8	-0.444	-0.543
B	4.32	-4.741	0.454
C	6.75	1.158	-0.329
A	5.25	1.751	-0.450
B	5.05	-1.276	-0.242

Section generation block

- Used to divide waveform data into multiple intervals*.
- By setting the waveform start and end, the waveform is divided into separate waveform by similar patterns.
- MaiLab's specialty diagnosis diagnoses the features of waveforms with repeated similar patterns.

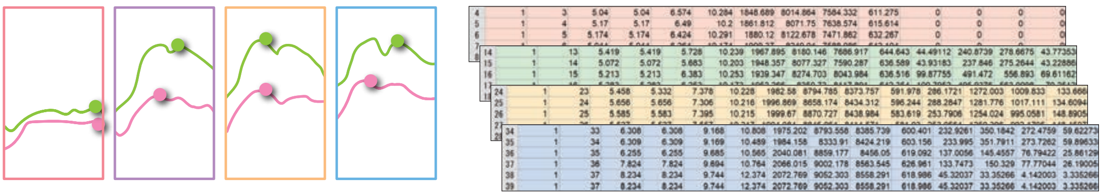


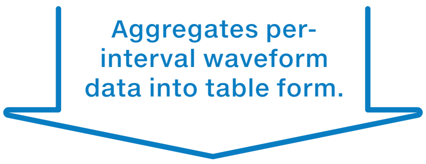
* Interval: For waveform data with repeated patterns, 1 segment of the waveform data divided by pattern is called an "interval" in MaiLab. In addition, the length of an interval is called "interval length".

Section feature generation block

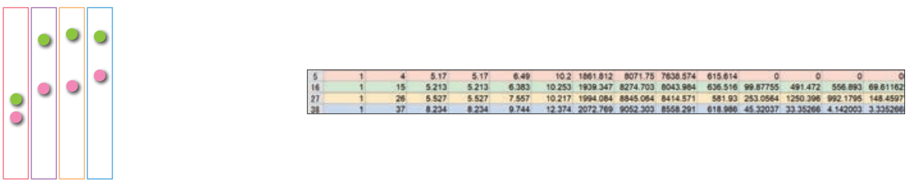
- Used to calculate waveform feature quantities for each interval.
- When diagnosing waveform features, the waveforms themselves are not diagnosed. Instead, the feature quantities (statistical quantities) that express the waveform features of each interval are used for diagnosis.
- MaiLab's specialty diagnosis diagnoses the features of waveforms with repeated similar patterns.

For each input interval, it calculates the feature quantities and aggregates them into 1 point.
(Example) Maximum value extraction





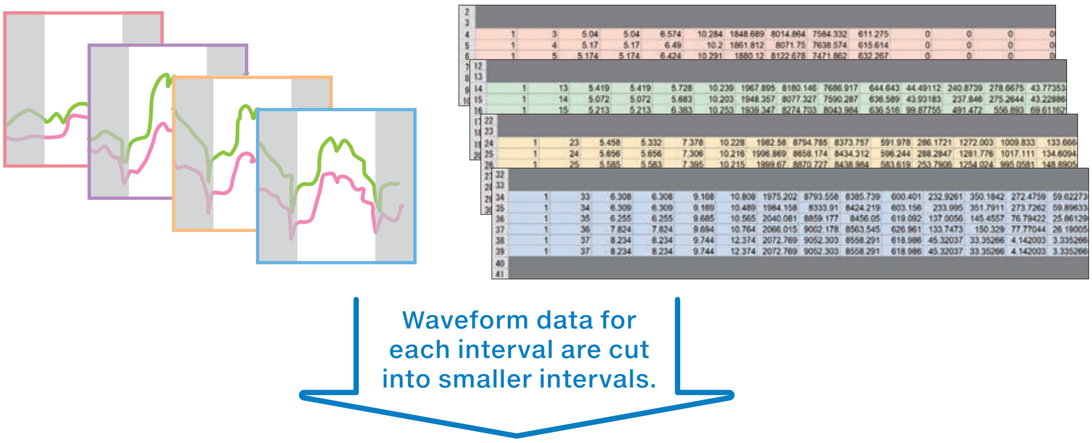
Calculates 1 record of feature quantities from 1 interval and converts them into table form.



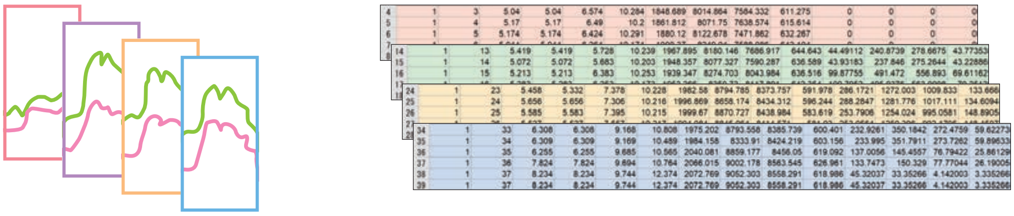
Sub-section forming block

- Used to improve the accuracy of AI models.
- Even when interval creation conditions are set in the section generation block, in some cases the expected interval cannot be created. The sub-section forming block further shapes the interval by setting shaping conditions to complement the section generation block.

Set the conditions and divide each input interval into smaller intervals.



It is divided into multiple intervals and output as ripple data in the same way as the input.



Date and time encoding block

- Used when creating variables of year, month, day, hour, minute, second.
- When data includes a variable of date/time information (time stamp), that information is processed to extract the year or day information, etc. By using date/time information, the data can be divided according to conditions, such as dividing data by day and night when tendencies are different between morning and night operation, dividing data by weekday and weekend, etc. Doing this may improve the accuracy of AI models.

Input: Data that includes date/time information

Date/time
2022/1/24 12:34:56
2022/1/24 14:34:56
2022/1/24 16:34:56
2022/1/24 18:34:56
2022/1/24 22:34:56
2022/1/25 0:34:56
2022/1/25 2:34:56
2022/1/25 4:34:56

Date/time encoding

Output: Data with day and time information divided into each value.

Date/time	Date/time_year	Date/time_month	Date/time_day	Date/time_hour	...	Date/time_dow
2022/1/24 12:34:56	2022	1	24	12	...	0
2022/1/24 14:34:56	2022	1	24	14	...	0
2022/1/24 16:34:56	2022	1	24	16	...	0
2022/1/24 18:34:56	2022	1	24	18	...	0
2022/1/24 22:34:56	2022	1	24	22	...	0
2022/1/25 0:34:56	2022	1	25	0	...	1
2022/1/25 2:34:56	2022	1	25	2	...	1
2022/1/25 4:34:56	2022	1	25	4	...	1

Category encoding block

- Used to improve the accuracy of AI models.
- Since some analysis methods cannot use category type data (character string information such as part material during manufacturing, etc.), this block converts it to numerical value data. Information to be used for learning can be extracted from category variables as well and used.

Input: Character string information

Product
Arm
Body
Door
Arm
Arm
Cube
Body
Cube
Arm

Category encoding

Output: Converted to numerical value data

Product_enc
0
1
3
0
0
2
1
2
0

Add category statistic block

- Used to improve the accuracy of AI models.
- In addition to using numerical value data (other than category variables) as is, statistical quantities can be calculated from the numerical value data for each of the same category type of category data. By adding the calculated statistical quantities as new variables, the variables that can be used for a method can be increased.

Input: Data containing multiple records with the same category variable.

Area	Amount
Aichi	100
Brazil	1100
Dominica	23
Aichi	90
Aichi	130
Cuba	400
Brazil	1200
Cuba	350
Aichi	110

Category statistics addition

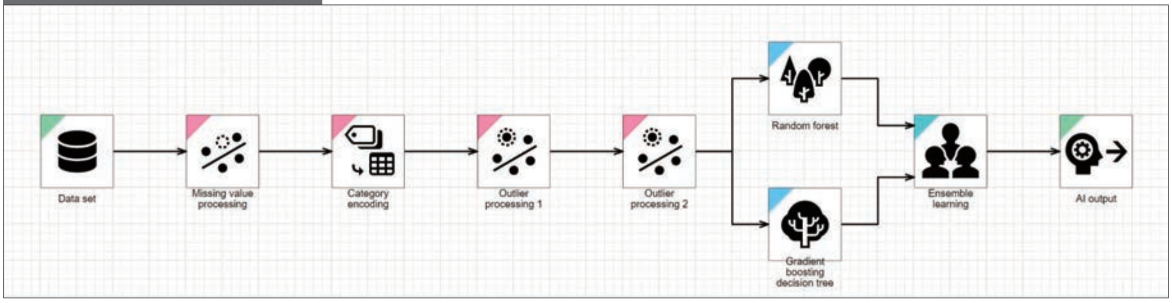
Output: Statistical quantity variables calculated for each category variable are added.

Area	Amount	Area_Amount_max	Area_Amount_std
Aichi	100	130	16.99
Brazil	1100	1200	50
Dominica	23	23	0
Aichi	90	130	16.99
Aichi	130	130	16.99
Cuba	400	400	25
Brazil	1200	1200	50
Cuba	350	40	25
Aichi	110	130	16.99

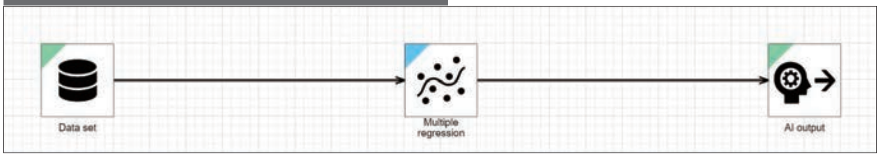
4.1.4 Constructing an original AI model

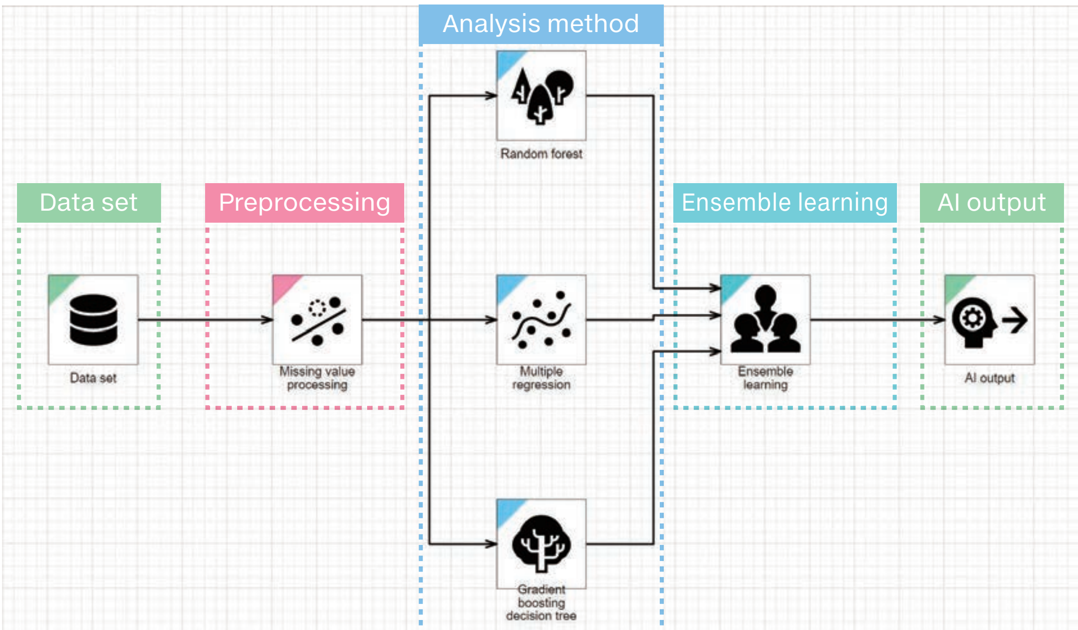
By editing an AI created by AutoML or a “Manual” created AI, the original AI model shown below can be constructed.

AI created by AutoML



“Manual” created AI





About block arrangement

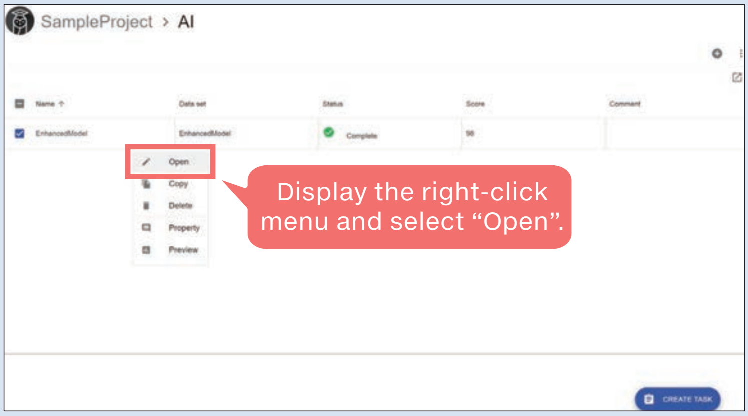
- The basic block sequence is as follows:

Data set ▶ Preprocessing ▶ Analysis method ▶ Ensemble learning ▶ AI output

- Utility blocks can be installed freely anywhere after the data set block.
- Preprocessing category blocks are installed as necessary.
- To use multiple analysis methods, install multiple analysis methods category blocks.
- When multiple analysis methods category blocks have been installed, install an ensemble learning block.

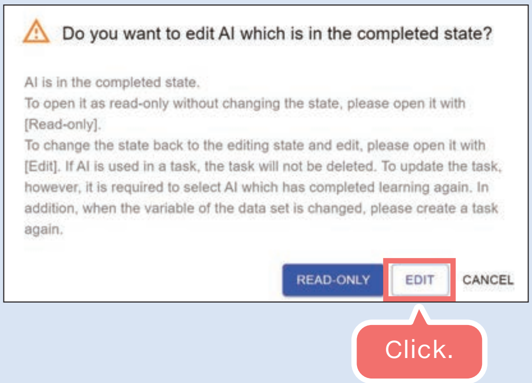
Here, the procedure for editing a “Manual” created AI and constructing the original AI model shown previously will be explained.

1



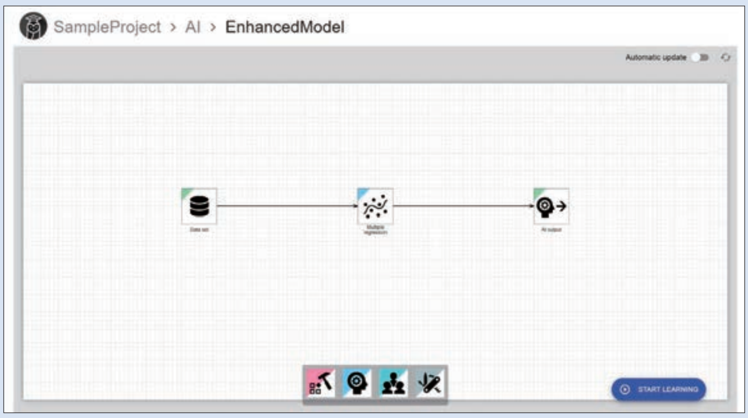
Select an AI that was previously created by AutoML and select “Open” from the right-click menu.

2



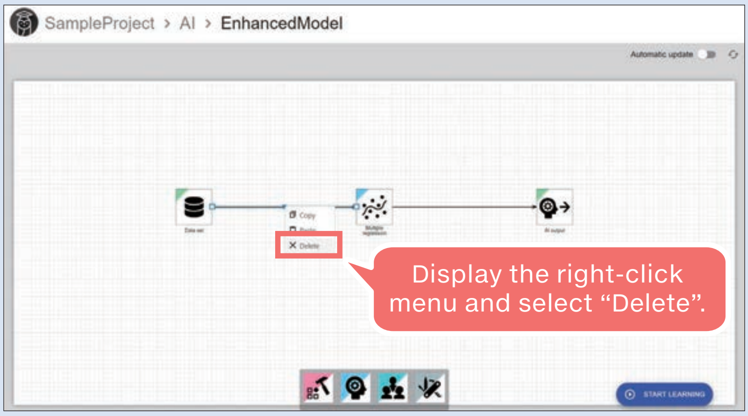
A pop-up will be displayed. Click the “EDIT” button.

3

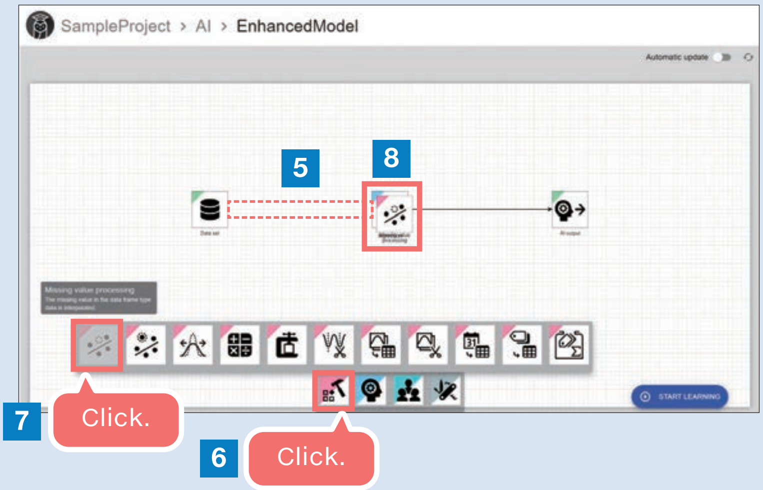


The AI Customization screen will appear.

4



Click on the connector to delete and select “Delete” from the right-click menu.



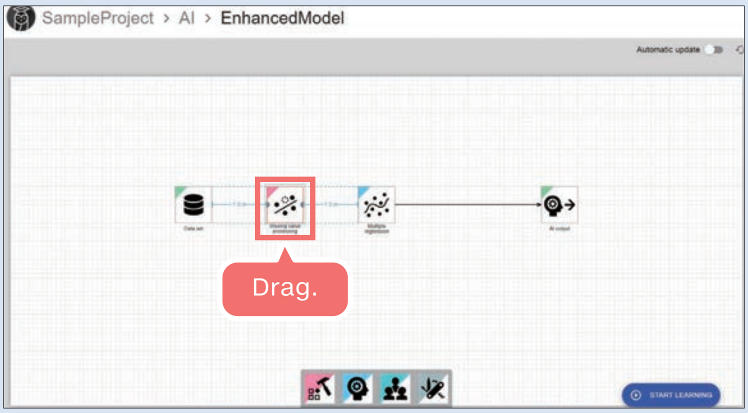
5 The connector was deleted.

6 Click on a “Preprocessing” category block.

7 Click on the “Missing value processing” block.

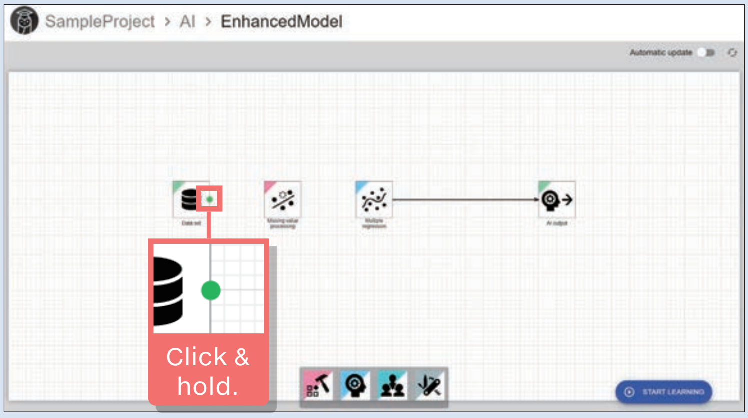
8 A “Missing value processing” block will appear near the center of the canvas.

9

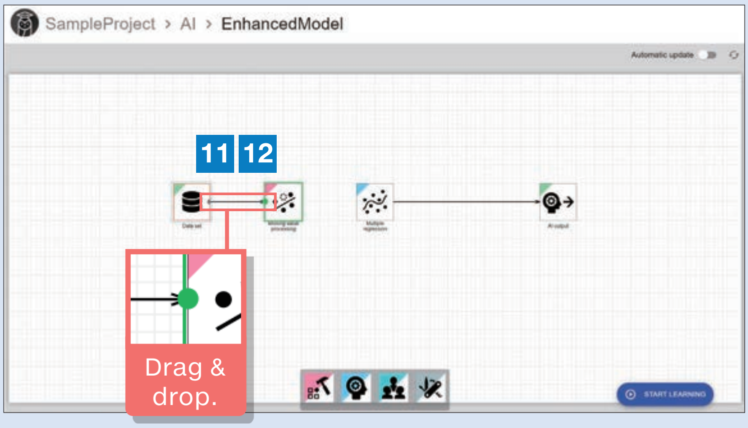


Drag the “Missing value processing” block to the desired location.

10



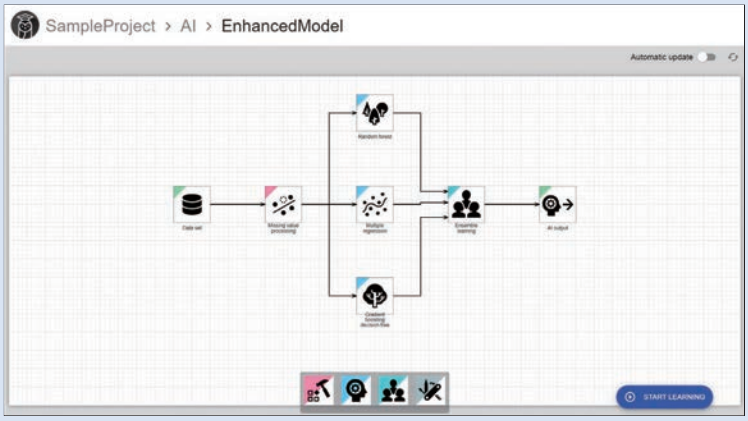
When the pointer is placed over the upstream block of data, a ● will appear on the right edge of the block. Click & hold the ●.



11 Drag it to the left edge of the block (on the downstream side of the data) that you want to connect to. An arrow connector following the pointer will appear.

12 A ● will appear on the left edge of the block. Dropping on top of the ● will connect the blocks to each other.

13



Repeat **4** to **12** to customize the AI model and create an original AI model.

4.2 Using the various function blocks

4.2.1 Customizing tasks

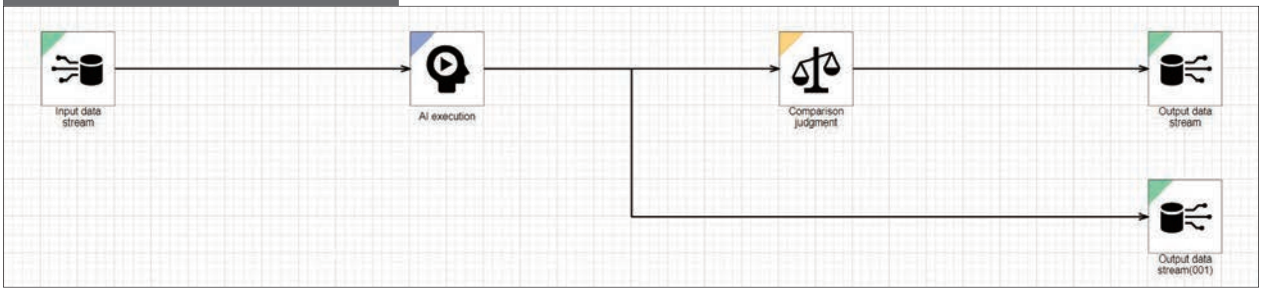
Tasks* can be created in MaiLab using simple functions and advanced functions.

- Simple task functions: Tasks can be created automatically using simple settings.
- Advanced task functions: Processes that cannot be achieved using simple tasks, such as performing diagnosis by applying multiple comparison judgments, outputting diagnosis results to multiple output destinations, etc. can be added.

Tasks created with simple task functions can be converted into advanced tasks and customized. Operations to create advanced tasks is the same as the operations used to construct original AI models.

*Task: A group of processes that diagnoses input data using AI and outputs diagnosis results.

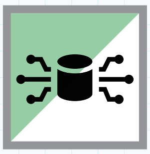
Advanced task example



4.2.2 Category dock and its content for creating tasks

Blocks are available for performing ideal processing using a task. In this paragraph, the types of blocks will be introduced.

Category dock and its content



▶ Data stream category

Refer to the next page for details.

Includes blocks that perform data input and output during task execution.

▶ Diagnosis category

Refer to chapter 5 for details.

Includes blocks that monitor input data and output diagnosis rules.

▶ Utility category

Includes blocks for various purposes such as processing data in the processing flow, etc.

■ Explanation of data stream category blocks

Input data stream block

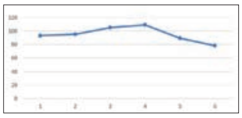
- Used to perform settings of collection sources for data to be diagnosed.
- Arranged on canvas in advance. So there is no need to arrange it from the dock.

Output data stream block

- Used to set output method and output destination for diagnosis results when executing a task and outputting diagnosis results.
- Multiple output methods and output destinations can be set.

Comparison judgment block

- Used to compare diagnosis target variables to threshold values or comparison conditions and judge whether or not they deviate from conditions.
- Multiple settings for comparison settings can be made.
- Judgment results are output in the form of additional columns joined to input data.
- Not only numerical values but also category values can be set as diagnosis target variables.



TimeStamp	Value
2021/2/5 14:00:00.01	93
2021/2/5 14:00:00.02	95
2021/2/5 14:00:00.03	105
2021/2/5 14:00:00.04	109
2021/2/5 14:00:00.05	89
2021/2/5 14:00:00.06	78

Waveform data

Comparison judgment conditions

- Setting 1: value > 100
- Setting 2: value < 80

Setting 1 result	Setting 2 result
0	0
0	0
1	0
1	0
0	0
0	1

Join.

TimeStamp	Value	Setting 1 result	Setting 2 result
2021/2/5 14:00:00.01	93	0	0
2021/2/5 14:00:00.02	95	0	0
2021/2/5 14:00:00.03	105	1	0
2021/2/5 14:00:00.04	109	1	0
2021/2/5 14:00:00.05	89	0	0
2021/2/5 14:00:00.06	78	0	1

The results obtained by logical combination are joined to input data and output.

Name	Gender	Position	Affiliation
Tanaka	Man	Manager	Sales
Hayashi	Woman	Chief clerk	Sales
Kojima	Man	Department	Sales
Ito	Man	General	Clerical work
Yamada	Man	General	Exploitation
Saito	Woman	Chief	Exploitation
Yamamoto	Woman	Chief	Exploitation

Category data

Comparison judgment conditions

- Setting 1: Affiliation = Administration, Development
- Setting 2: Position ≠ Section chief, Department manager

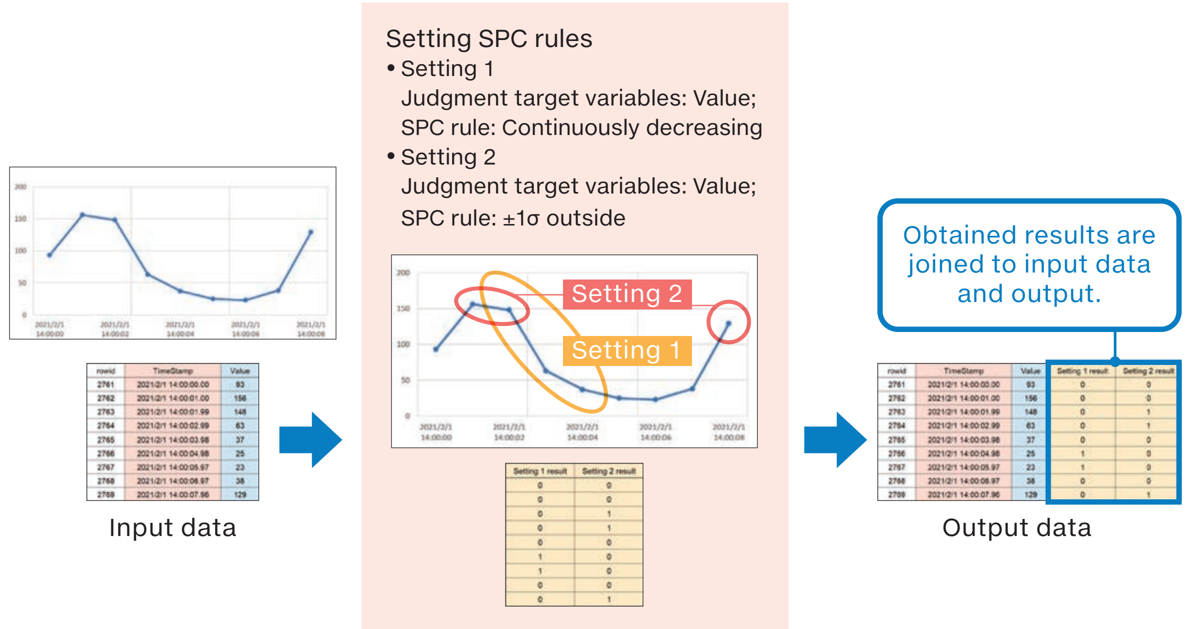
Setting 1 result	Setting 2 result
0	1
0	0
0	0
1	1
1	1
1	1
1	1

Join.

Name	Gender	Position	Affiliation	Setting 1 result	Setting 2 result
Tanaka	Man	Manager	Sales	0	1
Hayashi	Woman	Chief clerk	Sales	0	0
Kojima	Man	Department	Sales	0	0
Ito	Man	General	Clerical work	1	1
Yamada	Man	General	Exploitation	1	1
Saito	Woman	Chief	Exploitation	1	1
Yamamoto	Woman	Chief	Exploitation	1	1

SPC judgment block

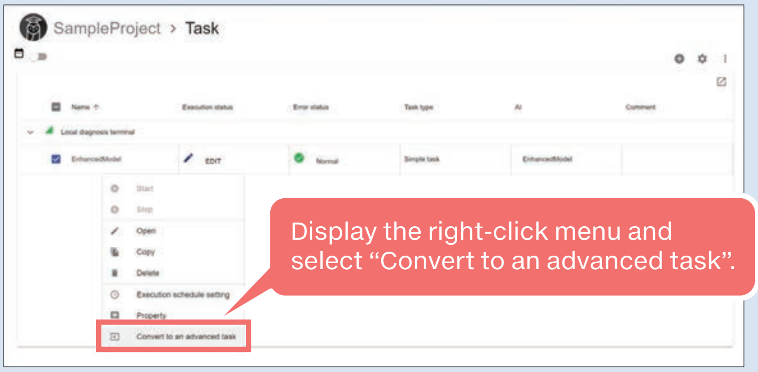
- Used to apply rules based on SPC (Statistical Process Control) when comparing diagnosis target variables to threshold values or comparison conditions and judging whether or not they deviate from conditions.
- Multiple SPC rules can be set.
- Judgment results are output in the form of additional columns joined to input data.



4.2.3 Creating simple to advanced tasks

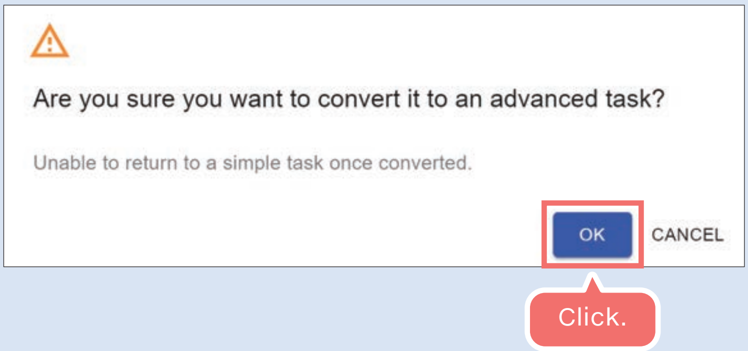
This paragraph will show the procedures for converting a task created using simple task functions into an advanced task and customizing it.

1



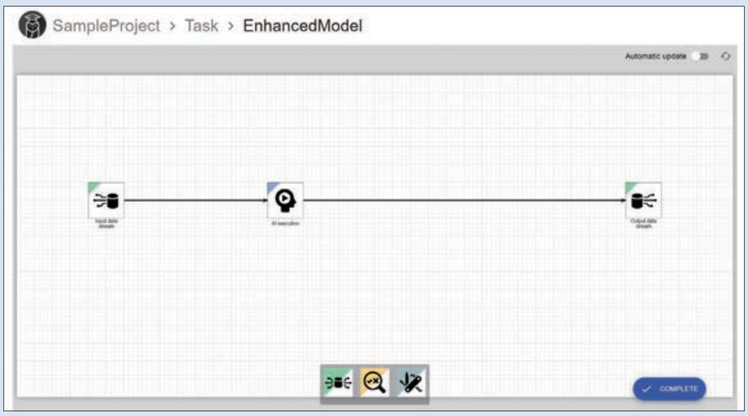
Select a task that was previously created by Simple task functions and select "Convert to an advanced task" from the right-click menu.

2



A pop-up will be displayed. Click the "OK" button.

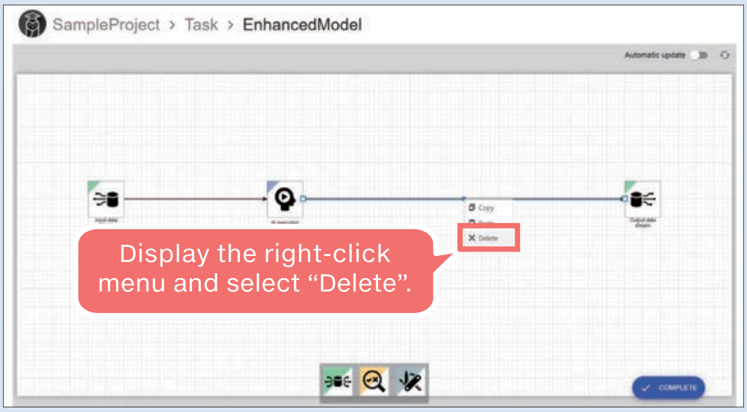
3



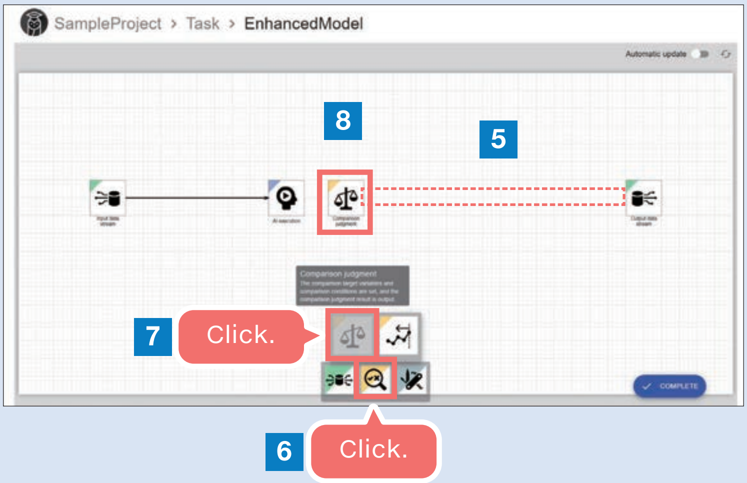
The task flow will be displayed.

* The screen configuration is the same as the "AI Customization Screen".

4



Click on the connector to delete and select “Delete” from the right-click menu.



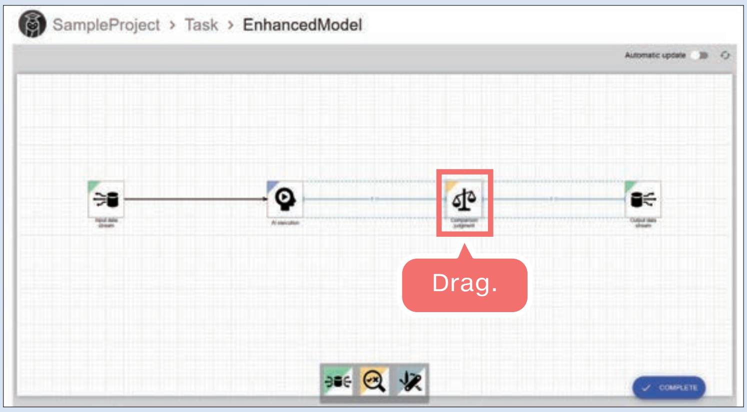
5 The connector was deleted.

6 Click on a “Diagnosis” category block.

7 Click on the “Comparison Judgment” block.

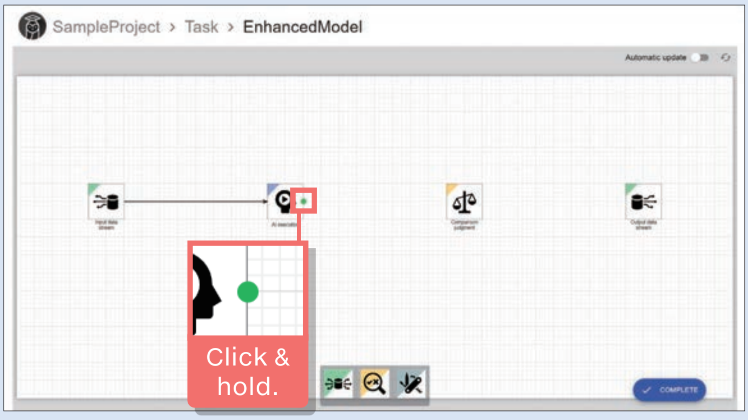
8 A “Comparison Judgment” block will appear near the center of the canvas.

9

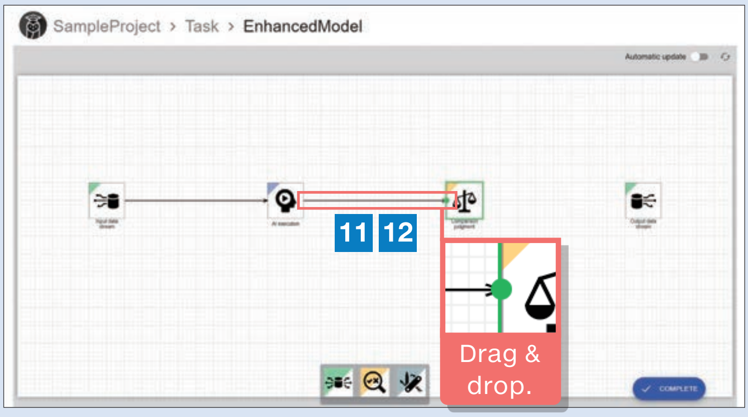


Drag the “Comparison Judgment” block to the desired location.

10



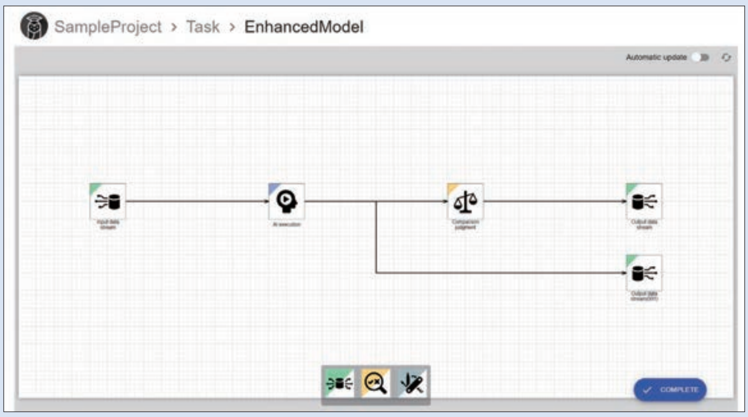
When the pointer is placed over the upstream block of data, a ● will appear on the right edge of the block. Click & hold the ●.



11 Drag it to the left edge of the block (on the downstream side of the data) that you want to connect to. An arrow connector following the pointer will appear.

12 A ● will appear on the left edge of the block. Dropping on top of the ● will connect the blocks to each other.

13



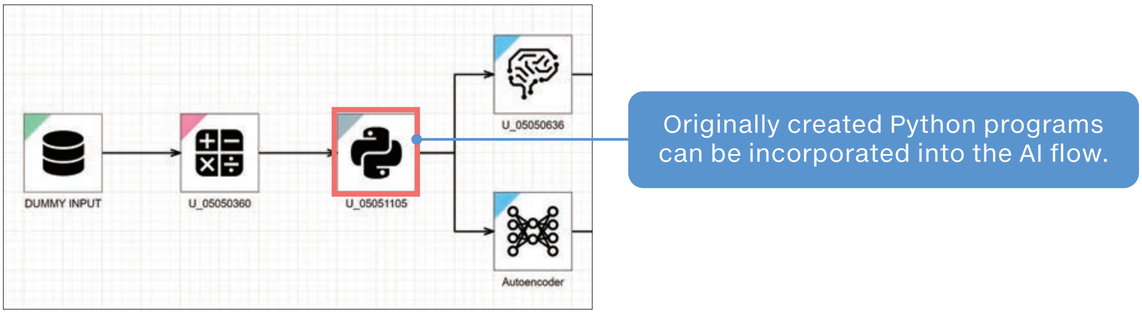
Repeat **4** to **12** to customize the simple task and create an advanced task.

4.3 Executing original processing using Python blocks

So far, the flow of general data analysis and methods for improving accuracy have been introduced. However, in some cases original processing is necessary depending on the data and targets being handled.
For example, consider the following situations:

- When it is necessary to use information such as product lot numbers that include product type information or values calculated from sensor data using a certain formula and used for manufacturing, etc. as new feature quantities.
- When data such as FFT (fast Fourier transform) itself will be transformed and it is necessary to perform the transformation between the reading in of data and its input into the analysis method block.

In MaiLab, original processing can also be incorporated into analysis and diagnosis flows according to individual needs.

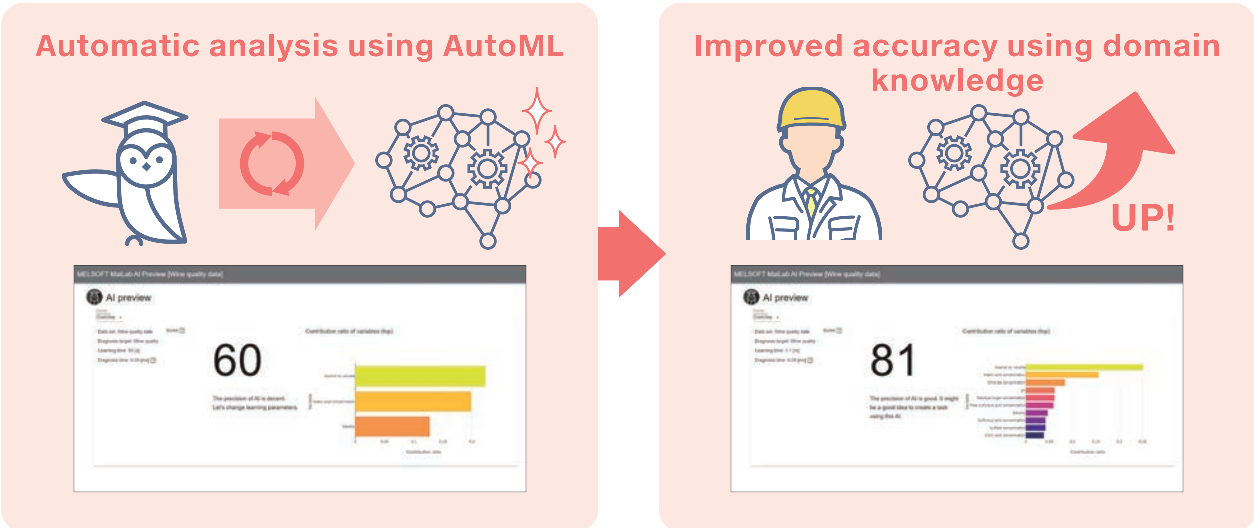


chapter 5

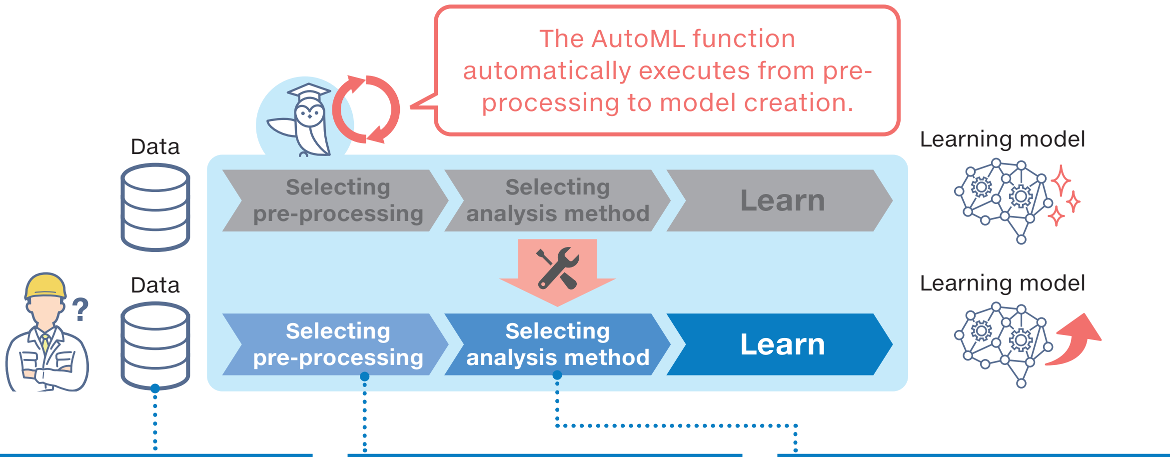
Improving the accuracy of the diagnosis model

In this chapter, methods for improving the accuracy of diagnosis models created using the AutoML function will be introduced.

By utilizing the AutoML function, diagnosis models can be easily and automatically created without the need for specialized knowledge or manual data processing. Accuracy can be further improved by incorporating knowhow such as characteristics related to the prediction targets or data, etc. into the learning model automatically created using the AutoML function.



Here are 3 points to review in order to improve diagnosis accuracy.
Increase the power of learning models created using the AutoML function by implementing improvement measures corresponding to each situation.



5.1 Checking the prepared data

5.1.1 Checking whether the data are as expected
5.1.2 Expanding the data set

5.2 Checking and adding pre-processing

5.2.1 Cleansing data
5.2.2 Feature quantity engineering: (1) Joining data to create new data
5.2.3 Feature quantity engineering: (2) Taking a specific part of the data and extracting features

5.3 Checking learning parameters and analysis methods

5.3.1 Selecting the analysis method

5.1 Checking the prepared data

If there are deficiencies in the prepared data (insufficient data quantity, inclusion of error data, etc.), the accuracy of the diagnosis model will be low.
First, check that the expected data are included, that there is no extreme bias in the data, etc. Also, if additional data can be provided, expand the data set.

5.1.1 Checking whether the data are as expected

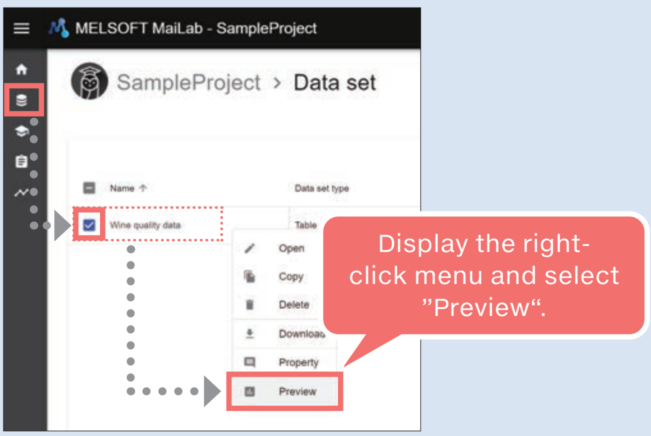
If unexpected data are included, correct analysis and learning cannot be performed.
Check for the following situations:
- Data does not include necessary variables.
- Lots of data from when devices were stopped are included, and there are few meaningful data.
- There is bias in the data distribution.

■ Data check methods and relearning procedures

An example of a method for checking data and procedures from data set re-registration to relearning will be introduced.

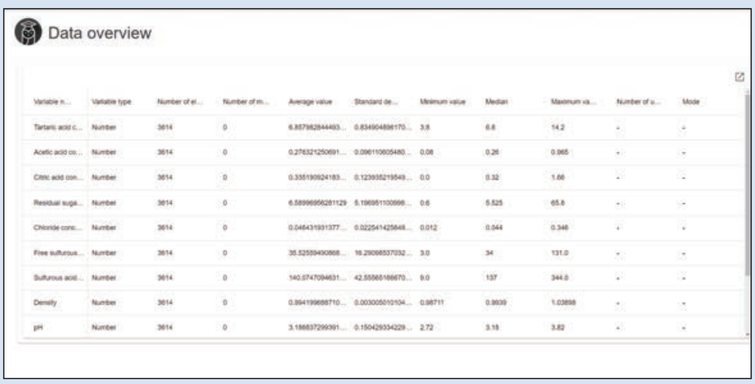
(1) Check for "Data does not include necessary variables".

1



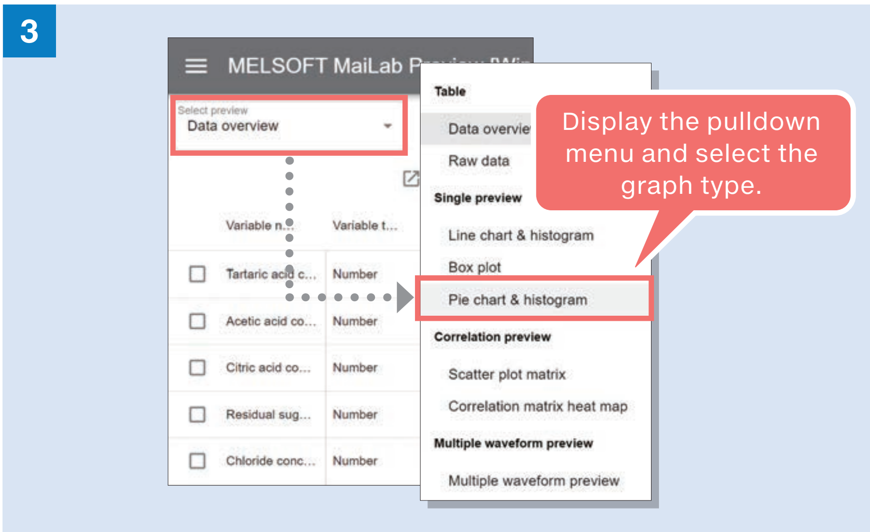
In the Data Set Management screen, select the data set to check and select "Preview" from the right-click menu.

2



In the Data Overview screen, check whether "Data includes necessary variables", "Variable types are as expected", etc.

(2) Check whether “There is bias in the data distribution.”



From the preview selection pulldown menu, select the graph type to display.

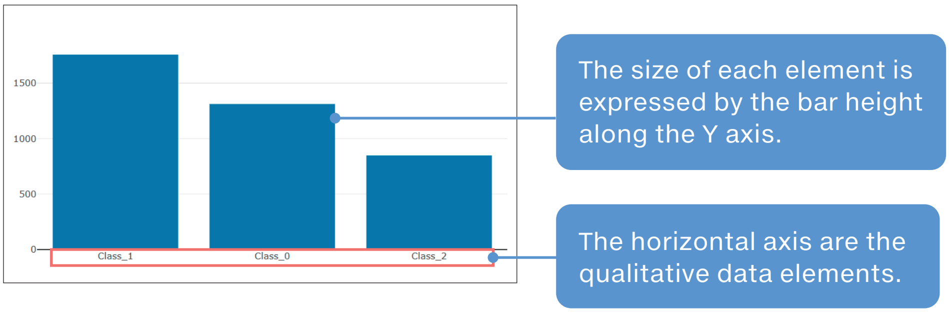
Select the variables to graph and check whether “There is bias in the data” using the pie chart, histogram/bar graph, etc.

* Histogram: If the display target are category variables, a bar graph will be used.

Suitable visualization methods for category type data

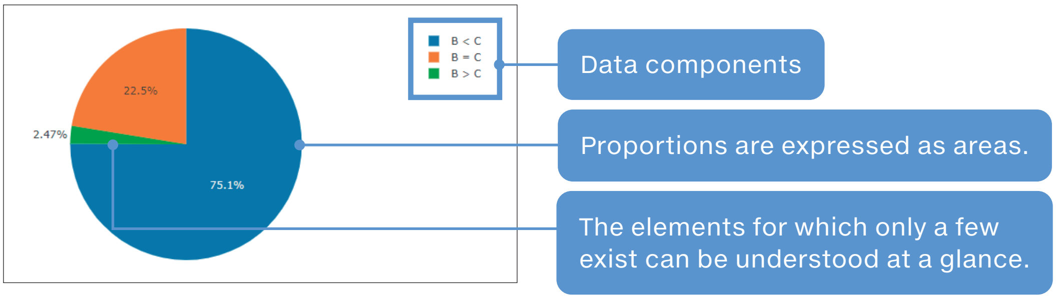
Bar graph

A graph suitable for comparing the size relationships of data for each category type element. Category type elements are assigned along the horizontal axis, and the data size for each element are expressed as bar height in the vertical axis. It is used for checking the size relationships of various data, such as monthly sales trends of each product, the number of times an element appears in the data, etc.



Pie chart

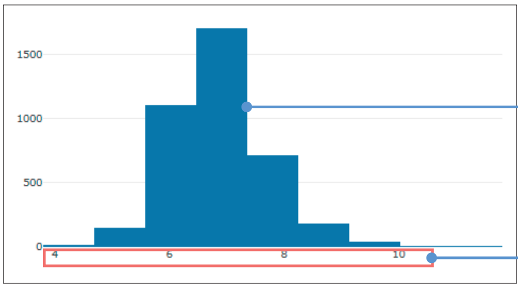
A graph in which the components of data and their proportions can be checked at a glance.



Suitable visualization methods for numerical value type data

Histogram

Graph expressing the number of numerical types (frequency of occurrence) The division ranges (classes) of quantitative data are assigned along the horizontal axis, and the frequency of quantitative data (frequency of appearance) within each division range is assigned along the vertical axis. The number of divisions along the horizontal axis is called the bin number. By setting appropriate values for each data, how quantitative data are distributed can be known.

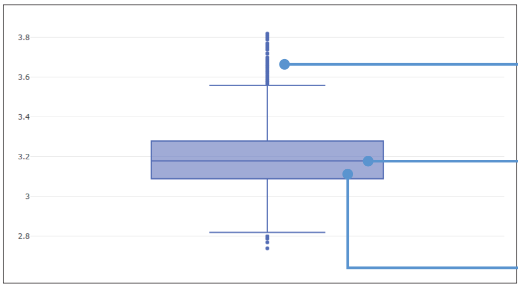


The frequency of quantitative data within each class is assigned along the vertical axis.

The classes of quantitative data are assigned along the horizontal axis.

Box plot

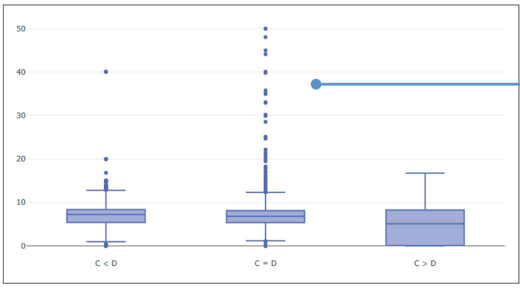
A graph that expresses the variability of numerical value data and the presence/absence of outlier values under certain conditions in an easy-to-understand way.
The variations within each category can be compared.



The presence/absence of outlier values from the overall distribution can also be seen at a glance.

The line inside the box is the data median value.

The box encloses 50% of all data.



Comparison of each category
Example:When C > D, the values are clustered to low values, but when C = D, the values are scattered to large values. There are differences in variability between C < D and C = D, but almost all data are clustered near the median value.

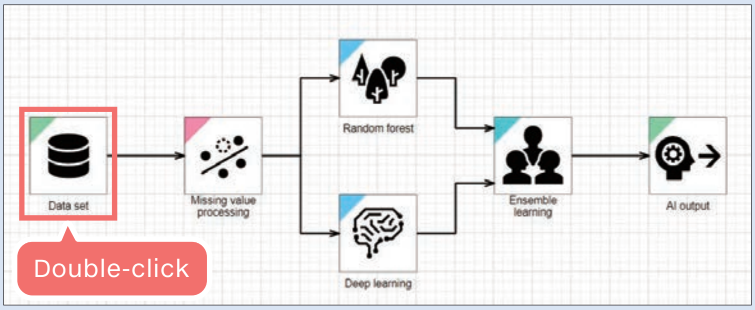
(3) Performing relearning with a correct data set

When it is necessary to revise data, perform relearning using a data set in which correct data are registered.

1 Create a new data set using the revised data. 3.1 Creating the data set

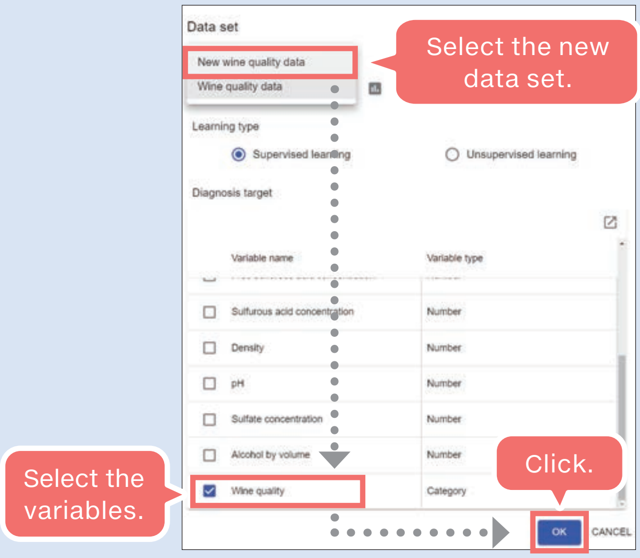
*When creating the AI automatically using the AutoML function, the following operation is unnecessary.
Execute the procedure in “3.2 Creating the AI” again.

2



Open the AI in the AI Editor and double-click on the data set block to open the properties.

3

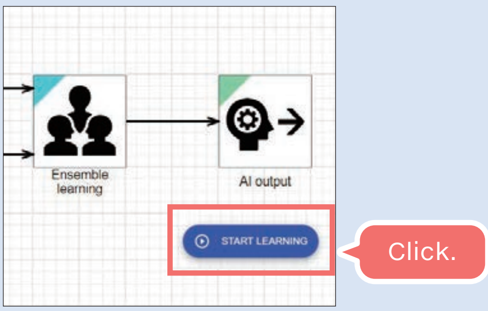


Change to the data set in which the correct data are registered, and set the objective variables again.

When variables have been added:
When the added variables will also be used in downstream function blocks, it is necessary to add the process to each block.

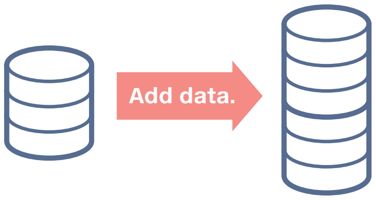
When variables have been deleted:
When the deleted variables were also used in downstream function blocks, it is necessary to delete the process from each block.

4



Perform relearning using the new data set and check the score.

5.1.2 Expanding the data set

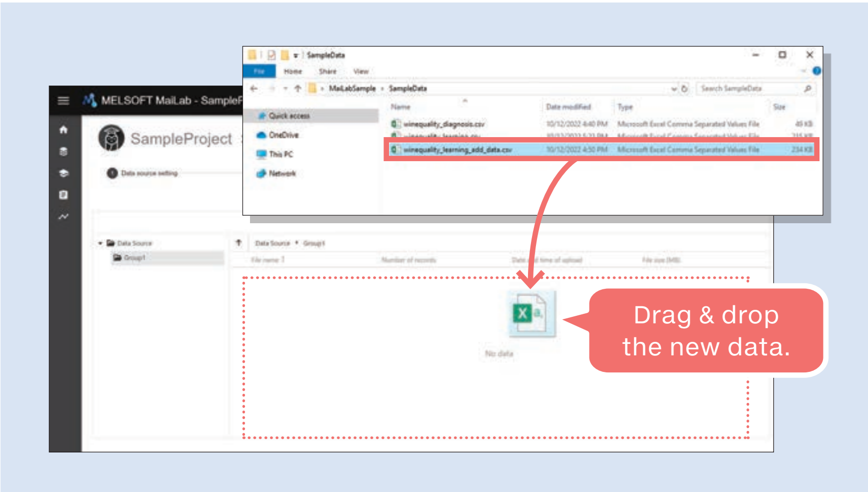


When data that can be used for learning can be prepared separately, the data can be added to expand the data set.
Register the additional data and perform relearning according to the following procedure.
•Add data to the data set.
•Perform relearning in the AI using the new data set.

■ Procedure for adding data to a data set

The procedure for adding data to a data set will be introduced.

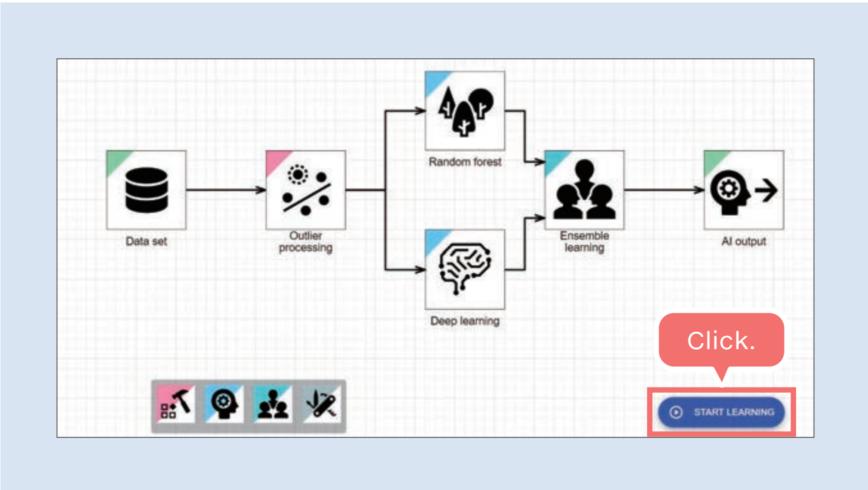
(1) Add data to the data set.



In the Data Source Management screen, upload the text file to be added.
After uploading, create a data set by performing the same operations as in “3.1 Creating the data set”.

* Text files which can be added are limited to text files with the same configuration (header row number, data start row number, number of variables, variable names) as was used when the data set was created.

(2) Perform relearning using the data set to which data was added.



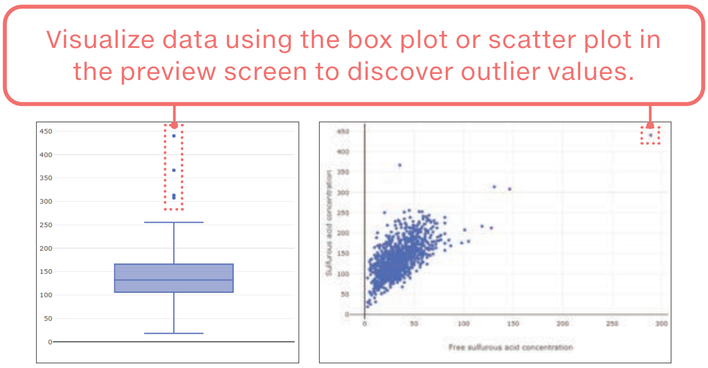
Open the AI in the AI Editor and start learning.

5.2 Checking and adding pre-processing

When predicting quality or signs of abnormality, AI prediction accuracy will be improved if there are data indicating that situation.
Although it would be nice if we could sense every bit of data indicating a situation, but in many cases it is difficult to achieve due to problems such as a basic inability to perform sensing, high costs, etc. In such cases, data can be made more predictable for AI by processing the data available, combining data and extracting features, etc.
In this section, an example of effective pre-processing in the manufacturing industry will be introduced.

5.2.1 Cleansing data

The process of tidying up data with missing values, abnormal values, etc. to make the data clean is called data cleansing.

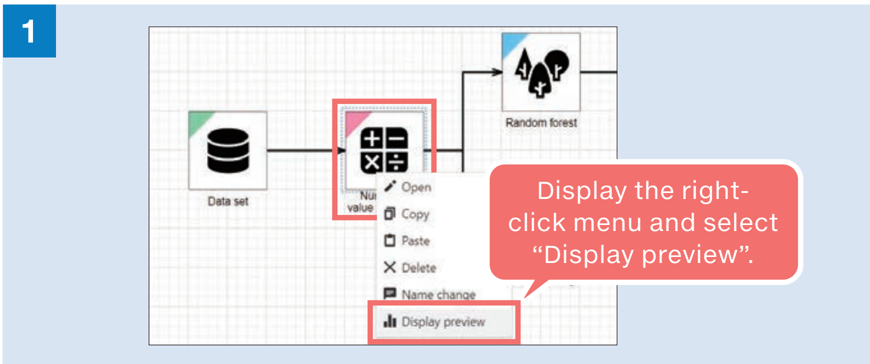


The collected data may contain abnormal values due to missing values or unexpected behavior. In such cases, unexpected AI training results may occur due to the abnormal values. Because of this, check for the presence of abnormal values and consider how to handle them

- Check for the presence of such values in data set preview.
- Perform cleansing using a pre-processing block.

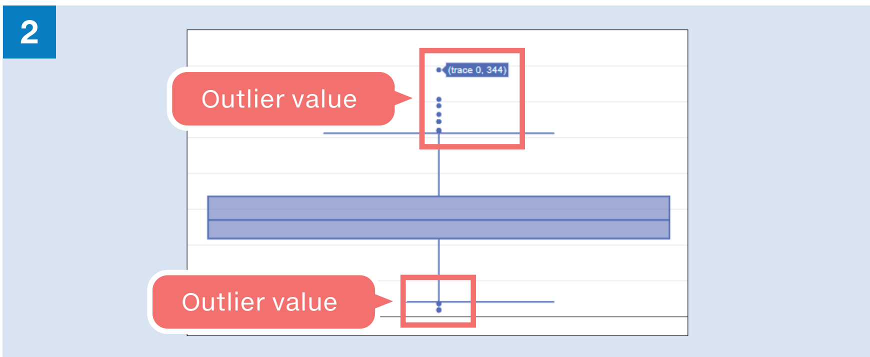
■ Outlier value processing

(1) Preview the block to be input to the analysis method and check for outlier values.



Right-click on the data set block or the preprocessing category block and select "Display preview" from the menu.

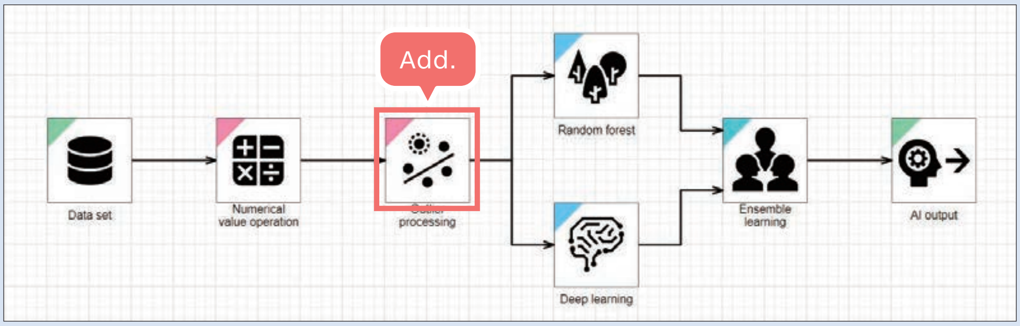
* In the preview for each block, check the output data (block processing results) using a graph.



Check in the graph for the presence of outlier values and their values.

(2) Add the outlier processing block.

3



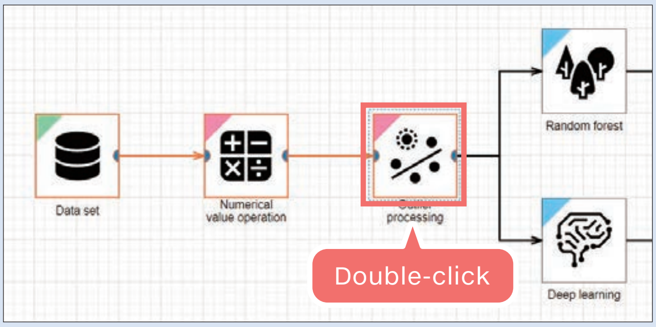
Add the outlier processing block before the Analysis methods category block.

TOPIC

When creating a diagnosis model, outlier values may cause reduced accuracy of the model. Removing the corresponding data or removing the sensor itself from the learning data is effective.
However, in some cases the outlier values are meaningful, such as when the occurrence frequency of outlier values is high before the occurrence of a malfunction, etc. It is necessary for the analyst to judge the final handling of such data based on an understanding of the data's background.

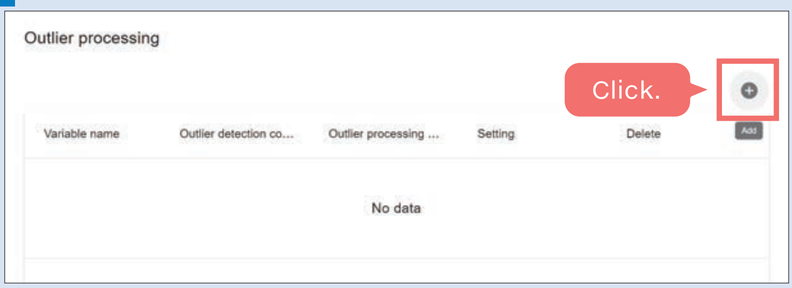
(3) Set the conditions for outlier value processing.

4



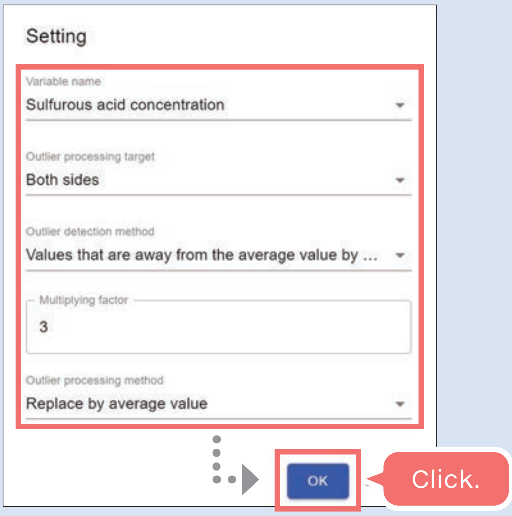
Double-click on the "Outlier processing" block to open the properties.

5



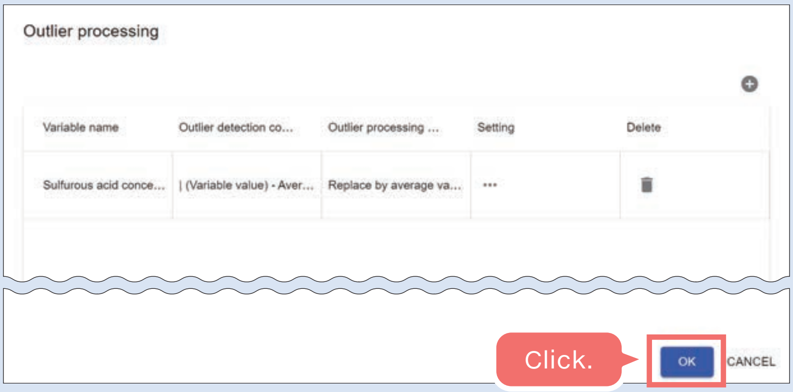
Click the "Add" button.

6



Set the target variables and processing conditions, and click the “OK” button.

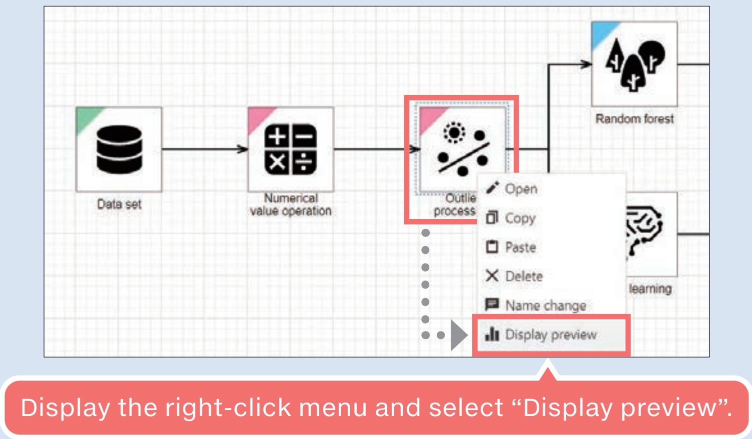
7



Set the outlier value processing conditions and click the “OK” button.

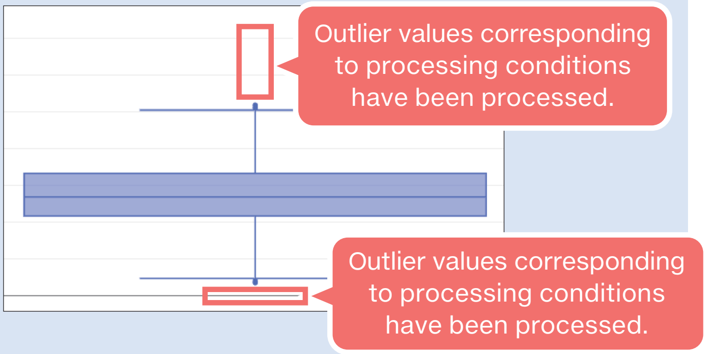
(4) In the block preview, confirm that outlier values have been processed.

8



Right-click on the outlier processing block and select “Display preview” from the menu.

9



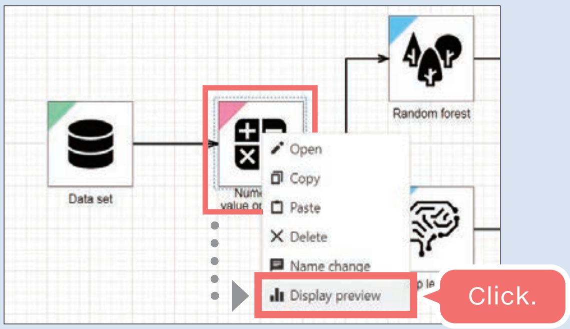
Confirm that outlier values have been processed and eliminated.

* The figure at left shows the processing when the settings in **6** are used to set detection targets on both the upper and lower sides, and values of more than 3 times the average are considered outlier values

■ Missing value processing

(1) Check for outlier values in the preview of the block to be input to the analysis method.

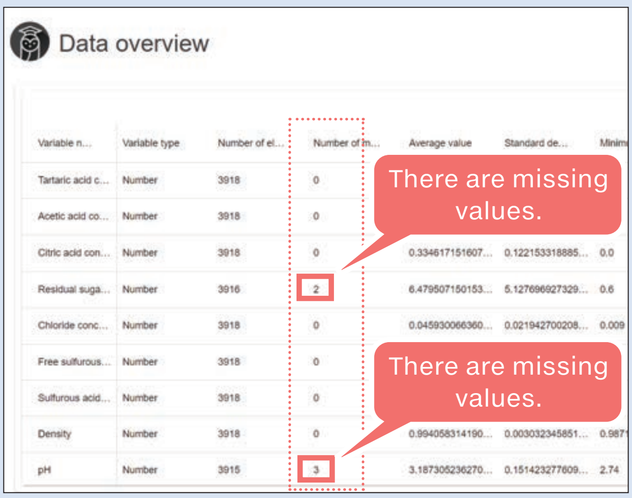
1



Right-click on the data set block or the preprocessing category block and select "Display preview" from the menu.

* In the preview for each block, check the output data (block processing results) using a graph.

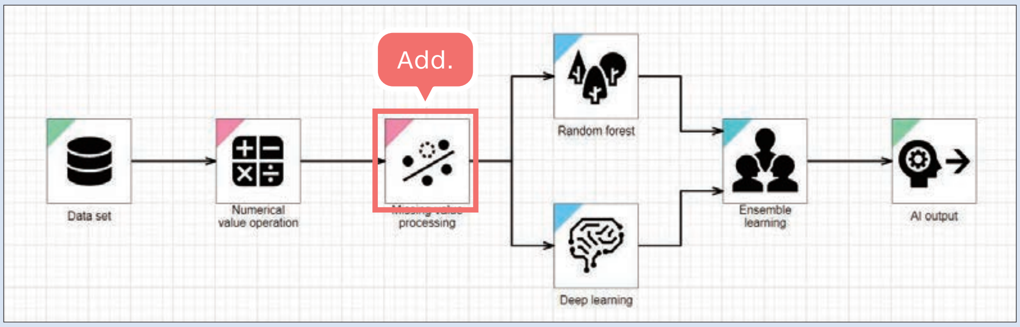
2



Check whether or not there are missing values (missing numbers) in the Data Overview screen.

(2) Add the missing value processing block.

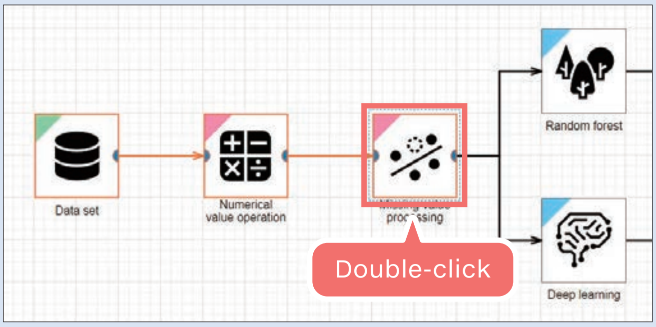
3



Add the missing value processing block before the Analysis methods category block.

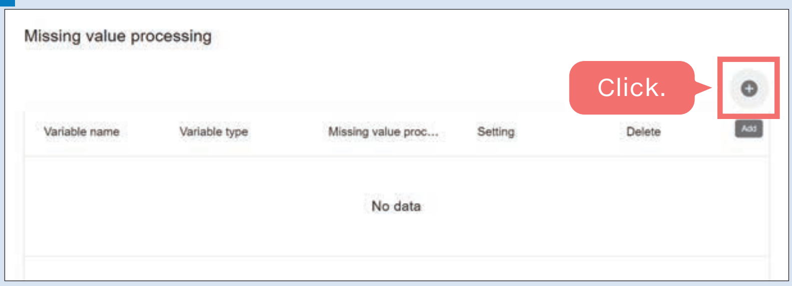
(3) Set the conditions for missing value processing.

4



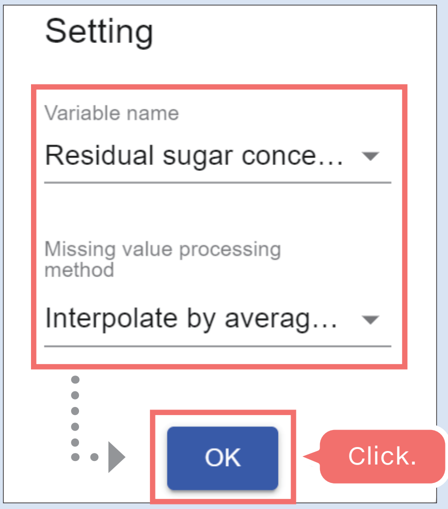
Double-click on the missing value processing block to open the properties.

5



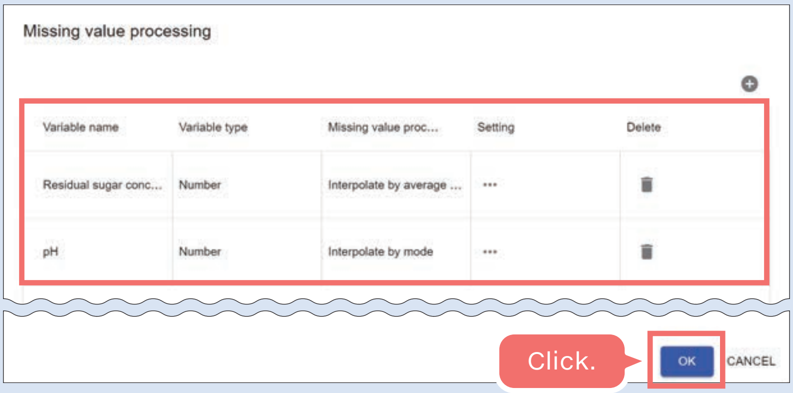
Click the “Add” button.

6



Set the target variables and processing method, and click the “OK” button.

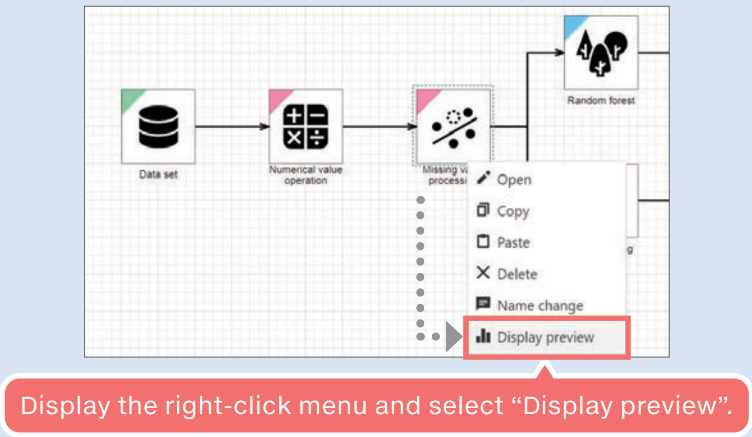
7



Set the processing method for all variables with missing values found in **2**, and click the “OK” button.

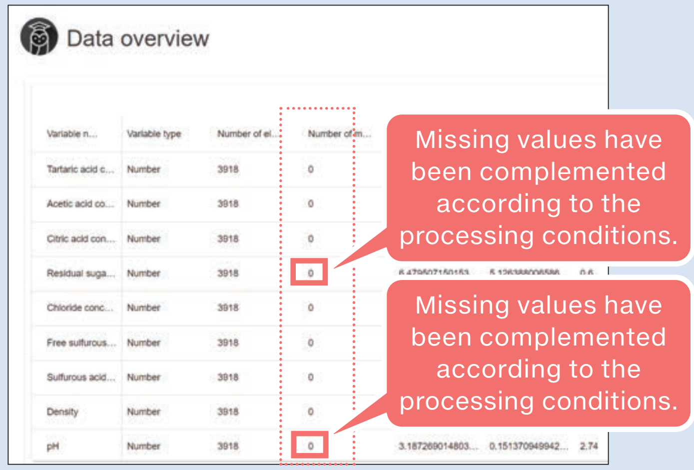
(4) In the block preview, confirm that missing values have been processed.

8



Right-click on the Missing value processing block and select “Display preview” from the menu.

9



In the Data Overview screen, confirm that missing values have been processed and eliminated.

* The figure at left shows the processing when missing values have been complemented according to the settings in **6** using the median value and mode value.

5.2.2 Feature quantity engineering: (1) Joining data to create new data

In some cases, new data related to the objective variables can be created by combining and processing mainly data with high contribution rates to the AI. Preparing data that is more related to the objective variables will improve the AI prediction accuracy.

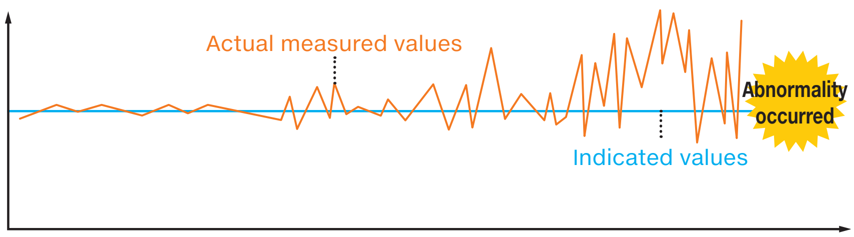
TOPIC

When humans look at data, combining existing data or rereading existing data to make it easier to understand is often performed. For example, in the case of a restaurant, “Sales amount” and “Number of customers” data are used to create new data called “Customer unit value” for evaluation and analysis.
The creation of data that is easy for humans to understand directly results in creating data that is easy for AI to understand, leading to increased accuracy.

In some cases, “Obtaining the difference between data” and “Calculating data proportions” are effective. Here, methods for creating new data using 4 arithmetic operations will be introduced.

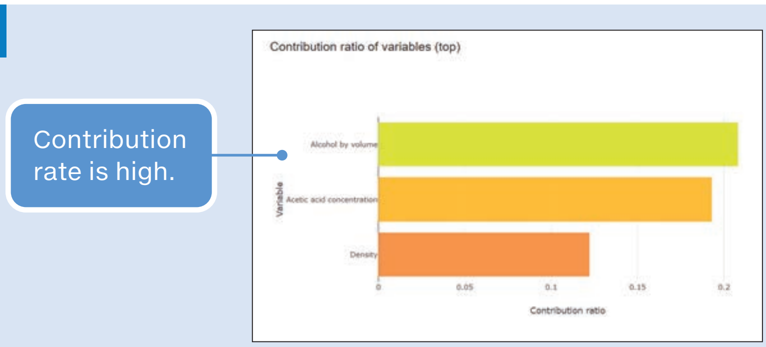
- Check important feature quantities using AI preview.
- Create combined feature values in a Pre-processing block.

In this paragraph, a case in which the contribution rate of actual measured values is high and indicated values are also collected will be used as an example.



(1) Check feature quantities with high contribution rates using AI preview.

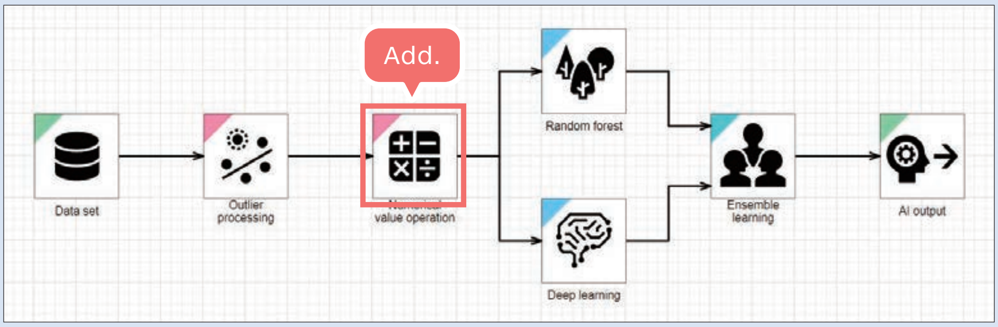
1



In the contribution ratio of variables (top) shown in the AI learning results, check the variables with high contribution rates.

(2) Add the Numerical value operation block.

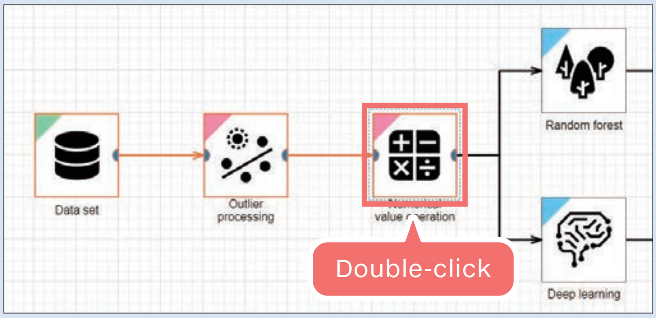
2



Add the Numerical value operation block before the Analysis methods category block.

(3) Set the conditions for numerical value arithmetic operations.

3

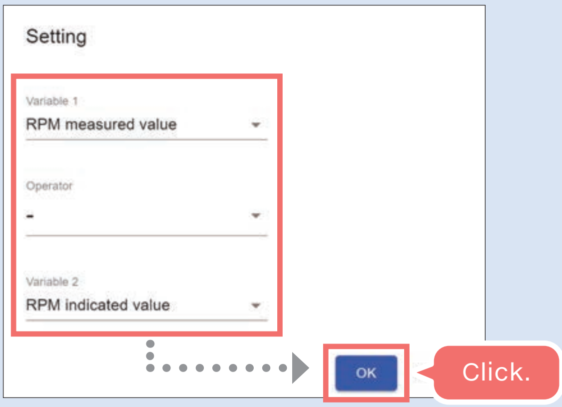


Double-click on the Numerical value operation block to open the properties.

4

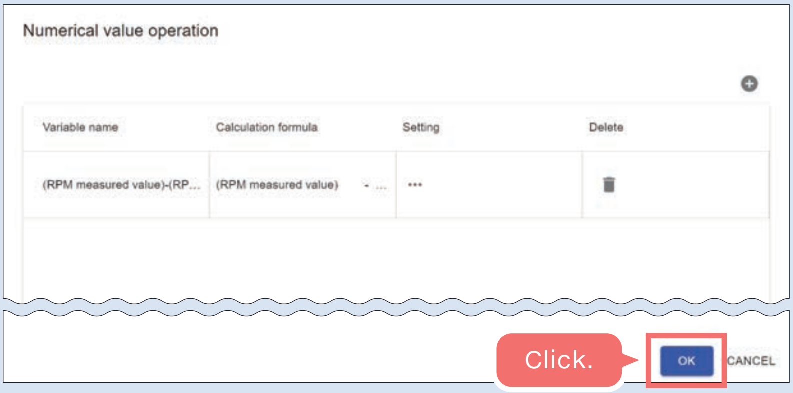
Click the “Add” button.

5



Set the arithmetic formula (Measured value - Indicated value) used by the variable that will become the new variable and click the “OK” button.

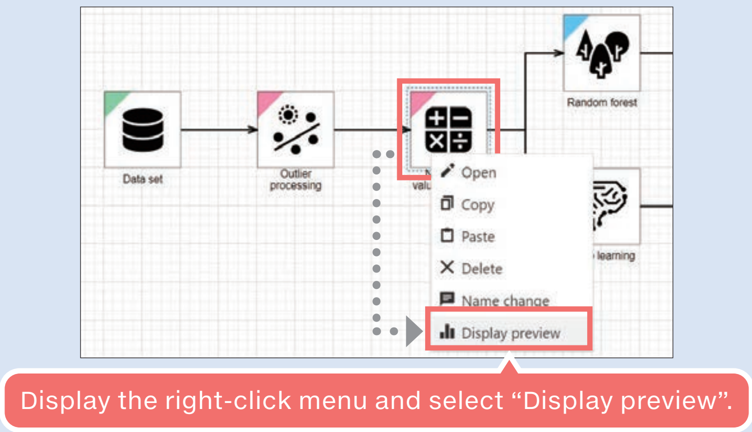
6



Set the arithmetic formula of the new variable and click the “OK” button.

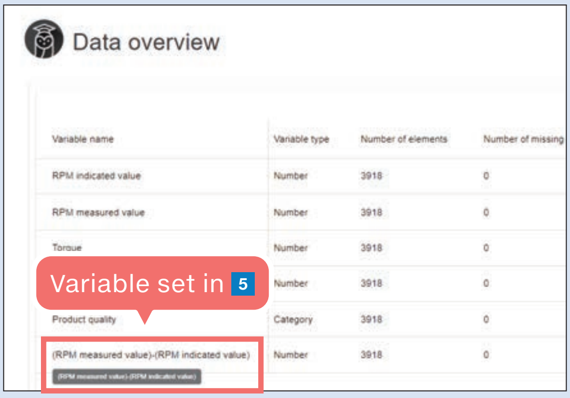
(4) In the block preview, confirm that the new variable has been added.

7



Right-click on the Numerical value operation block and select “Display preview” from the menu.

8



Variable name	Variable type	Number of elements	Number of missing
RPM indicated value	Number	3918	0
RPM measured value	Number	3918	0
Torque	Number	3918	0
	Number	3918	0
Product quality	Category	3918	0
(RPM measured value)-(RPM indicated value)	Number	3918	0

In the Data Overview screen, confirm that the new variable set in 5 has been added.

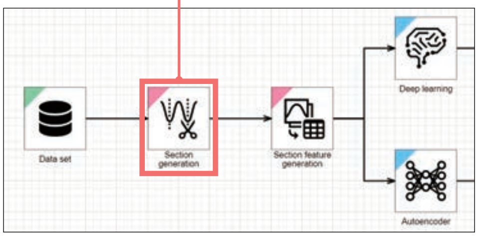
5.2.3 Feature quantity engineering: (2) Taking a specific part of the data and extracting features

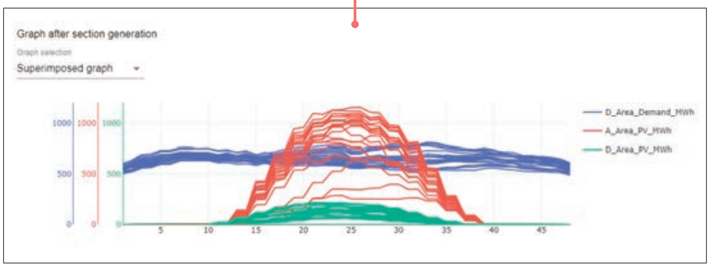
When data set type is “Waveform data set”, cut out specific intervals and create feature quantities.
If the waveform shape differs greatly between intervals or if the length of the data included in the interval expands or contracts greatly, the created feature quantities will vary. Revise the interval creation conditions so that cutting out can be performed at each specific interval.

- Check the extraction status of the “Section generation” block.
- Reset the interval creation conditions.

Open the Section generation block properties screen.

If similar intervals are not being cut out, change the interval creation conditions.

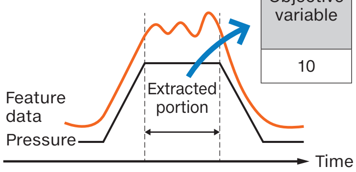




For example, when the same process is repeated such as for a press process, in many cases the shape data will be the same every time.
For data like this, focus on the specific portions of the data where features are likely to be expressed instead of on the entire data. In some cases, creating feature quantities from the data extracted from a specific portion is effective.

Example: When the interval during which pressure is applied is cut out

Cut out the portion where the pressure is constant after rising. The pressure and minimum value of feature data of the cut out portion are calculated. The calculation results will be joined to the objective variable as feature quantities.

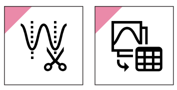


Objective variable
10

→

Objective variable	Minimum value (pressure) of extracted portion	Minimum value (feature data) of extracted portion
10	62.1	100.5

Waveform processing

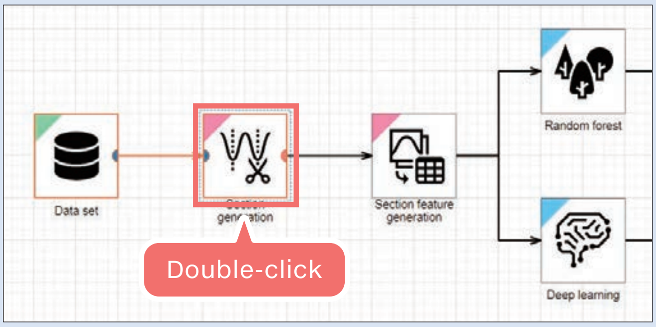


The portion where a value is constant and stable as in the above example is often linked to deterioration of mechanical systems.
However, since there are also cases in which the portion where the value is not constant but is increasing or decreasing is linked to abnormality, it is desirable to create feature quantities from multiple conditions.

The specific operations performed in the above example are introduced from the following page.

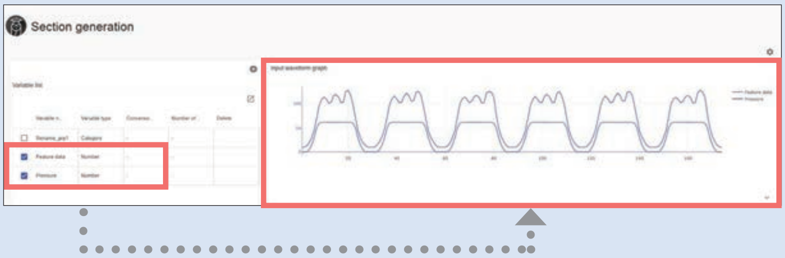
(1) In the Section generation block properties screen, set the conditions for cutting out the waveform.

1



Double-click on the “Section generation” block to open the properties.

2

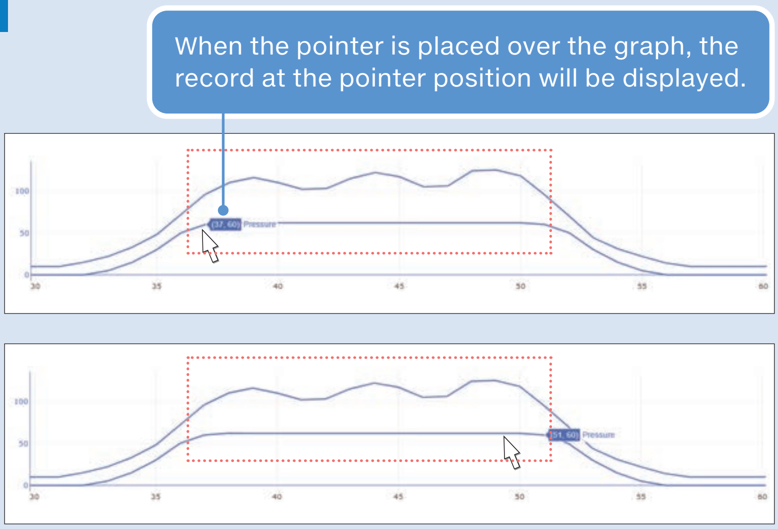


When the variable for checking the interval is selected, the input waveform graph for the selected variable will be shown.
In this example, the feature data and pressure variables are used.

3

Drag on the graph to select 1 cycle of the input waveform.

4



The graph will be enlarged. On the graph, check the value of the range where pressure is stable.

In this example, since it is stable when it exceeds 60, the extraction conditions are set as follows:

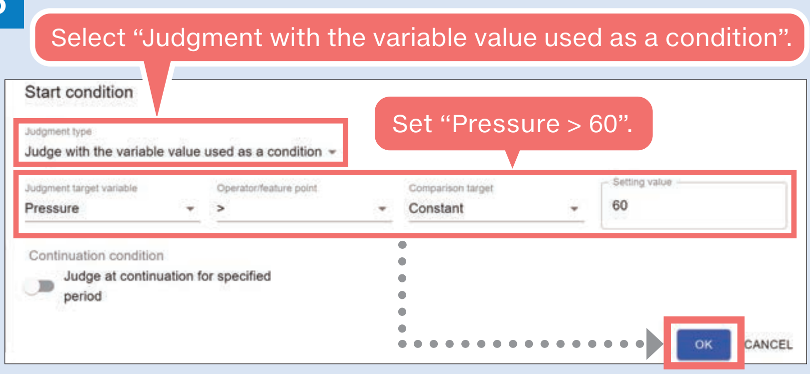
- Start condition: Pressure > 60
- End condition: Pressure > 60

(2) Change the conditions for cutting out the waveform.

5

In the Start conditions, click the “Setting” button.

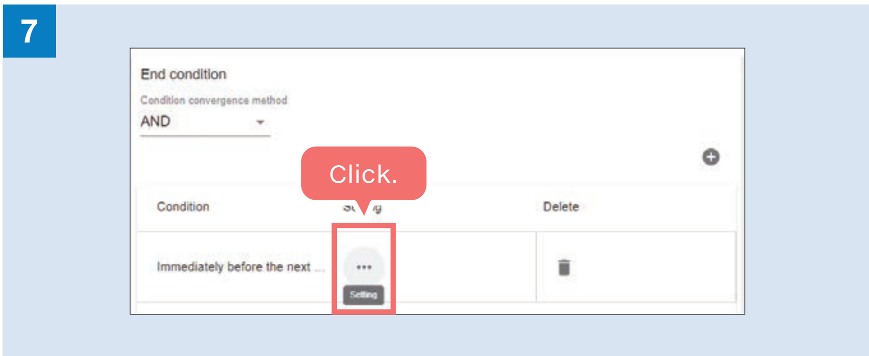
6



Set the start conditions as follows and click the “OK” button.

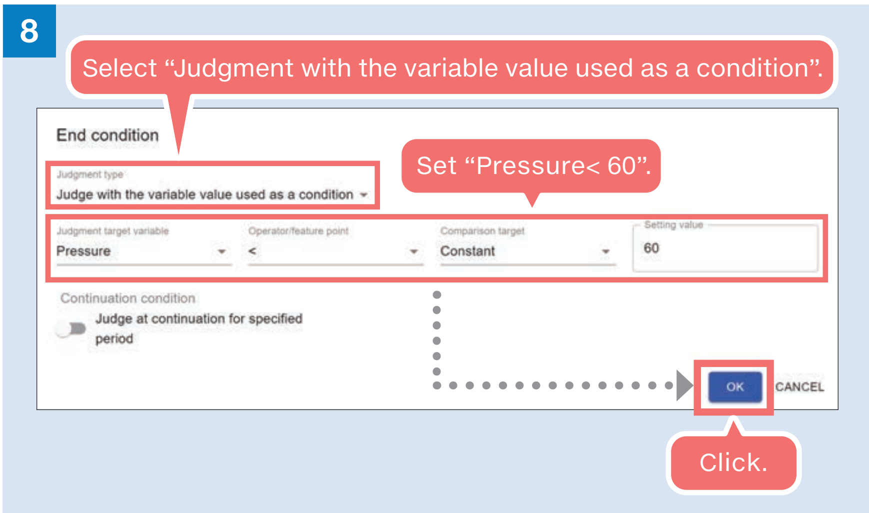
- Judgment type: Judgment with the variable value used as a condition
- Judgment target variable: Pressure
- Operator/feature point: >
- Comparison target: Constant
- Setting value: 60

7



In the End conditions, click the “Setting” button.

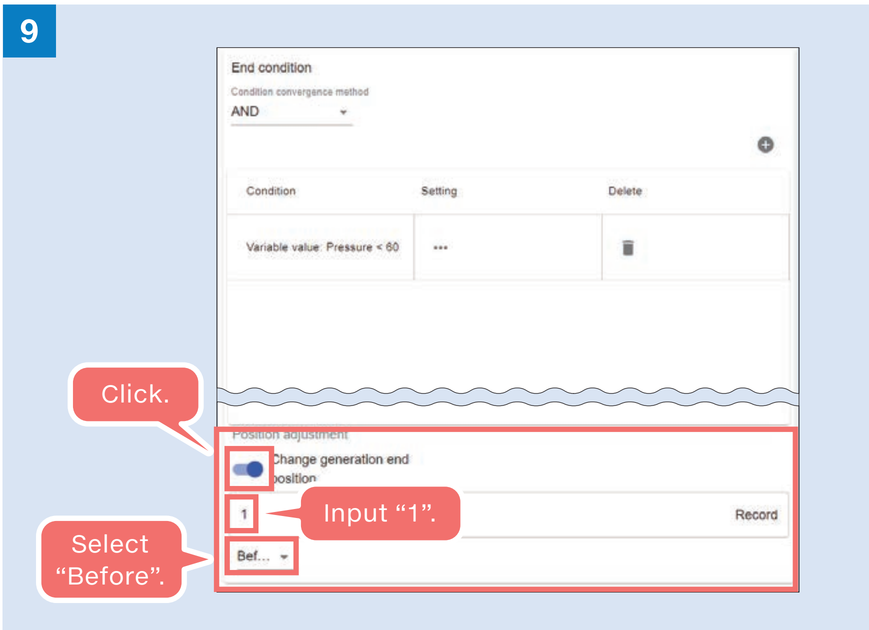
8



Set the end conditions as follows and click the “OK” button.

- Judgment type: Judgment with the variable value as a condition
- Judgment target variable: Pressure
- Operator/feature point: <
- Comparison target: Constant
- Setting value: 60

9



With the condition “Pressure < 60”, the portion until the first record where pressure is below 60 will be the cut out target.
Therefore, fine-tuning will be performed to cut out until 1 record before the condition is met.

(3) Check the cut out interval.

10

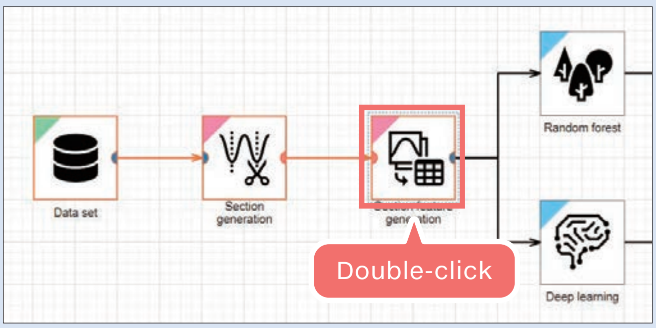
Click the "SECTION GENERATION" button.

11

After interval creation, confirm that the interval which exceeds 60 was cut out, and click the "OK" button.

(4) Perform feature quantification for the cut out interval in the Section feature generation block.

12



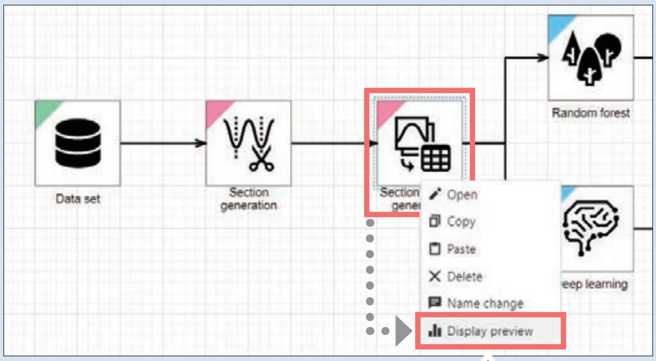
Double-click on the "Section feature generation" block to open the properties.

13

Select the variable feature quantity type and click the "OK" button.
In this example, the feature data and pressure feature quantity is "Minimum value".

(5) Check the results in the Section feature generation block preview.

14

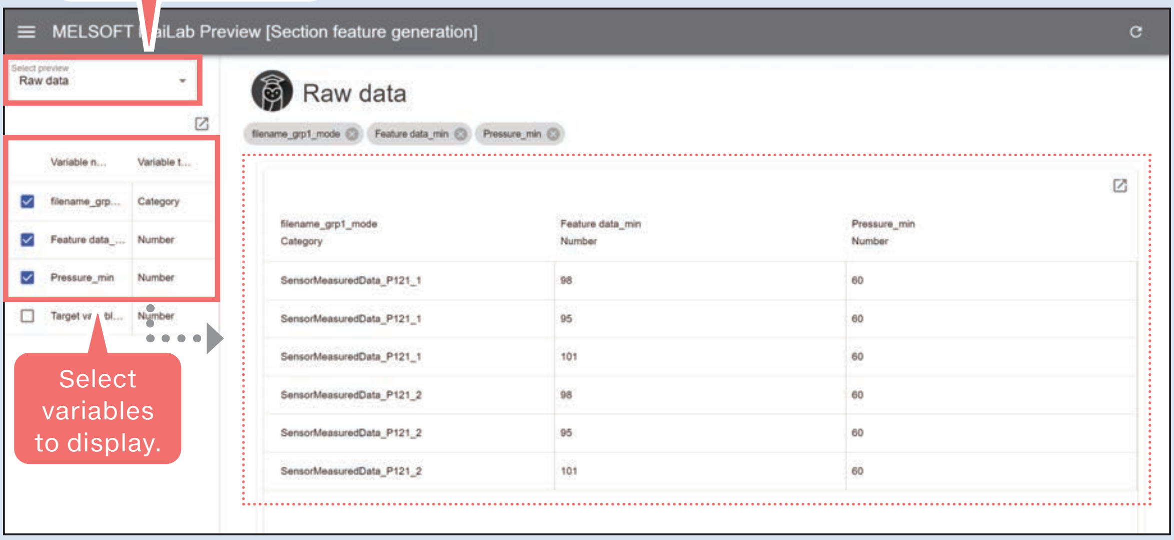


Display the right-click menu and select “Display preview”.

Right-click on the “Section feature generation” block and select “Display preview” from the menu.

15

Select “Raw data”.



filename_grp1_mode Category	Feature data_min Number	Pressure_min Number
SensorMeasuredData_P121_1	98	60
SensorMeasuredData_P121_1	95	60
SensorMeasuredData_P121_1	101	60
SensorMeasuredData_P121_2	98	60
SensorMeasuredData_P121_2	95	60
SensorMeasuredData_P121_2	101	60

Select “Raw data” in Preview select. Data of the selected variables will be displayed. Confirm that they are the values that were expected.

5.3 Checking learning parameters and analysis methods

In the AutoML function, optimization is automatically performed from multiple machine learning methods.
However, if the purpose is clear from the start and there is a method suitable for that purpose, specifying the analysis method is more effective in some cases.
In this section, the characteristics of the analysis methods will be briefly introduced.

5.3.1 Selecting the analysis method

For example, if periodic sensor data with the same waveform shape are always input, guard band that set thresholds along the waveform shape are more suitable than deep learning that extracts features from the waveform.
Search for the suitable method from the characteristics of unsupervised learning method, unsupervised learning method for waveforms, and supervised learning method.

■ Check method

☑ Methods for diagnosing conditions that are different from usual (unsupervised learning method)

MT method (Mahalanobis-Taguchi method)

- A method in which learning is performed using only usual-condition (normal-condition) data and considers “Other than normal is abnormal”.
- It converts numerous variables into a single numerical value called “Mahalanobis distance", and can quantitatively detect signs of abnormality.

Autoencoder

- A neural network that encodes (encodes, compresses) input data and converts it into separate data, and recovers and outputs the original data.
- When abnormal data that were not learned are input, they cannot be recovered correctly. As a result, it can judge and detect whether input data are normal or abnormal.

☑ Methods for diagnosing conditions that are different from usual (unsupervised learning method for waveforms)

Similar waveform recognition

- Learning is performed using the waveform patterns of normal data. A method in which conditions that are different than usual can be detected by judging the similarity between the input data waveform pattern and learned patterns.
- It can detect signs of abnormality that simple upper/lower threshold value judgment cannot detect.

Guard band

- A method in which the normal value range is defined based on standard waveform data, and if the diagnosis target value is outside the normal range, it is judged as abnormal.
- Suitable for applications where high-speed abnormal judgment of periodic waveform pattern data is performed.

Methods for predicting specific quality or defect factors (supervised learning methods)

Deep learning

- A method which uses a multi-layer neural network to automatically extract features from multiple input data and perform diagnosis and learning.
- In MaiLab, a compact network has been designed to operate at relatively high speed with low memory.

Multiple regression

- A method which derives the value of the objective variable based on polynomial equations using multiple explanatory variables.
- Suitable for cases with simple data in which there is a linear relationship between objective variables and explanatory variables. Because of this, high-speed diagnosis and learning can be performed.

Decision trees

A method that improves accuracy by combining multiple decision trees that classify data using Yes/No and showing it in a hierarchal diagram.
[Gradient boosting decision tree]
With a good balance between learning accuracy stability and calculation speed, it exhibits stable learning performance for any type of data.
[Random forest]
Although the learning accuracy may be less than gradient-boosting decision tree in some cases, learning can be performed faster.

k-nearest neighbors algorithm

- A method that performs estimation by judging whether the data to be estimated are similar to learned data.
- The following properties make it is suitable when the number of learning data and variables is small.
  - Estimation accuracy decreases as the number of variables increases.
  - Estimation time increases as the number of learning data increases.

chapter 6

How to create diagnosis models with higher accuracy

One approach to further improve the accuracy of models created based on the steps in previous chapters is to reconsider the data used to create the model. Even if data unrelated to solving the problem (data unrelated to objective variables) are used in the model, improvements in accuracy cannot be expected.
Therefore, examining whether data related to objective variables are being collected appropriately and whether appropriate data are being used in the model may lead to accuracy improvements in some cases.
At that time, perform the examination while refer to the following flow as a reference. The flow concept is explained below.

Approaches to solving issues utilizing data

Definition of issues

Examination of what the target is, understanding the current situation, and what should be solved.

- Extraction of business issues
- Breaking down into on-site issues

Data preparation

Examine what data should be collected and how they should be collected.

- Factor analysis of on-site issues
- Selection/collection of necessary data

Analysis/diagnosis model creation

Examination of validity of data analysis and analysis results

- Check data trends
- Create/verify diagnosis model

Operation

Diagnosis based on analysis results and examination of the results

- Diagnosis system operation
- Verification of results

Definition of issues

Examination of what the target is, understanding the current situation, and what should be solved.

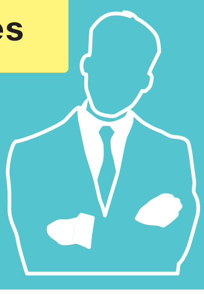
Extraction of business issues

↓

Breaking down into on-site issues

Definition of issues and breaking down into on-site issues

Business issues



This year's goal: Improving quality

For business issues, even broad, slogan-like goals are fine. However, with such broad goals, what each person should do as an individual is unclear. If the goals are not broken down so that what individuals should do can be clearly understood, people won't take any actions. Data analysis is the same. It is necessary to break it down to the level at which the analyst can move their hands, or in other words to break down the issue to be solved into on-site issues.

↓

On-site issues

This year's goal: Improve the defect rate for Defect B on Equipment A by 10%.

The target equipment and defect to be improved are clear, and the numerical goal is also stated. If the goal can be made this specific, people can shift to actual actions toward achieving the goal. Here, what is important is that the numerical goal is stated. The fact that the numerical value is stated has the meaning that the data representing the issue are being measured.

■ What are data representing the issue?

The data that represent the issue in data analysis and are the subject of prediction or estimation are referred to as objective variables.
For example, the definitions of FA site data are as follows:

- Objective variables: When processing accuracy will be predicted, processing accuracy.
- Explanatory variables: Data related to the objective variables (data which seem to be related to processing accuracy, such as ambient temperature, device current, etc.).

■ Supervised learning and unsupervised learning

Analysis methods can be broadly classified into two categories: Supervised learning and unsupervised learning. At FA sites, since the defect occurrence frequency is low and only small amounts of abnormal data can be collected, unsupervised learning may be used, making it difficult to verify the model and ensure accuracy. In some cases, collection of abnormal data is considered to improve accuracy.

Supervised learning	This refers to training a model using data for which explanatory variables and objective variables have been measured. As stated before, "teacher" is synonymous with objective variables.
Unsupervised learning	This refers to training a model using only normal data, or data with only small amounts of abnormality. It is used when the number of data to be detected is extremely small or when there are multiple patterns of detected data and definition is difficult, not performing training with data for which objective variables have not been measured. For example, the MT method of training using only normal data, etc. corresponds to unsupervised learning.

■ Objective variables and explanatory variables

Not only for supervised learning and unsupervised learning, for data analysis it is necessary to consciously collect both objective variables and explanatory variables.
As one approach to further improve model accuracy, it is important that the objective variables be set correctly. If the objective variables are not set correctly, or if variables that do not precisely represent the issue are set as objective variables, the relationship with explanatory variables becomes weak. As a result, high accuracy cannot be expected.
Objective variables and explanatory variables can be thought of as follows:

Objective variables	Since they are data that represent issues, they are equivalent to the results of cause-and-effect relationships (inspection results in the inspection process, etc.)
Explanatory variables	Since they are data related to issues, they are equivalent to causes and factors of cause-and-effect relationships (manufacturing conditions in the manufacturing process, etc.)

From the above, it is not uncommon for objective variables to be obtained from the inspection process and explanatory variables to be obtained from the manufacturing process. Ideally, inspection process data and manufacturing process data would be measured simultaneously. However, in most cases the processes themselves are separate and measurement timings are when time is free.
When objective variables and explanatory variables are measured separately, being able to link objective variables and explanatory variables to each other is important. If linking is vague, the cause-and-effect relationship between objective variables and explanatory variables in the data will be weak, and high accuracy in the model cannot be expected. Therefore, when collecting objective variables and explanatory variables, it is necessary to also be conscious of how they will be linked.

Data preparation

Examine what data should be collected and how they should be collected.

Factor analysis of on-site issues

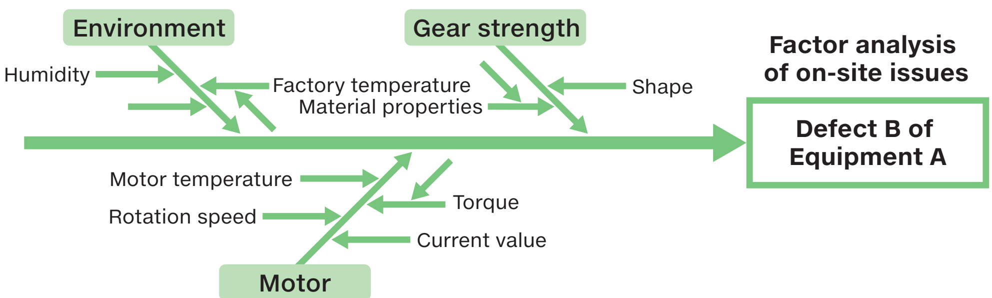
▼

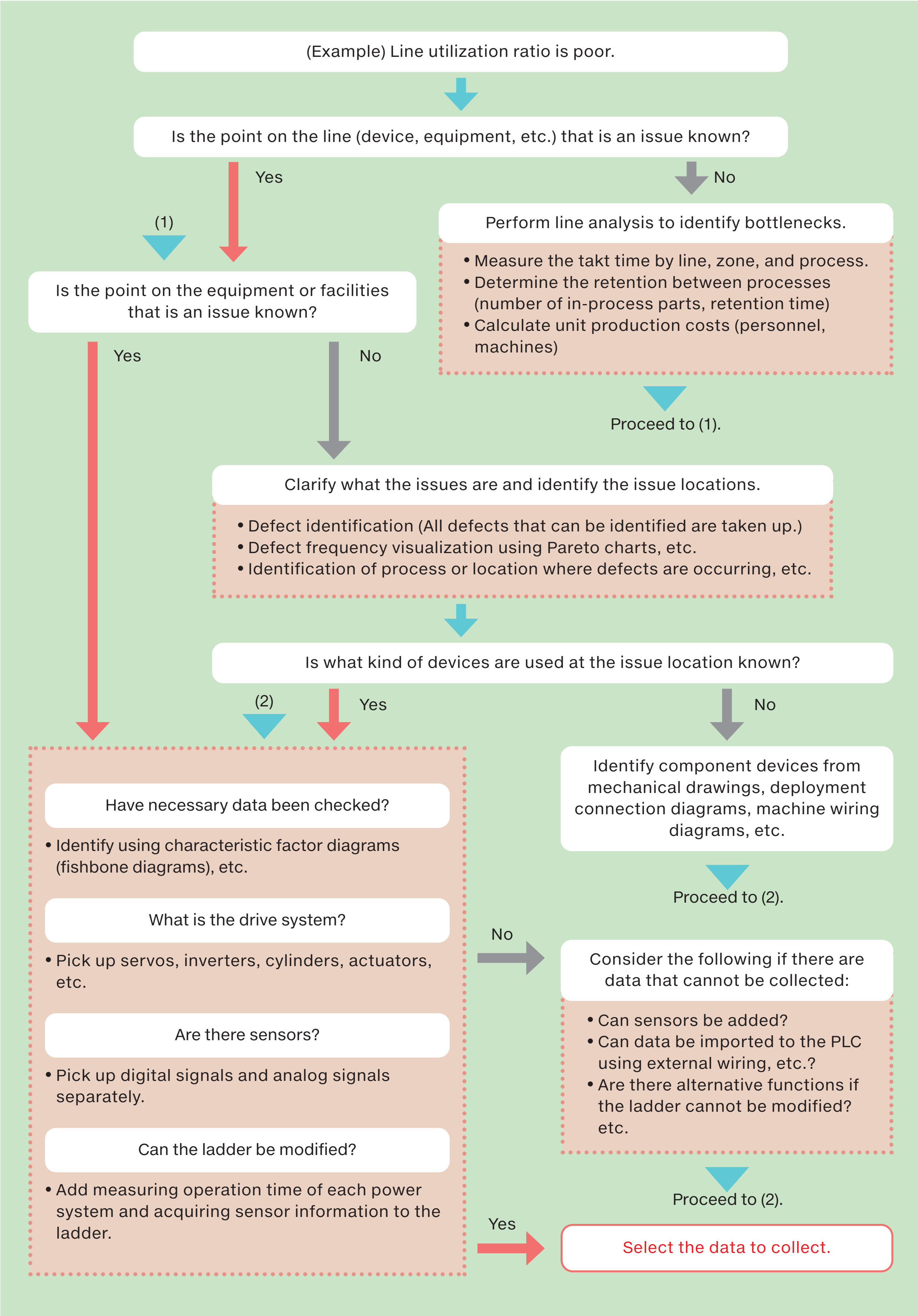
Selection/collection of necessary data

Concept of explanatory variables

■ Performing factor analysis of on-site issues and extracting explanatory variables

When using data analysis to achieve improvements in the defect rate of Equipment A, unrelated Equipment C data would probably not be used. To solve Equipment A issues, it is necessary to use appropriate data obtained from Equipment A.
When solving issues using data analysis, it is necessary to collect data related to the issue in advance. For example, create a characteristic factor diagram as shown below or a flow as shown on the following page. Begin by examining the primary factors of the issue to be solved and check that data representing those primary factors are being obtained. If necessary, install sensors, etc. and collect data.
What is important is not “Since there is data, data analysis should be performed”, but “In order to solve the issue using data analysis, data related to the issue should be selected and collected, and then data analysis should be performed.” At this time, as stated before it is important to link the data representing the issue (objective variables) and data related to the issues (explanatory variables).

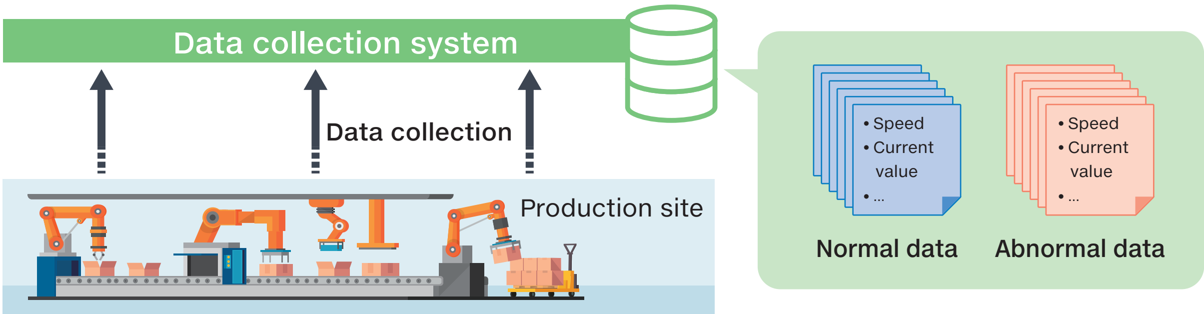




(Example) Line utilization ratio is poor.
Is the point on the line (device, equipment, etc.) that is an issue known?
Yes
No
Perform line analysis to identify bottlenecks.
• Measure the takt time by line, zone, and process.
• Determine the retention between processes (number of in-process parts, retention time)
• Calculate unit production costs (personnel, machines)
Proceed to (1).
(1)
Is the point on the equipment or facilities that is an issue known?
Yes
No
Clarify what the issues are and identify the issue locations.
• Defect identification (All defects that can be identified are taken up.)
• Defect frequency visualization using Pareto charts, etc.
• Identification of process or location where defects are occurring, etc.
Is what kind of devices are used at the issue location known?
(2)
Yes
No
Identify component devices from mechanical drawings, deployment connection diagrams, machine wiring diagrams, etc.
Proceed to (2).
Have necessary data been checked?
• Identify using characteristic factor diagrams (fishbone diagrams), etc.
What is the drive system?
• Pick up servos, inverters, cylinders, actuators, etc.
Are there sensors?
• Pick up digital signals and analog signals separately.
Can the ladder be modified?
• Add measuring operation time of each power system and acquiring sensor information to the ladder.
No
Consider the following if there are data that cannot be collected:
• Can sensors be added?
• Can data be imported to the PLC using external wiring, etc.?
• Are there alternative functions if the ladder cannot be modified? etc.
Proceed to (2).
Yes
Select the data to collect.

Data collection/accumulation

Collect and accumulate data



In order to perform data analysis, it is necessary to accumulate sufficient quantities and types of past data. Accumulate data in the constructed data collection system and use it for offline analysis. In performing offline analysis, the amount of data is involved as follows:

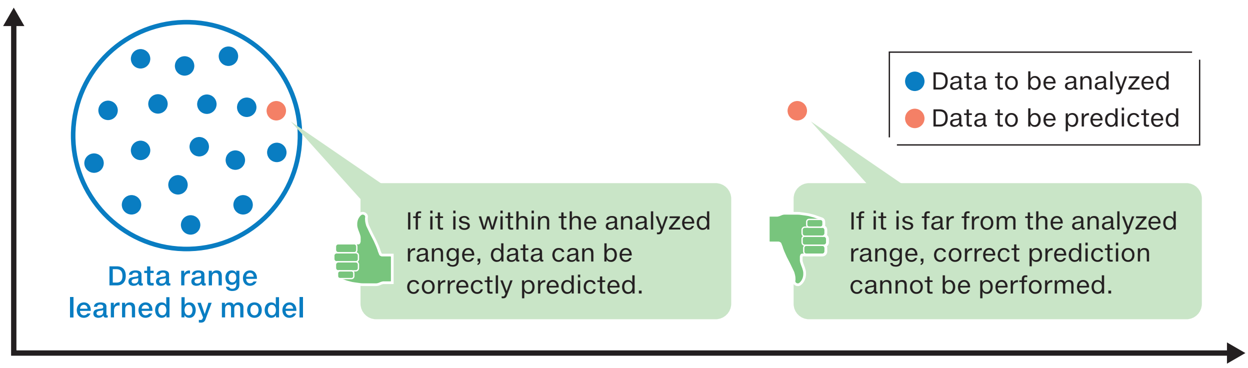
☑ **Having defective product data or abnormal data will enable more effective offline analysis.**
☑ **In general, having larger amounts of accumulated data will enable more effective offline analysis.**
☑ **The amount of accumulated data required will vary according to the issue and data characteristics.**

Based on the above, model accuracy may be improved by doing the following:

☑ **Add more data, revise and add to the types of data, revise the measurement timing of data for analysis.**
☑ **In particular, add defective product data and abnormal data.**

Data analysis is not a cure-all.

Data analysis is based on what was stated before: “In order to solve the issue using data analysis, data related to the issue should be selected and collected, and then data analysis should be performed.”
However, data analysis is not a cure-all.



For example, when predicting a certain data, if the data to be predicted is within the range of data learned by the model in advance as shown in the above diagram, data can be predicted with high accuracy. However, for data which is far away from the analyzed data range, high prediction accuracy cannot be expected.
It is necessary to note that in data analysis, high-accuracy prediction cannot be performed for data which is not similar to existing known data. This means that if the data that the model learns does not include data for predicting defective products or abnormalities, it will be difficult to make predictions with high accuracy. (Particularly for supervised learning, it is important to collect data for predicted defective products and abnormalities.)
For example, it is difficult to use Defect B data to create a model to predict Defect D. To create a model to predict Defect D, it is necessary to also collect Defect D data and use it for learning.

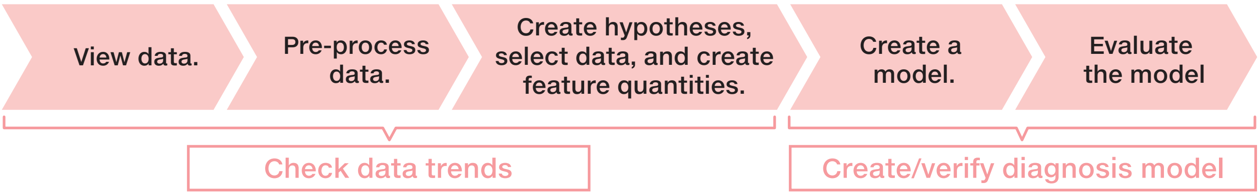
Analysis/diagnosis model creation

Examination of validity of data analysis and analysis results

Check data trends

Create/verify diagnosis model

Offline analysis



The 5 steps above are the general flow of offline analysis, in which collected data are analyzed and a model for solving issues is created. The above steps are not necessarily followed in one direction from left to right. In many cases there is a lot of going back and forth between steps.
MaiLab automatically performs each step of “Create/verify diagnosis model.” in AutoML. However, improvements can be made in each step of “Check data trends.” MaiLab automatically performs each step in AutoML. However in each step of “Check data trends.”, the accuracy of the model may be improved in light of domain knowledge and physical background by doing the following:

- ☑ **Manually remove variables that should be removed in advance to avoid discarding necessary variables.**
- ☑ **Create and add new “variables” and “feature quantities” based on data trends, characteristics, and hypotheses identified in advance.**
- ☑ **Perform pre-processing of data such as removing noise and outlier values, performing missing-value processing, etc.**

■ Think about the data

Another approach is to check the correctness of the data. For example, for time-series data included in the previously stated “Create hypotheses, select data, and create feature quantities”, review the concepts of separators/windows, binning of numerical data. In addition, the following viewpoints can be raised, and the "quality" of the obtained data greatly affects the accuracy of the model.

- ☑ **Are the various sensors used for measurement properly calibrated?**
- ☑ **Are the various sensors used for measurement appropriately measuring data related to the issues? (Are the measurement method, means, and installation positions appropriate?)**
- ☑ **Is it necessary to apply compensation to the obtained data?**
- ☑ **Are the clocks among multiple facilities or processes synchronized and correct? (Important when date/time data are used for linking data)**
- ☑ **Are data that have been converted to Log, etc. being used as is, or is it necessary to perform Log conversion, etc. on data?**
- ☑ **Are data measurement intervals appropriate? (Should measurements be taken at finer intervals or are the intervals too fine?)**
- ☑ **Are the measurement intervals for multiple time-series data aligned?**
- ☑ **Are there any flags or other indicators that equipment or processes are in operation, and if so, can they be obtained and utilized as needed to make better divisions?**
- ☑ **Is it necessary to compensate for differences between equipment or devices, and if so, is compensation being correctly applied?**
- ☑ **Is information obtained from sources other than data being converted into variables? (for example, information obtained from serial ID naming rules)**
- ☑ **Are there any useful information recorded on paper that is not stored in a database or file, and is it being obtained?**
- ☑ **Do trends change due to presence/absence of equipment maintenance, material changes, etc., and if there are changes, are they being taken into consideration during analysis?**
- ☑ **Are the results analyzed without preconceptions? (However, the obtained analysis results should be interpreted and understood in light of domain knowledge and physical background.)**

Operation

Diagnosis based on analysis results and examination of the results

Diagnosis system operation

Verification of results

Operation

When the model (analysis rules) created based on analysis results is actually operated at the production site, the following should be considered.

■ The analysis results and model should be understood and should be explained correctly and in an easy-to-understand way to responsible parties.

In order to apply the created model to the production site, it is necessary for the person in charge of the site and other people involved to understand the model.
When explaining, think carefully about the analysis results and the model and be sure that you understand them so that you can explain them correctly and in an easy-to-understand way.

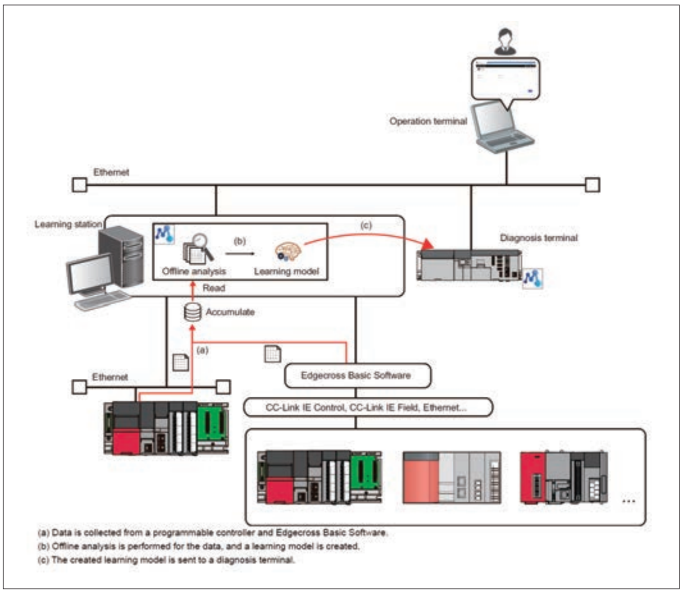
■ Diagnosis system construction

Construct a system for performing real-time diagnosis and consider actions to be taken when symptoms are detected.
Example: When a detection occurs, light a patrol lamp at the production site to notify everyone immediately

■ Diagnosis system operation

Introduce a diagnosis system to the production site and operate it.
Depending on the importance of the diagnosis target and diagnosis accuracy, start operation according to on-site operation policies, such as by first verifying the system on a prototype line.
Example: Start operating real-time diagnostics on a trial basis on one of multiple devices.

Diagnosis system construction example



(a) Data is collected from a programmable controller and Edgecross Basic Software.
(b) Offline analysis is performed for the data, and a learning model is created.
(c) The created learning model is sent to a diagnosis terminal.

■ Verification of results

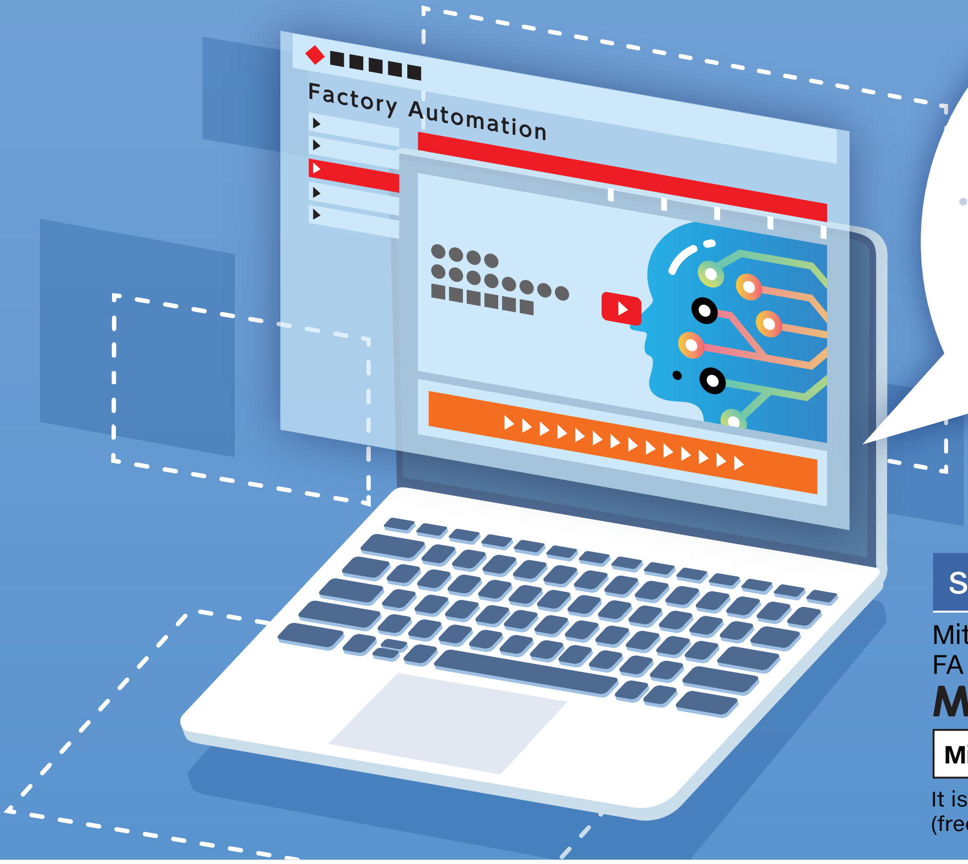
Evaluate the degree of operation effect on on-site issues.
In addition, since data trends may change due to long-term operation of production facilities, changes in the environment or materials, etc., periodic review is necessary.
Example: The defect rate for Defect B on Equipment A was improved by 11%.

Mitsubishi Electric Data Science Tool
Data Analysis Textbook

MELSOFT MaiLab Trial Edition

Available free of charge for 90 days

You can download the MELSOFT MaiLab Trial Edition and the sample data from Mitsubishi Electric FA Global Website to thoroughly check the operating procedures. Please give it a try!



Download the MELSOFT MaiLab Trial Edition!

The sample data* is also available.

*Used in Chapter 2 of this document.

Scan this QR code to download.

Mitsubishi Electric
FA Global Website
MELSOFT MaiLab

Mitsubishi Electric MaiLab

It is necessary to register as an "FA member" (free of charge) to download items.

L(NA)08871ENG-A 2301(MEE)

Printed Jan. 2023