

# WiseProphet

WiseProphet v1.0

**WISEiTECH**

Copyrights 2019 WISEiTECH co., Ltd. All rights reserved.

## **Copyright Notice**

Copyrights 2019 WISEiTECH co., Ltd. All rights reserved.

## **Restricted Rights Legend**

All WiseProphet Software and documents are protected by copyright laws and the Protection Act of Computer Programs, and international convention. WiseProphet software and documents are made available under the terms of the WISEiTECH License Agreement and may only be used or copied in accordance with the terms of this agreement. No part of this document may be transmitted, copied, deployed, or reproduced in any form or by any means, electronic, mechanical, or optical, without the prior written consent of WISEiTECH Co., Ltd.

## **Trademarks**

WiseProphet is a registered trademark of WISEiTECH Co., Ltd. Other products, titles or services may be registered trademarks of their respective companies.

The Information about the user manual.

The title of a guide: the user manual on Wise Prophet

Initial issue date: Apr 26th, 2019

Software version: Wise Prophet v1.0

Guide version: 1.0.1

## Contents

About manual.....	1
Product information and recommended specifications .....	4
The basic concept of machine learning .....	5
Definition of Terms .....	10
Chapter 1. Introduction of WiseProphet.....	12
1.1. Summary.....	12
1.2. Features.....	12
1.2.1. Ease of use .....	12
1.2.2. Support to every data .....	13
1.2.3. Support with various algorithm .....	13
1.2.4. Feature Extraction.....	13
1.3. Main Function .....	14
Chapter 2. Login.....	15
2.1. Screen Layout.....	15
2.2. Signing up.....	16
Chapter 3. Model Training.....	17
3.1. Screen Layout.....	17
3.2. Data Input.....	18
3.2.1. Create File .....	18
3.2.2. Local File.....	19
3.2.3. Database.....	19
3.3. Data Discovery .....	20
3.3.1. Exploratory Data Analysis .....	21
3.3.2. Data Scaling.....	22
3.3.3. Data Distribution.....	23
3.3.4. To change the Data Type .....	24
3.3.5. To use variables .....	25

3.3.6. To set the target.....	25
3.4. Feature Selection .....	27
3.4.1. The Effect by Feature.....	27
3.4.2. Correlation analysis.....	28
3.5. Model Selection .....	29
3.5.1. Clustering.....	30
3.5.2. Classification .....	31
3.5.3. Regression .....	32
3.5.4. Parameter Optimization.....	34
3.6. Setting the ratio of Validation Data.....	34
3.6.1. Train-Validation Holdout.....	34
3.6.2. Cross-Validation .....	35
3.7. Model execution .....	36
3.7.1. Clustering.....	36
3.7.2. Classification .....	36
3.7.3. Regression .....	38
3.7.4. Model Logs .....	39
Chapter 4. Model Management.....	40
4.1. Screen Layout.....	40
4.2. Model Management.....	40
4.3. Model Modification .....	42
Chapter 5. Model Operation.....	45
5.1. Screen Layout.....	45
5.2. Settings and Functions.....	45
5.2.1. Add Schedule .....	45
5.2.2. Run and verify the schedules.....	46
Chapter 6. Monitoring.....	47
6.1. Screen layout.....	47
6.2. the information of user and model .....	47

Chapter 7. Setting.....	49
7.1. User Setting.....	49
7.2. Connection Setting .....	50
Chapter 8. FAQ .....	51

# About manual

## For whom are the manual

You can learn with this manual how to classify and predict data after preprocessing it and building model with **WiseProphet**. The algorithm WiseProphet provides is based on the one offer the open sources: Scikit-learn, Keras, Tensorflow.

## Precondition for the manual

This manual describes the necessary process for generating model and predicting data with WiseProphet. So, to understand our descriptions well, you should know the following in advance.

- reference to “Understanding machine learning – Basic concept of machine learning and Definition of terms”

## Constraint for the manual

This manual does not include all the details for applying to business practice with WiseProphet. So, if you want to know about management like installation and preferences, you should refer to each product guide.

---

Reference

You can refer to “management manual” to learn how to install and configure WiseProphet.

---

# Organization of manual

WiseProphet manual for user has 6 chapters and Appendix.

Each chapter deals with the following.

- Before starting:

## **Product information and recommended specifications**

Guiding product information and recommended hardware specifications for WiseProphet.

## **Basic concept and terms about machine learning**

Outlining basic concept about machine learning before applying WiseProphet.

- **Chapter 1. The introduction to WiseProphet**

To describe main features of WiseProphet.

- **Chapter 2. Login**

To outline the login screen of WiseProphet.

- **Chapter 3. Model training**

To explain the process for training models.

- **Chapter 4. Model management**

To describe how to predict new data and handle with models with WiseProphet.

- **Chapter 5. Monitoring**

To outline the monitoring screen of WiseProphet,

- **Chapter 6. FAQ**

To provide the FAQ about WiseProphet.

## Our contact information

WISEiTECH Co., Ltd.

253,Pangyo-ro, Bundang-gu,

Seongnam-si, Gyeonggi-do

South Korea

Tel: + 82-02-6246-1400

Fax: + 82-02-6246-1415

Email: [contact@wise.co.kr](mailto:contact@wise.co.kr)

Web (Korean): <http://www.wise.co.kr/>

Technical support: <http://www.wise.co.kr/>

# Product information and recommended specifications

## 1. Composition and installation

This product consists of a CD, an executable, and the user manual. It is our trained representative to be in charge of installing the product. And you can receive a CD, an executable, and the user manual from him.

## 2. Warranty period

We provide maintenance for free within a year of the purchase. After the warranty period, the service will be paid.

## 3. Hardware specification

Hardware specification depends on the volume of data you want to analyze. So, we made it out on 100GB basis. GPU is not necessary, but if you want to maximize performance, you need it.

CPU	8Core or above
Memory	128GB or more
HDD	1TB or more (the number of users * 1TB)
GPU	RTX 2080 Ti or above

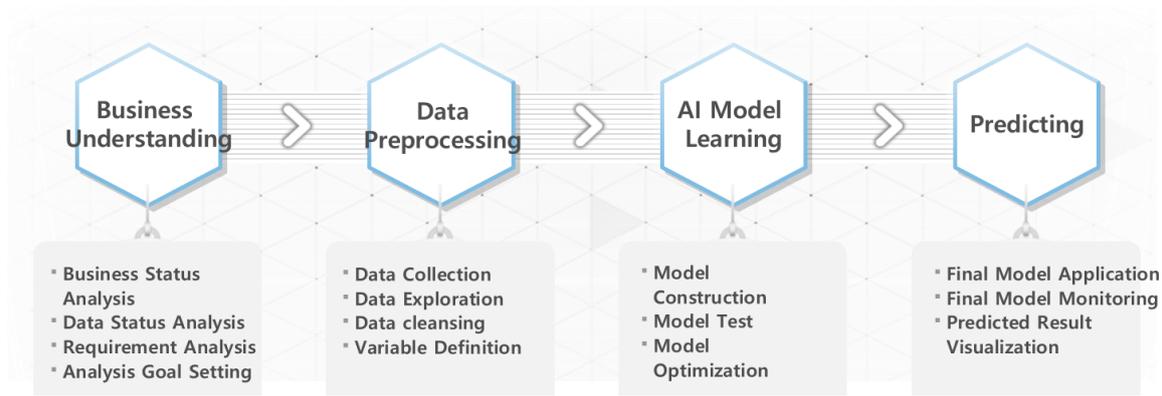
## 4. Supported OS

- CentOS Linux
- OS X
- Redhat Enterprise Linux
- Solaris
- SUSE Linux
- Ubuntu Linux
- Microsoft Windows

# The basic concept of machine learning

## 1. Machine Learning

Machine learning means that a program trains data and finds out its pattern by automatizing the model generation for data analysis. It would minimize human intervention and make decision through training.



[Graphic] Process of Machine Learning

## 2. Data Preprocessing

Data preprocessing means to select features of data, cleanse and converse data before generating the training model. It is an important task you should perform to construct excellent predictive model. The process of data preprocessing is as follows.

Classification	Content
1) Data Exploration	<ul style="list-style-type: none"><li>- The phase to explore data before analyzing it.</li><li>- To grasp data distribution and basic statistics: average, dispersion.</li><li>- To check data state: missing values or outlier.</li><li>- To identify data type: numeric, category, text etc.</li></ul>
2) Data Cleansing	<ul style="list-style-type: none"><li>- To fill the missing values or remove noisy data and outliers.</li></ul>

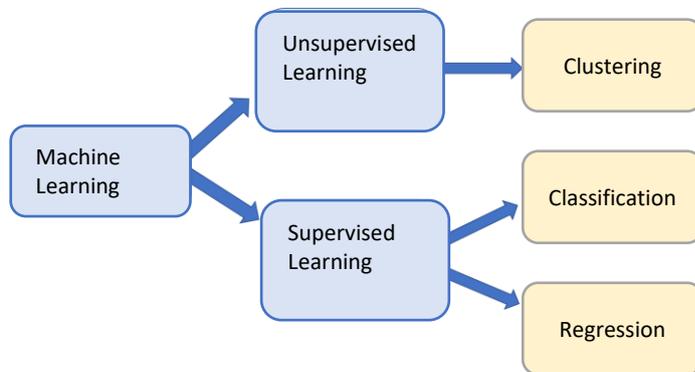
---

3) Data Conversion	- Data Scale: A task to alter range of different data to specified range. Ex) To alter the range of height/weight to range from 0 to 1. - Changing Categorical Data: To change the categorical data to numerical data(0, 1)
4) Feature Selecting	- To select the best feature for sorting and predicting data. - To add the necessary features or eliminate the unnecessary ones.
5) Feature Extraction	- To produce the new features from combination of ones.

---

### 3. Learning

The learning way can be divided into Supervised Learning and Unsupervised Learning.



[Graphic] How to classify Machine Learning Model

#### 1) Supervised Learning

- Supervised Learning is the way to learn under the condition of having the right answer.
- The right answer for each data is called 'label', a group of labeled data is called 'Training Set'.
- After generating model based on learning data, it can be predicted that what feature of data belongs to which of right answer: labels, values

- The types of Supervised Learning

Supervised Learning is classified into Classification and Regression according to the type.

- Classification

- : To divide by category. Ex) spam/normal mail classification

- Typical algorithm: Support vector machine, Decision tree etc.

- Regression

- : To predict continuous values. Ex) Prediction of House prices, Stock and Demand.

- Typical algorithm: Linear regression analysis, Logistical regression etc.

#### 2) Unsupervised Learning

Unsupervised Learning is the way to learn without the right answer: label. It is used for figuring out hidden patterns or groups with only features of data.

Typical algorithm: k-means clustering, PCA.

Ex) google news service: classifying news by similar topic.

Word clustering: putting similar words together.

## 4. Algorithm Selecting

The learning way depends on user's purpose, algorithm is selected according to it: classification, regression. Prediction accuracy, training time and usability should be considered in selecting the algorithm.

Typical algorithm by training way is as follows.

Learning Type	Algorithm
Classification	Random forest
	Decision tree
	Support vector machine
Regression	Linear regression analysis
	Elastic net
	k-means

## 5. Model Evaluation

Model Evaluation is the phase to conduct efficiency test of learning model and to predict data with model. To evaluate model is divided into Training Set and Test Set. It is the point that the Test Set shouldn't be used in the algorithm selection and model learning process.

How to evaluate model includes Splitting Training/Test Set and Cross Validation.

- 1) Training/Test Set Splitting: It is the most common method to divide into Training Set and Test Set. Usually at a ratio of 80:20.
- 2) Cross Validation: Not once division into Training Set and Test Set, but several times. One of them would be Training Set, the others Test Set. After many times, the most accurate model would be selected.

Model Evaluation Metrics are classified into Classification, Regression and Clustering by learning type.

Learning Type	Evaluation Metric	Description
---------------	-------------------	-------------

Classification	Accuracy	A percentage of how much the Forecast Value and the Actual Value match.
	Precision	The rate at which the Actual Value occurs among the Forecast Value.
	Recall	The rate of how much accurate the value separated from model is.
Regression	RMSE	The square root of the Forecast Value minus the Actual Value.
	MSE	The absolute value of the forecast value minus the actual value.
	MAPE	A percentage of the forecast value and the actual value.

# Definition of Terms

The terms about machine learning in this manual are written in Korean referring to Google Machine Learning Glossary(<https://developers.google.com/machine-learning/glossary/>).

Name(Korean)	Name(English)	Meaning
학습 데이터	Training Set	Data set for model training.
평가 데이터	Test Set	Data set for evaluating model at an early stage.
특징	Feature	Input variable to be used for performing forecast.
특징 선택	Feature Selection	Selecting the input variable for model.
특징 추출	Feature Engineering (=Feature Extraction)	Figuring out what feature is useful for training model and changing log files and source data into each characteristic.
특징 별 영향	Feature Importance	The relative importance of each variable for a target.
데이터 스케일	Data Scale	To make the range of the characteristic value match the other one of Data Set. For example, all the floating point of Data Set can be set in the 0-1 range. If the range of certain feature is in the 0-500 range, you can regulate it by dividing each value by 500.
1 분위	Q1	The lowest 25% of data Which means 25% of data is less than or equal to the Value.
3 분위	Q3	The highest 75% of data Which means 75% of data is less than or equal to the Value.
유일값	Unique Value	The only value after data deduplication.
결측값	Missing Value	When data is empty.
목표 변수	Target Value	The target to predict.
정규화	Normalization	To convert the range of the actual value into the range of the standard value: usually -1~+1 or 0~1. For example, if the original range of certain feature is in the 800~6,000 range, you can normalize the range to -1~+1 which goes through subtraction and division
매개변수	Parameter	A variable of model training by itself. Value of weight parameter trains depending on repeating training.
범주형 데이터	Categorical Data	Feature with discontinuous set of possible values Ex) gender(male, female)

원-핫 인코딩	One-hot Encoding	To convert categorical data into numeric one: 0, 1
모델 학습	Model Training	To decide the best model
L1 정규화	L1 regularization	The regularization type to give a penalty to a weighting in proportion to the sum of the weighting's absolute values.
초평면	Hyperplane	The boundary to divide a space into two spaces.
분류 행렬	Confusion Matrix	Matrix comparing to the actual value and the forecast value.
배치	Batch	To repeat model training

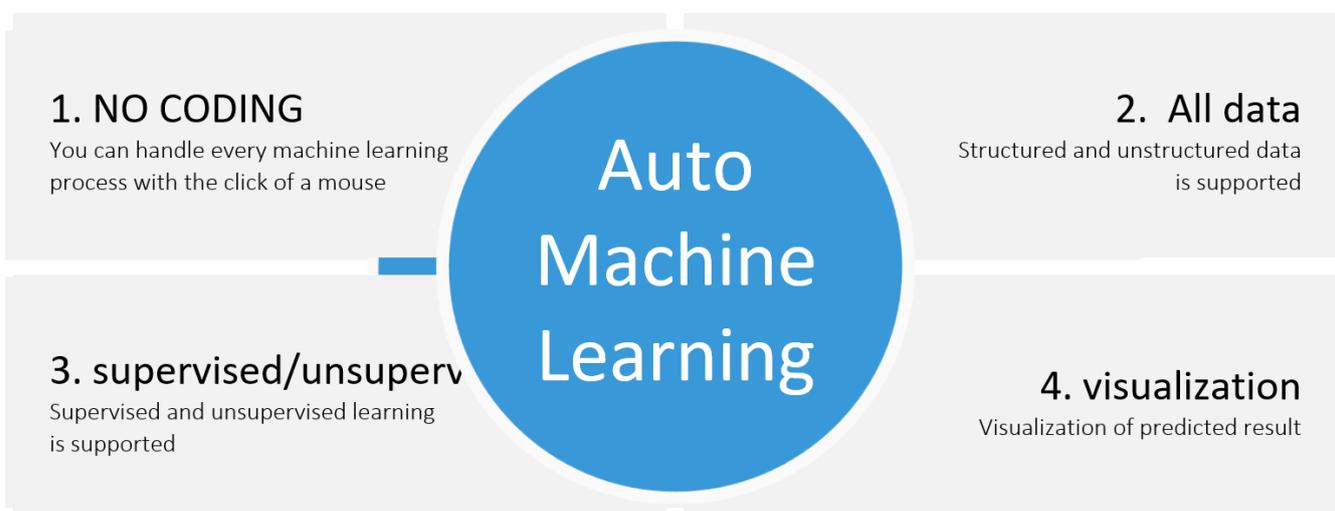
# Chapter 1. Introduction of WiseProphet

This chapter introduces simply main functions and features of WiseProphet before using it.

## 1.1. Summary

WiseProphet is the platform to au automate machine learning which gets prediction result easily with data. After collecting data and conducting various models automatically, the predictive result comes up with the optimized algorithm.

## 1.2. Features



### 1.2.1. Ease of use

- Even nonprofessionals can handle machine learning with the click of a mouse
- User-friendly interface

### 1.2.2. Support to every data

- Support not only structured data but also unstructured data: text, images etc.
- Able to analyze features from unstructured data

### 1.2.3. Support with various algorithm

- Providing algorithms: Classification, Regression, Clustering
- Able to connect to the Supervised Learning with various algorithm and the Unsupervised Learning.
- Offering the outlier algorithm for discovering new patterns and the supervised learning algorithm for predictive analysis

### 1.2.4. Feature Extraction

- Systematic support to Data Preprocessing with feature engineering
- Optimizing features by selecting them and presenting importance

### 1.3. Main Function

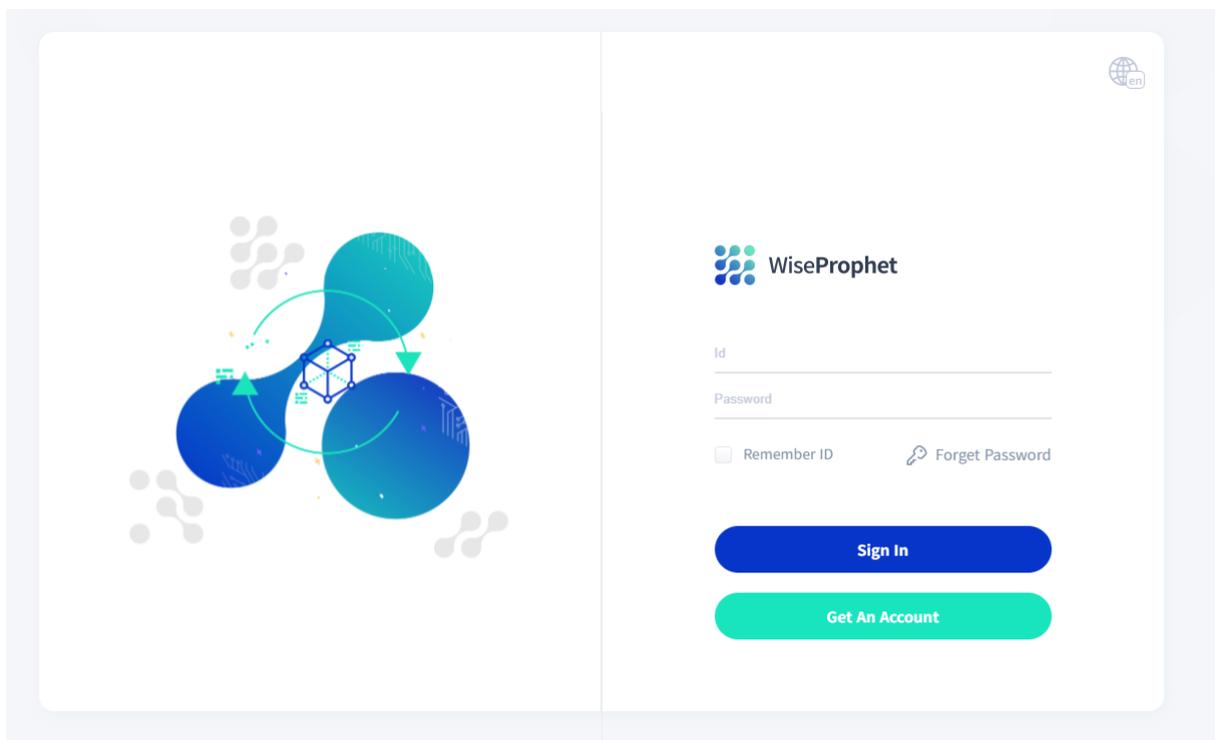
Main function of WiseProphet is as follows.

Function	Description
Preparing data	To provide original data of various Structured and Unstructured Data
Technical Data Preprocessing	Data conversion, data cleansing, variable scaling
Providing various algorithm	Various algorithm by type of prediction Diverse open-source-based algorithm qualified optimally.
Support with learning and tuning parameter	Reapeatable model learning Hyperparameter optimizing and tuning
Model management	Easy deployment of models into various environments Regular monitoring and managing models
Visualization of explaining model	Indicating intuitional Model Evaluation Metrics for understanding forecast results Providing the visualized result of model prediction.

# Chapter 2. Login

This chapter explains how to use WiseProphet for user registration and login.

## 2.1. Screen Layout

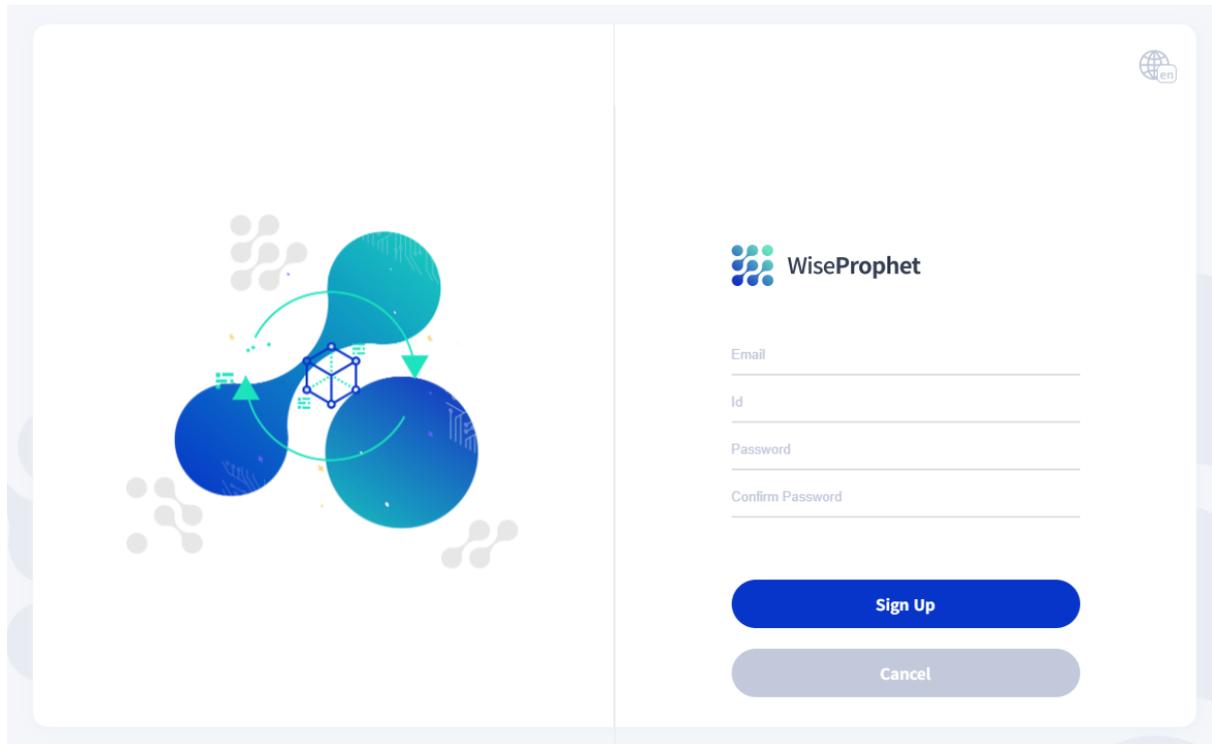


After user registration with 'sign up', the login screen appears.

- 1) The screen above is the main login screen. After entering your email and password, click the login button. Then, the main screen of model learning appears.
- 2) Time for keeping login is an hour. After an hour you should log in again.
- 3) Sign up for a membership with 'sign up' button.
- 4) If you enter the incorrect password 5 times, your account would be locked. To reuse it again, you should send an email to our administrator's account: wiseprophet@wise.co.kr

## 2.2. Signing up

- You can check the privacy policy by clicking 'sign up' button. If you agree it, you can sign up by setting email and password as follows.



The image shows a sign-up form for WiseProphet. The form is split into two columns. The left column contains a decorative graphic with blue and teal circles and lines. The right column contains the form fields and buttons. At the top right of the right column is a globe icon with 'en' next to it. Below the logo, there are four input fields: Email, Id, Password, and Confirm Password. At the bottom of the form are two buttons: a blue 'Sign Up' button and a grey 'Cancel' button.



 **WiseProphet**

Email

Id

Password

Confirm Password

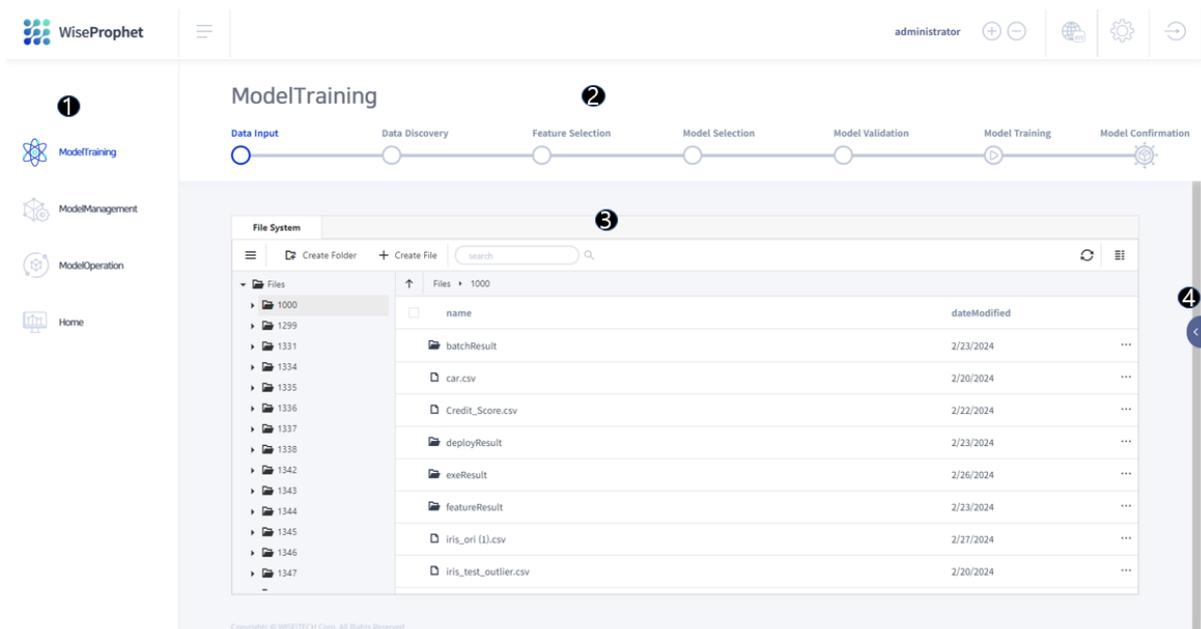
**Sign Up**

Cancel

# Chapter 3. Model Training

This chapter explains how to use WiseProphet for Model Learning.

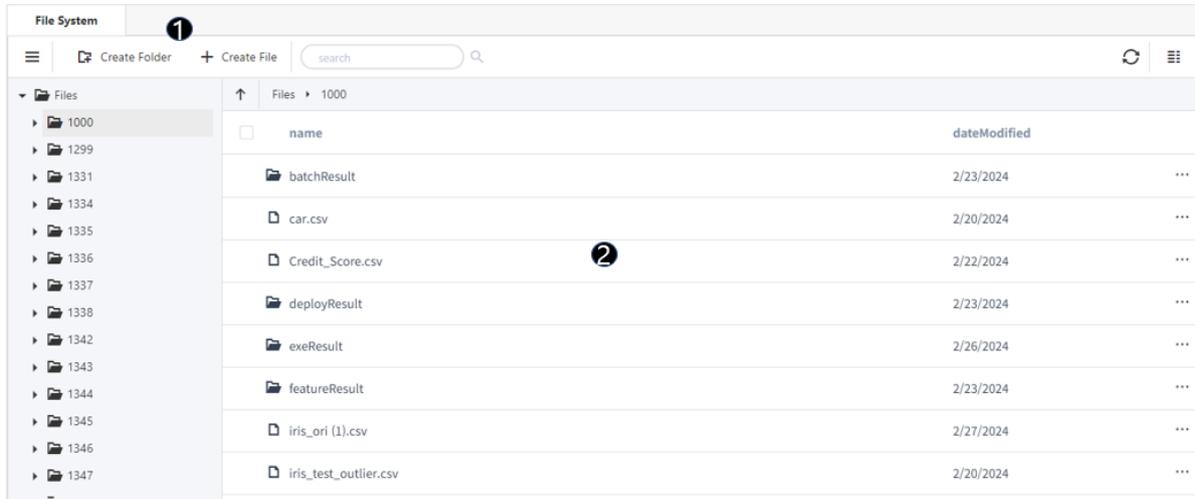
## 3.1. Screen Layout



The screen for selecting data and learning model consists of 1) menu, 2) navigation, 3) screen of model learning, 4) logs.

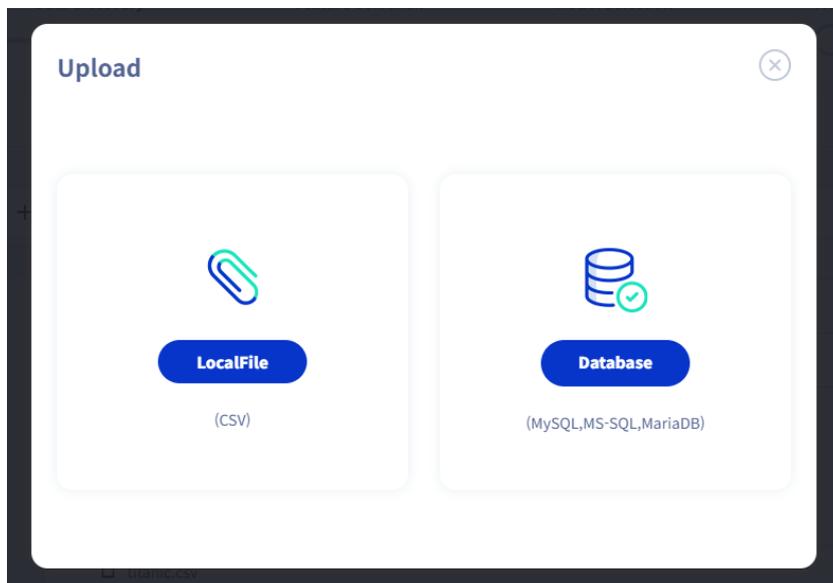
- 1) This is a menu, composed of Model Training, Model Management, Home.
- 2) It is a navigation showing the phases of model learning with 7 steps: Data Input, Data Discovery, Feature Selection, Model Selection, Model Validation, Model Training, Model Confirmation.
- 3) It is a screen of model learning which shows its process according to learning stage. Model learning could perform in order of Data Discovery, Feature Selection, Model Selection, Model Validation, Model Training, Model Confirmation.
- 4) It is space of logs where you can check all the logs to analyze data for model learning.

## 3.2. Data Input



The default screen for data selection consists of 1) create folder and upload file 2) a list of folders and data.

### 3.2.1. Create File



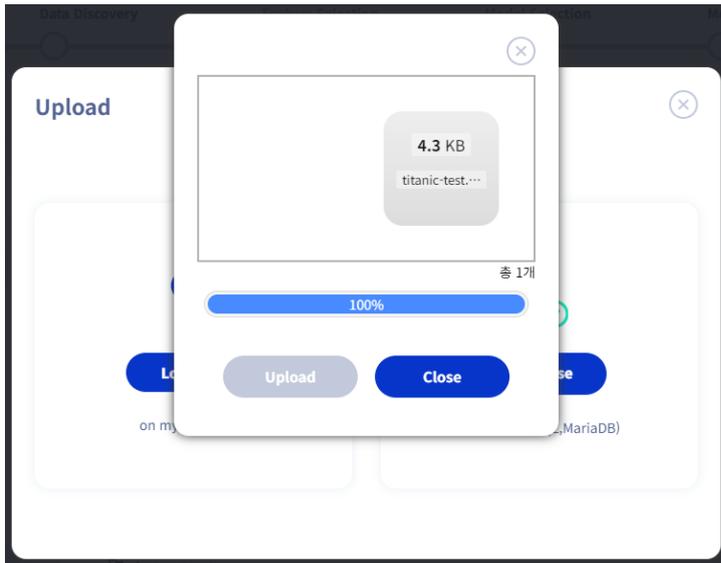
Local File

- 1) Click the LocalFile button to upload the data.
- 2) Supported file types include csv formats.

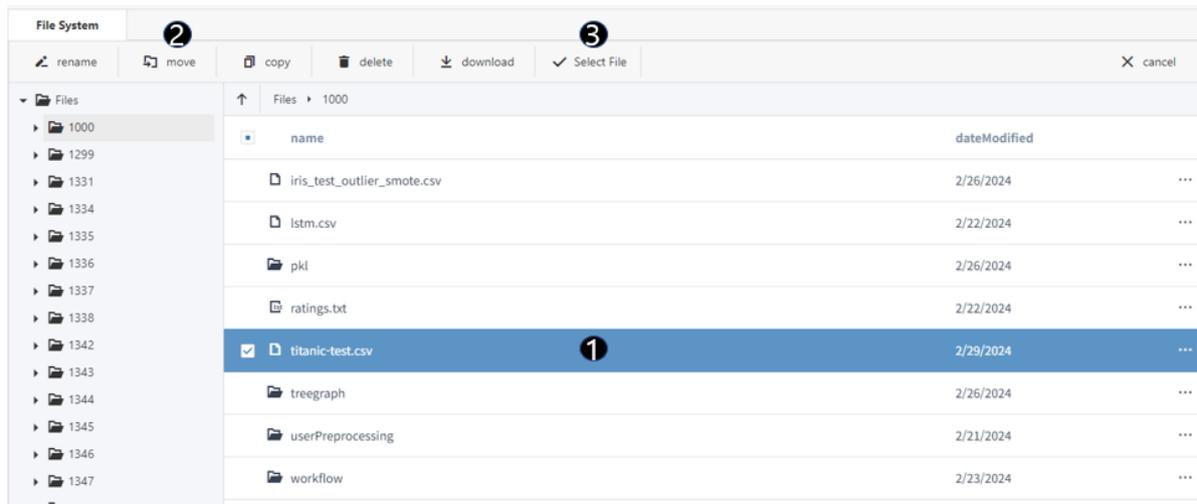
Database

- 1) The database retrieves data from the selected table.
- 2) Database supported include MySQL, MS SQL, and Oracle.

### 3.2.2. Local File



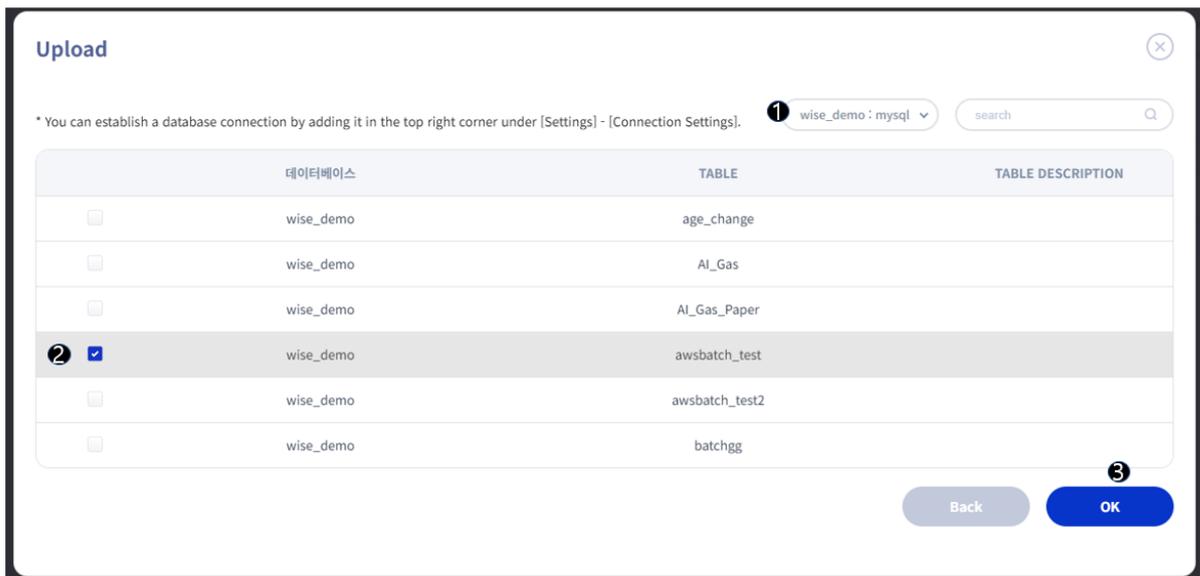
After selecting the data to upload, click the upload button to initiate the upload process. Once the file is uploaded, you can select the file from the data selection default screen.



- 1) Select the data for analysis.
- 2) Click choose file can start model training with the selected data.
- 3) Rename/move/delete/download the selected data as needed

### 3.2.3. Database

To upload data is supported by 'Sample Database'.



- 1) Select the name of database.
- 2) Choose the table name.
- 3) Click the 'SELECT' button.

### 3.3. Data Discovery

Data Discovery is the phase of figuring out the structure and feature of data and checking basic statistics.

ModelTraining

Data Input  Data Discovery  Feature Selection  Model Selection  Model Validation  Model Training  Model Confirmation

DATABASE

#### Data Discovery

Standard Scale Search keyword

Dataset Info ( Missing : 0 Duplicate : 0 Columns : 14 Rows : 19090 )

	NAME	TYPE	MIN	MAX	UNIQUE VA...	MISSING	MEAN	STDDEV	USE	TARGET
Q	age	Nu...	29	77	41	0	54.37	9.07	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Q	sex	Nu...	0	1	2	0	0.68	0.47	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Q	cp	Nu...	0	3	4	0	0.97	1.03	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Q	trestbps	Nu...	94	200	49	0	131.63	17.52	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Q	chol	Nu...	126	564	152	0	246.28	51.76	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Q	fbs	Nu...	0	1	2	0	0.15	0.36	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Copyrights © WISETECH Corp. All Rights Reserved.

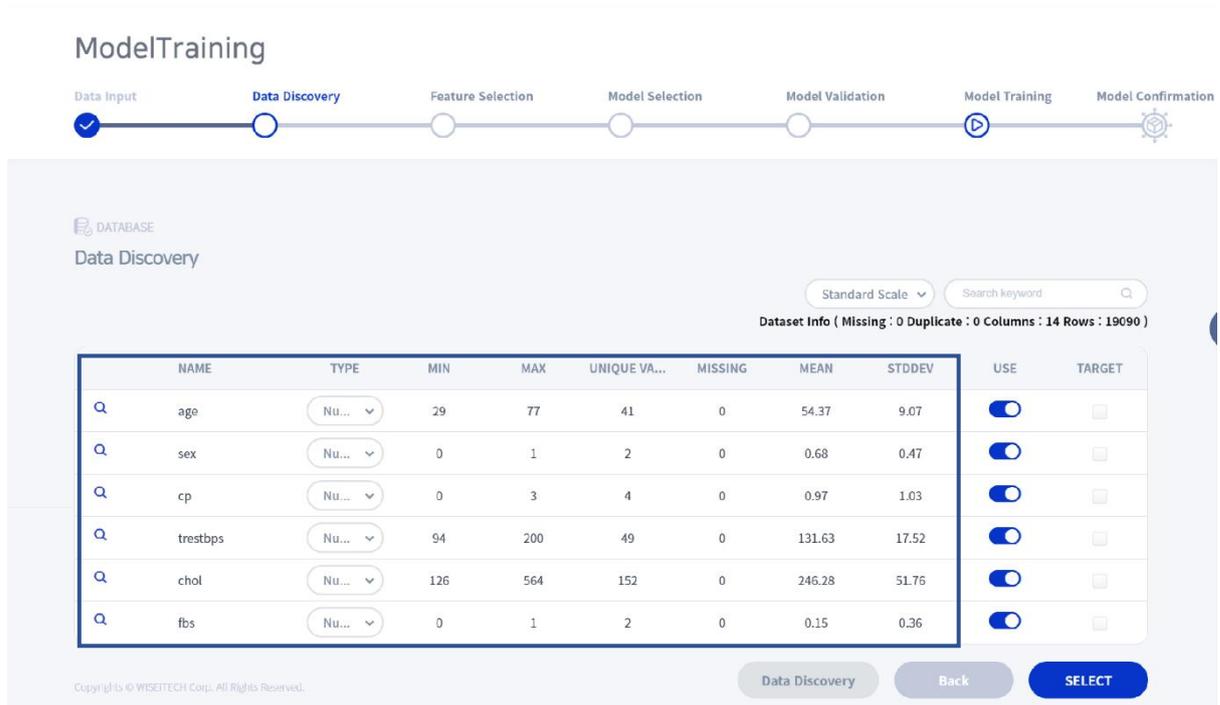
Data Discovery Back SELECT

- Here is the first page of Data Discovery.

- It consists of 1) Exploratory Data Analysis 2) Data Scaling 3) Data Distribution 4) To change the Data Type 5) Whether to use variables 6) To set the target.

### 3.3.1. Exploratory Data Analysis

In the phase of Exploratory Data Analysis, you can check basic statistics and a list of variables uploaded in the phase of Data Input. There is Dataset Info in the upper right of the table.



- In the page of Data Discovery, you can check basic statistics as follows.

Name	meaning
NAME	A variable name
TYPE	Types of variables(Numerical, Categorical)
MIN	The lowest value
MAX	The highest value.
First quartile	The lowest 25% of data
Third quartile	The lowest 25% of data
UNIQUE VALUE	The unique value after Data Deduplication
MISSING	No data value **
USE	Whether to use variables
TARGET	The target variable you want to predict

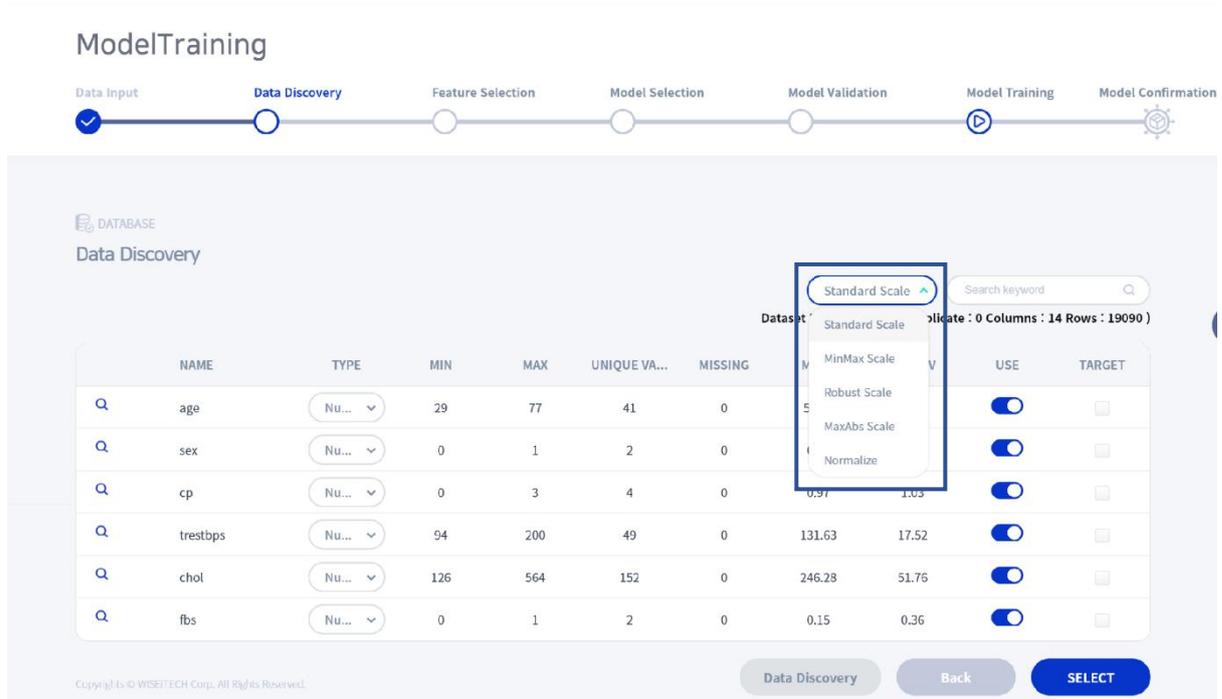
\*\* Without Missing Value Imputation it would be automatically replaced by 0.

### 3.3.2. Data Scaling

You should adjust the ranges of each data equally because they are different from each other. This work is called Scaling. You should set several variables in the same range by increasing or reducing their range.

For example, you are predicting it from height and weight how long 100 metres take for. The height and the weight have different ranges, so the height with its bigger value can affect the result. Therefore, you should convert the ranges of height and weight to 0~1. Ex) 5.7ft(175cm) → 0.175

There are the many methods for scaling: Standard scale, MinMax scale, Robust Scale, MaxAbs scale, and normalize.



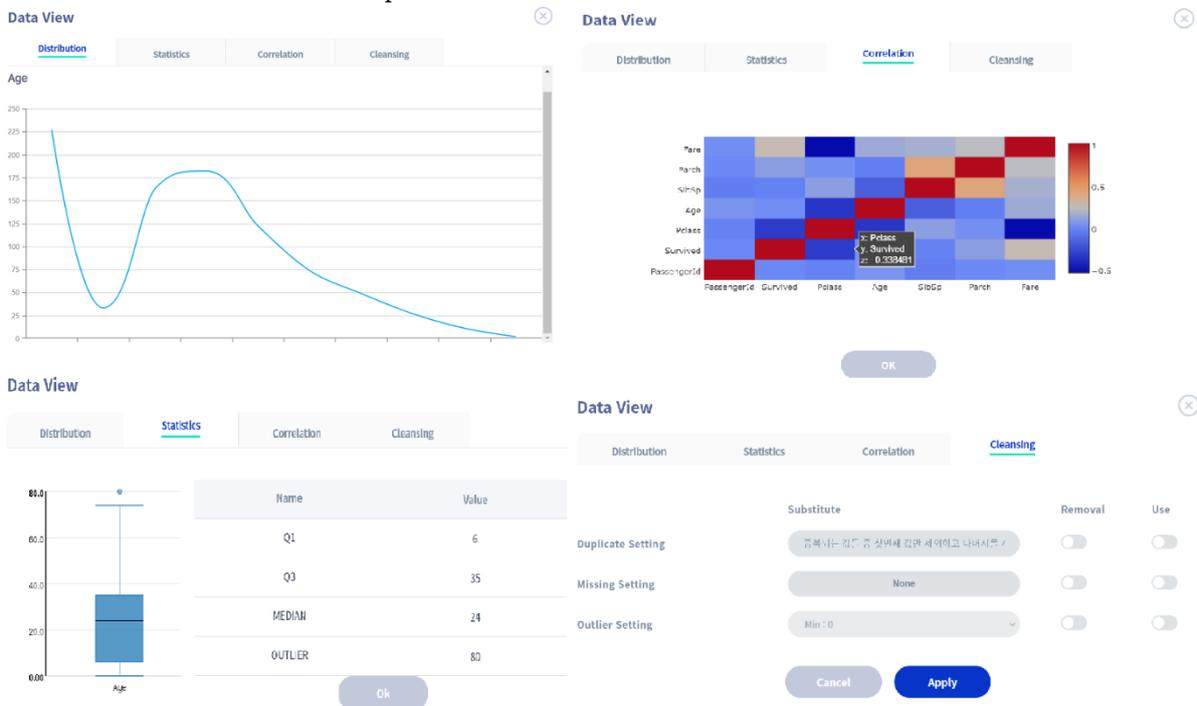
- This is a screen for Data Scaling.
- You can change how to scale data: Standard scale, MinMax scale, Robust Scale, MaxAbs scale, and normalize.
- How to scale data in WiseProphet is as follows.

Name	Description
Standard Scale	It converts data range to Normal distribution. Which has a distribution form of left-right symmetry based on the center value.
Min-Max Scale	It converts data range to 0-1.

Robust Scale	It converts data range based on the center value and the quartile range. When there are many outliers, Robust Scale is used to minimize effect of them.
Maxabs Scale	It converts data range so as to be max absolute value to 1 and 0 to 0. Data with only positive number works like MinMax Scale. But it could be sensitive to large outlier.
Normalization	It converts the actual range to the standard range: -1 ~ +1 or 0 ~ 1.

### 3.3.3. Data Distribution

If the number of data is small, you can look at its values. But if not, you need to check what shape data makes and what value is near to.



The shape of data is called Data Distribution.

- This is a screen for checking Data Distribution.
- Click each magnifying glass button, you can see its distribution, detailed statistics, correlations among variables. You can cleanse data about duplicate value, missing value, and outlier.
- You can check the histogram showing data distribution of variables at Distribution, the box plot representing Q1, Q3, MEDIAN, OUTLIER at Statistics, and the correlations among variables at Correlation.

- Histogram is a method which divides available range of data into several parts and calculates the frequency of each part.
- First quartile and third quartile are explained in a 3.3.2 table, outlier means the observed value smaller than 1<sup>st</sup> quartile or larger than 3<sup>rd</sup> quartile.
- Correlation indicates how correlated variable A and B are. It has a value between -1 and 1. The closer to 1 its absolute value gets, the higher correlation of them gets.
- You can decide whether to use duplicate value, missing value, and outlier or delete them at Cleansing. If you click the 'Use' button, 'Removal' and 'Substitute' are activated. You can remove it by clicking 'Removal' or cleanse data by setting alternative value.
- By applying Data Cleansing, 'Data Discovery' button is activated. If you want to restore data to its original state, click the same button.

### 3.3.4. To change the Data Type

You can change types of variables: Categorical and Numerical

ModelTraining

Data Input    Data Discovery    Feature Selection    Model Selection    Model Validation    Model Training    Model Confirmation

DATABASE  
Data Discovery

Standard Scale    Search keyword

Dataset Info ( Missing : 0 Duplicate : 0 Columns : 14 Rows : 19090 )

NAME	TYPE	MIN	MAX	UNIQUE VA...	MISSING	MEAN	STDDEV	USE	TARGET
age	Nu...	29	77	41	0	54.37	9.07	<input checked="" type="checkbox"/>	<input type="checkbox"/>
sex	Categorical	0	1	2	0	0.68	0.47	<input checked="" type="checkbox"/>	<input type="checkbox"/>
cp	Numerical	0	3	4	0	0.97	1.03	<input checked="" type="checkbox"/>	<input type="checkbox"/>
trestbps	Text	94	200	49	0	131.63	17.52	<input checked="" type="checkbox"/>	<input type="checkbox"/>
chol	Date	126	564	152	0	246.28	51.76	<input checked="" type="checkbox"/>	<input type="checkbox"/>
fbs	Image	0	1	2	0	0.15	0.36	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	Nu...								

Copyright © WISETECH Corp. All Rights Reserved.

Data Discovery    Back    SELECT

### 3.3.5. To use variables

You can choose whether to use variables for prediction. You can exclude unnecessary variables when generating analysis model.

The screenshot shows the 'Data Discovery' step in a model training workflow. At the top, a progress bar indicates the current step is 'Data Discovery', with other steps being 'Data Input', 'Feature Selection', 'Model Selection', 'Model Validation', 'Model Training', and 'Model Confirmation'. Below the progress bar, the 'Data Discovery' section displays a table of dataset variables. The table has columns for NAME, TYPE, MIN, MAX, UNIQUE VA..., MISSING, MEAN, STDDEV, USE, and TARGET. The 'USE' column contains toggle switches, and the 'TARGET' column contains checkboxes. A blue box highlights the 'USE' column. At the bottom right, there are buttons for 'Data Discovery', 'Back', and 'SELECT'.

NAME	TYPE	MIN	MAX	UNIQUE VA...	MISSING	MEAN	STDDEV	USE	TARGET
age	Nu...	29	77	41	0	54.37	9.07	<input checked="" type="checkbox"/>	<input type="checkbox"/>
sex	Nu...	0	1	2	0	0.68	0.47	<input checked="" type="checkbox"/>	<input type="checkbox"/>
cp	Nu...	0	3	4	0	0.97	1.03	<input checked="" type="checkbox"/>	<input type="checkbox"/>
trestbps	Nu...	94	200	49	0	131.63	17.52	<input checked="" type="checkbox"/>	<input type="checkbox"/>
chol	Nu...	126	564	152	0	246.28	51.76	<input checked="" type="checkbox"/>	<input type="checkbox"/>
fbs	Nu...	0	1	2	0	0.15	0.36	<input checked="" type="checkbox"/>	<input type="checkbox"/>

- This is a screen for deciding whether to use variables
- You can choose not to use variables after checking unnecessary variables.

### 3.3.6. To set the target

- You can TARGET the variable you want to predict.
- Ex) A target of stock prediction is Stock. A target of predicting the survivors from Titanic is whether they survive.

## ModelTraining

Data Input — Data Discovery — Feature Selection — Model Selection — Model Validation — Model Training — Model Confirmation

DATABASE
Standard Scale v
Search keyword Q

Dataset Info ( Missing : 0 Duplicate : 0 Columns : 14 Rows : 19090 )

	NAME	TYPE	MIN	MAX	UNIQUE VA...	MISSING	MEAN	STDDEV	USE	TARGET
<span style="font-size: 0.8em;">Q</span>	age	Nu... <span style="font-size: 0.8em;">v</span>	29	77	41	0	54.37	9.07	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<span style="font-size: 0.8em;">Q</span>	sex	Nu... <span style="font-size: 0.8em;">v</span>	0	1	2	0	0.68	0.47	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<span style="font-size: 0.8em;">Q</span>	cp	Nu... <span style="font-size: 0.8em;">v</span>	0	3	4	0	0.97	1.03	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<span style="font-size: 0.8em;">Q</span>	trestbps	Nu... <span style="font-size: 0.8em;">v</span>	94	200	49	0	131.63	17.52	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<span style="font-size: 0.8em;">Q</span>	chol	Nu... <span style="font-size: 0.8em;">v</span>	126	564	152	0	246.28	51.76	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<span style="font-size: 0.8em;">Q</span>	fbs	Nu... <span style="font-size: 0.8em;">v</span>	0	1	2	0	0.15	0.36	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Data Discovery
Back
SELECT

Copyrights © WISETECH Corp. All Rights Reserved.

- This is a screen for targeting.
- You can target variables for purpose of analysis. But cannot in clustering.
- If you decide a target, you can go to the phase of Feature Selection. If not, you will go to the phase of Model Selection.

## 3.4. Feature Selection

Feature Selection means to select the optimal features when you are about to classify or predict a target variable. For example, when you predict 100 metres runners' record—a target—, it would be affected by their averages, ages, heights, weights, experiences of winning. You can add some feature necessary to prediction to those factors or exclude one unnecessary from them.

At the phase of Feature Selection, you can use only features set to 'use' at the phase of Data Exploration and extract feature by importance to the target. If you did not choose a target, this phase would be skipped.

How to select the optimal feature is (1) the effect by feature and (2) Correlation analysis

### 3.4.1. The Effect by Feature

The Effect by Feature means each variables' relative importance to a target. For example, if the importance value indicates it to predict runners' record that the average is 0.4, the age is 0.3, the height is 0.1, and the weight is 0.1, it means that 'the average' affects 40% for prediction.

The screenshot displays the 'ModelTraining' interface. At the top, a progress bar shows the following steps: Data Input (checked), Data Discovery (checked), Feature Selection (active), Model Selection, Model Validation, Model Training, and Model Confirmation. Below the progress bar, the 'Feature Selection' section is active, showing a bar chart of feature importance and a table of selected features.

**Feature Selection**

Forward Selection | Backward Elimination

Correlation

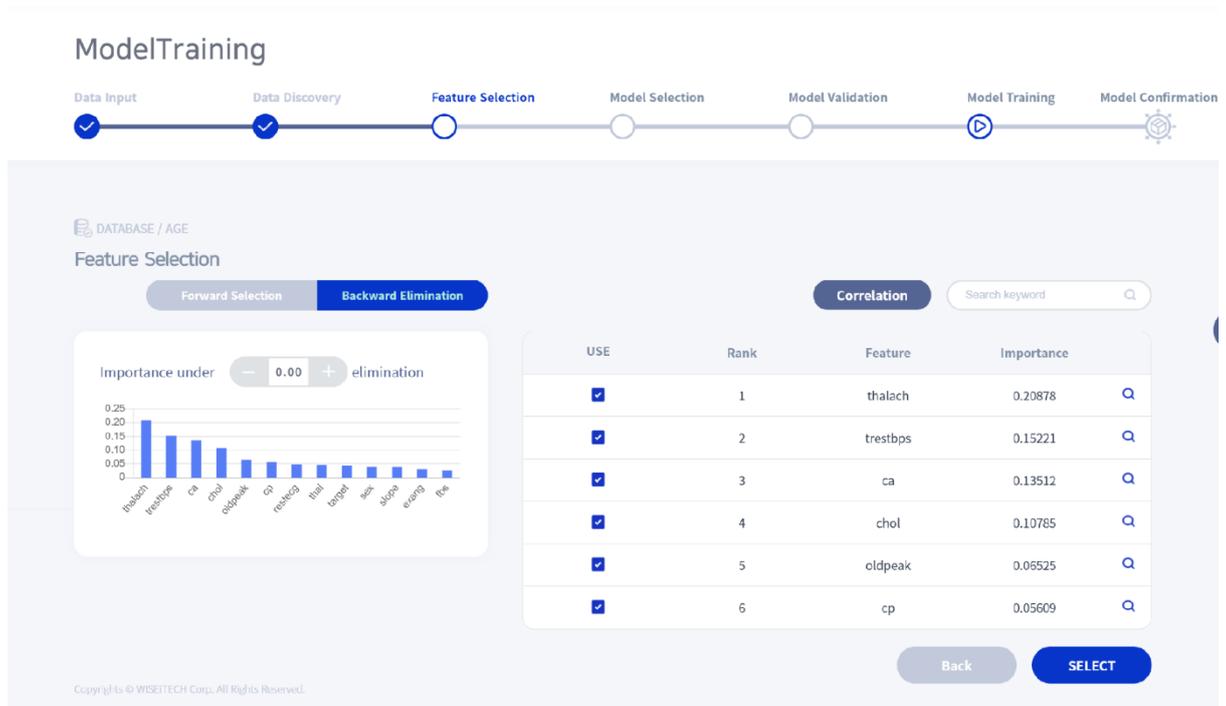
Feature Rank: 13

USE	Rank	Feature	Importance
<input checked="" type="checkbox"/>	1	thalach	0.20878
<input checked="" type="checkbox"/>	2	trestbps	0.15221
<input checked="" type="checkbox"/>	3	ca	0.13512
<input checked="" type="checkbox"/>	4	chol	0.10785
<input checked="" type="checkbox"/>	5	oldpeak	0.06525
<input checked="" type="checkbox"/>	6	cp	0.05609

Back | SELECT

- This is a screen which 'Forward Selection' is clicked at the Effect by Feature in the phase of Feature Selection.
- 1) If you click Forward Selection, you can select features in order of the importance.

- 2) With plus button, variables are added in order of importance, with minus button, they are in reverse order.
- 3) Click each magnifying glass icon at the right of variables, you can see each data distribution of features scaled by methods set at the phase of Data Exploration.



- This is a screen which 'Backward Elimination' is clicked at the Feature Selection.
- 1) If you click plus or minus button, it increases or decreases by 0.01. And features below the certain importance are excluded from variables.
- 2) You can also customize importance and exclude features below the customized value.

### 3.4.2. Correlation analysis

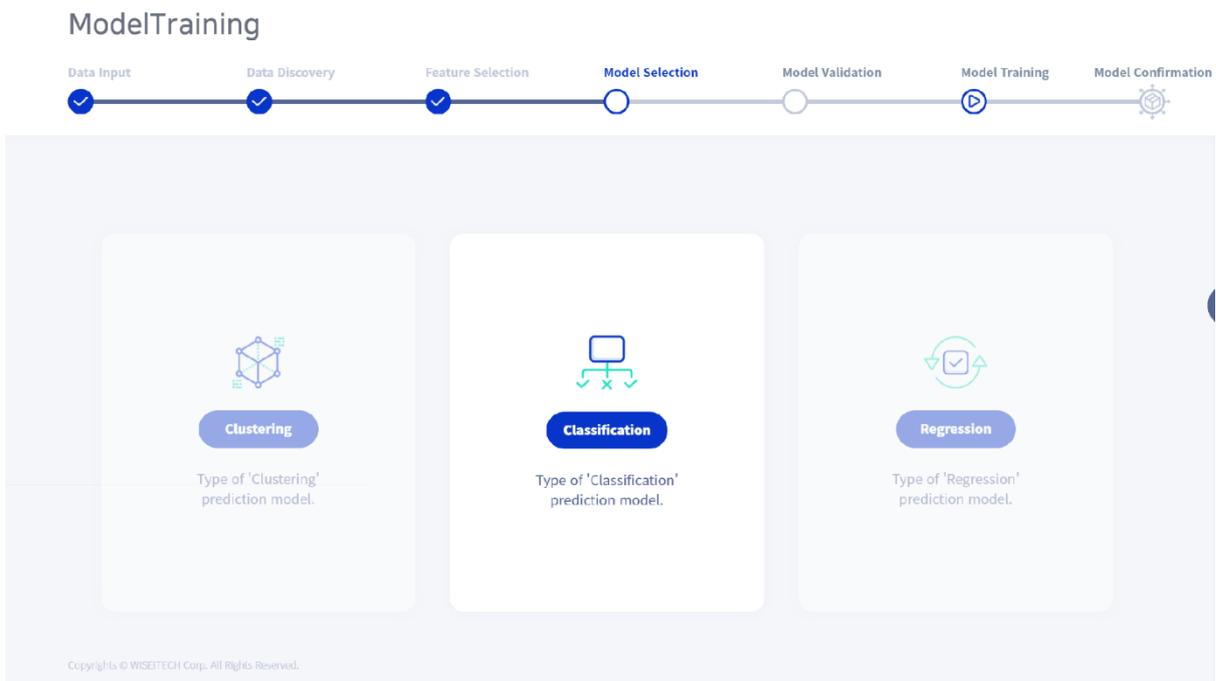
Correlation indicates how correlative two variables are. It is between  $-1$  and  $1$ . The closer its absolute value gets to  $1$ , the higher their correlation is. In selecting a variable with correlation, if correlation of two variables is too high, one of them should be excluded.



- Click the result of correlation analysis, and you can check a target variable and correlation of other variables.

### 3.5. Model Selection

You can select model to use for prediction at the phase of Model Selection. There are types of prediction: Clustering, Classification, and Regression.

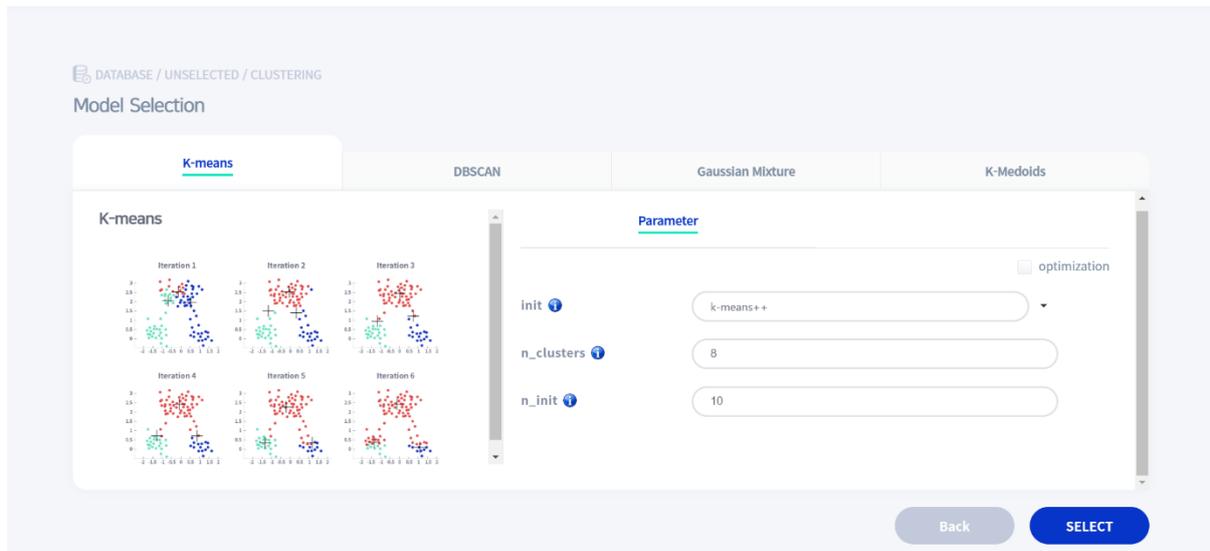


- This is a screen for Model Selection.
- 1) Type of prediction model consists of Clustering, Classification, and Regression. By the type of a target set at the phase of Data Exploration, the numeric activates Regression, the categorical does Classification, and the undefined does Clustering.
- 2) If you choose a type of prediction, you can see a screen of its own Model Selection.

### 3.5.1. Clustering

Clustering means to group similar data together. It is one of typical unsupervised learning, and not necessary to set a target.

- ① K-means: One of clustering model to make K clusters based on K Centroids.



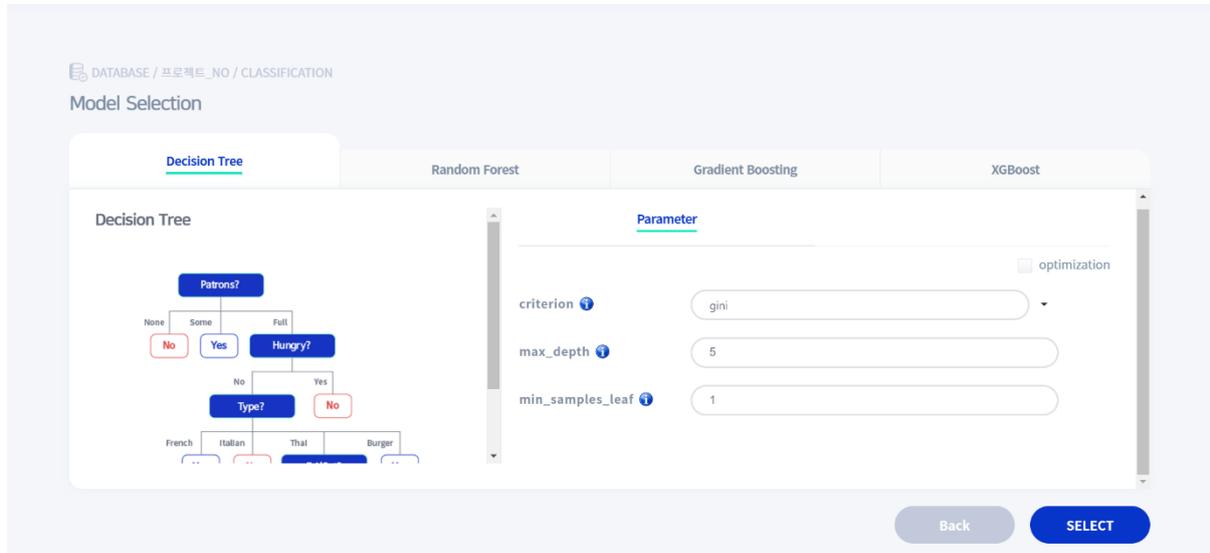
- 1) Choose “Clustering” from model types.
- 2) Select K-means.
- 3) Input parameters. Especially you should do n\_clusters, which means the number of clusters.

Name of parameter	Description
init	A cluster-centered method to initialize. - k-means++ : to select a center of initial cluster for raising speed of collection. - random: to select K at random.
n_clusters	- The number of cluster
n_init	- The number of cluster-centered initialization

### 3.5.2. Classification

Classification means the way to classify what category it belongs to. It is used to predict categorical variables.

- ① Decision Tree: It is called Decision Tree, cause it shows data’s pattern as combination of predictable rules, and it looks like trees. It is similar to the twenty questions game to narrow the range of answer by questions.



- 1) Choose “Classification” in the model types.
- 2) Select Decision Tree.
- 3) Input parameters referring to the instructions below.

Name of parameter	Description
criterion	The standard to divide tree -gini: impurity, a measure of how much impurities are mixed -entropy: a measure of uncertainty
max_depth	-a standard of how deep a tree is built up
min_samples_leaf	-The number of minimum sample to have for being a node

### 3.5.3. Regression

Regression is used to predict continuous values.

- ① Random Forest: It is a kind of ensemble learning to be used for Classification and Regression. It works by printing out average predicted value or classifying from many decision trees formed in learning. After a number of decision trees are made up, the result is decided by majority vote.

DATABASE / PCLASS / REGRESSION

### Model Selection

Decision Tree

**Random Forest**

Elastic Net

Gradient Boosting

XGBoost

**Random Forest**

The diagram illustrates the ensemble nature of a Random Forest. It shows several individual decision trees, labeled  $\phi_1, \dots, \phi_n$ , each receiving input  $X$ . The outputs of these trees are summed together, represented by a  $\Sigma$  symbol, to produce the final model output.

**Parameter**

optimization

max\_depth ⓘ

n\_estimators ⓘ

random\_state ⓘ

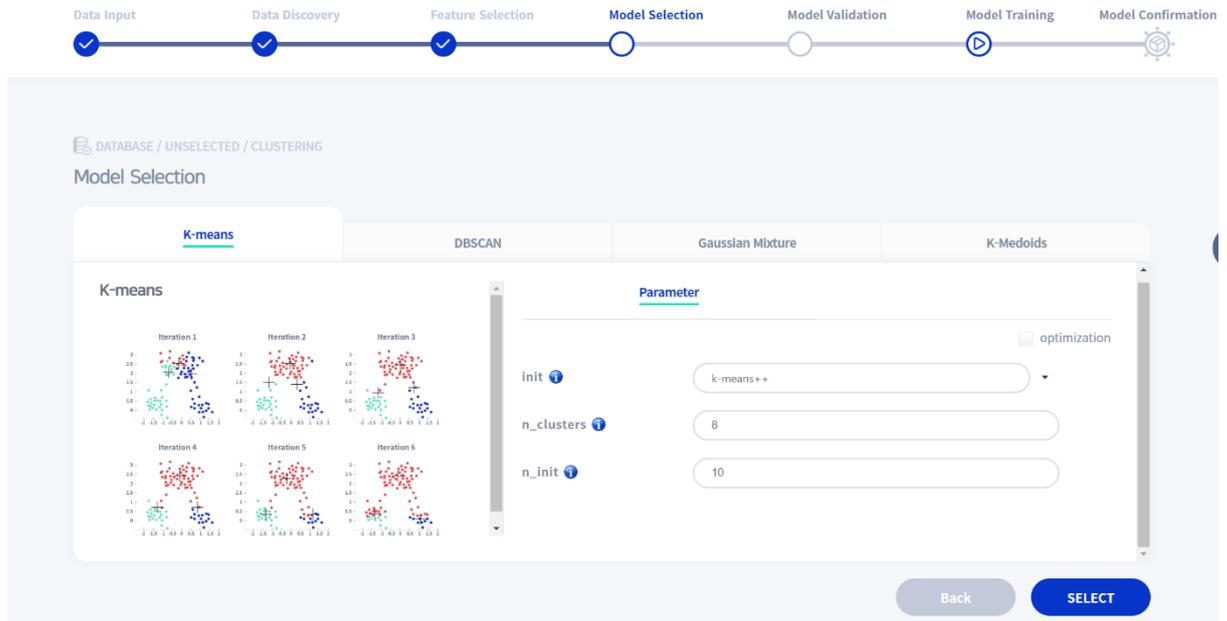
Back
SELECT

- 1) Select “Regression” from model types.
- 2) Choose Random Forest.
- 3) Input parameters referring to the instruction below.

Name of parameter	Description
max_depth	- a standard of how deep a tree is built up
n_estimators	- the number of trees
random_state	- the initial value of random number

### 3.5.4. Parameter Optimization

The optimization is used to optimize parameters of algorithm in generating model. WiseProphet supports the optimal Model Generation with Parameter Optimization.



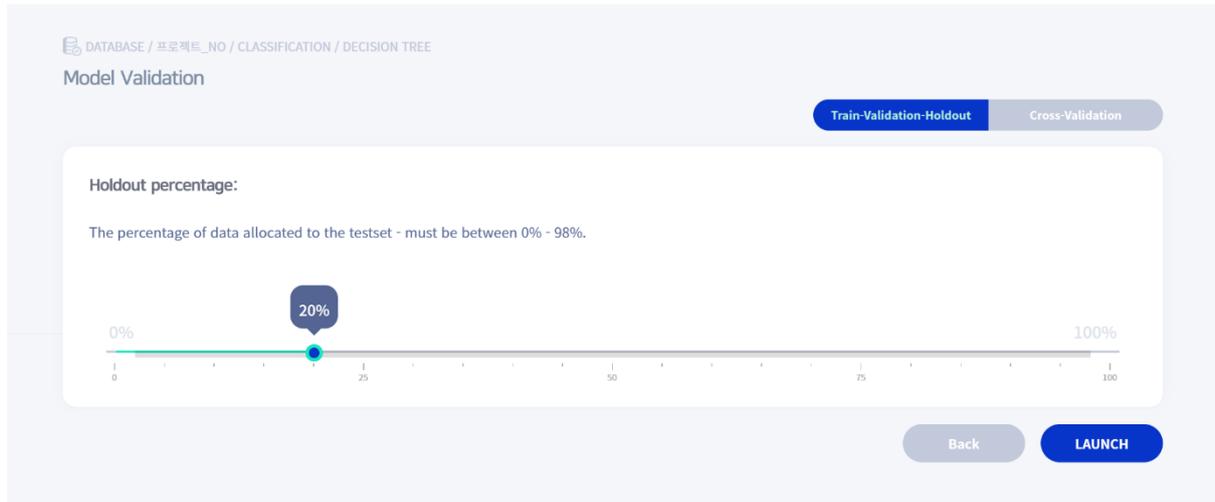
- This is a screen of Parameter Optimization.
- 1) Click the optimization button at the Model Selection.
- 2) Designate the ranges of each parameter of algorithm.
- 3) If you click the SELECT button, the optimal parameter shows up in the range designated
- 4) You can check the optimal parameter in the Logs after executing model.

## 3.6. Setting the ratio of Validation Data

After selecting algorithm, data is verified by dividing into Train and Validation Data. How to validate model is Train-Validation-Holdout and Cross-Validation.

### 3.6.1. Train-Validation Holdout

It is the way to validate the original data dividing into two randomly. One is used as training data for model learning, and the other is as validation data. If you selected the Optimization at the phase of Model Selection, only Train-Validation-Holdout can be used.



- This is a screen of Train-Validation-Holdout in setting the ratio of validation data.
- 1) Click Train-Validation-Holdout
- 2) Set the ratio of validation data. If the ratio is set into 20%, training data would be 80%, and validation data would be 20%.
- 3) Click LAUNCH button, and predictive model will be executed.

### 3.6.2. Cross-Validation

After dividing into subsets of similar size called Folds, each Fold is validated as test data, the others are as train data by the number of Folds times.



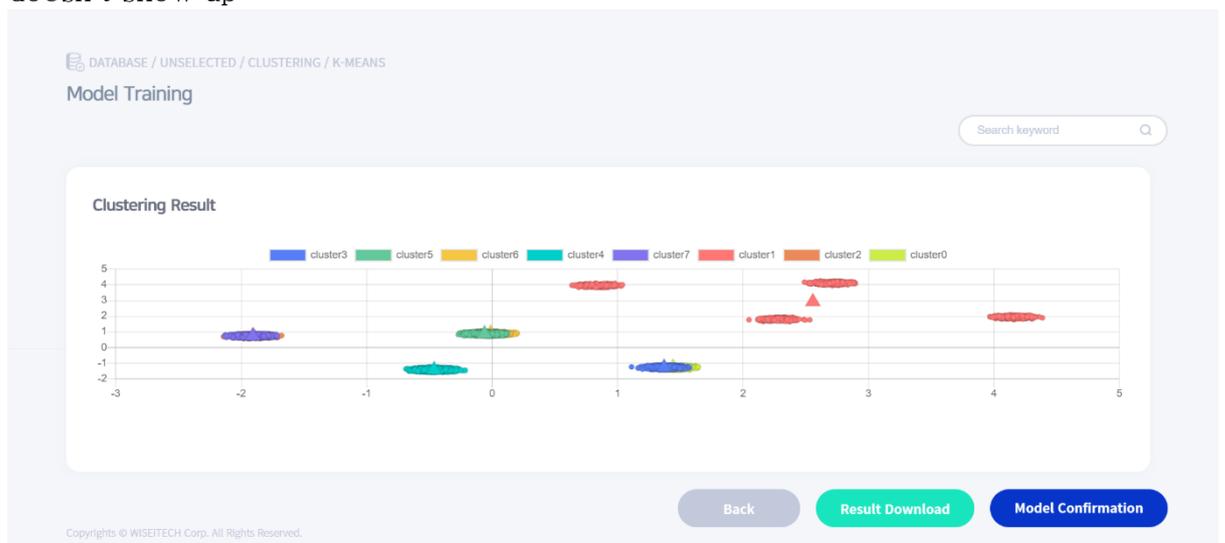
- This is a screen of Cross-Validation in setting the ratio of validation data.
- 1) Click Cross-Validation.
- 2) Set the number of cross-validation Folds
- 3) Click LAUNCH button, and predictive model will be executed.

## 3.7. Model execution

It shows Model Evaluation Metrics depending on Clustering, classification, and regression.

### 3.7.1. Clustering

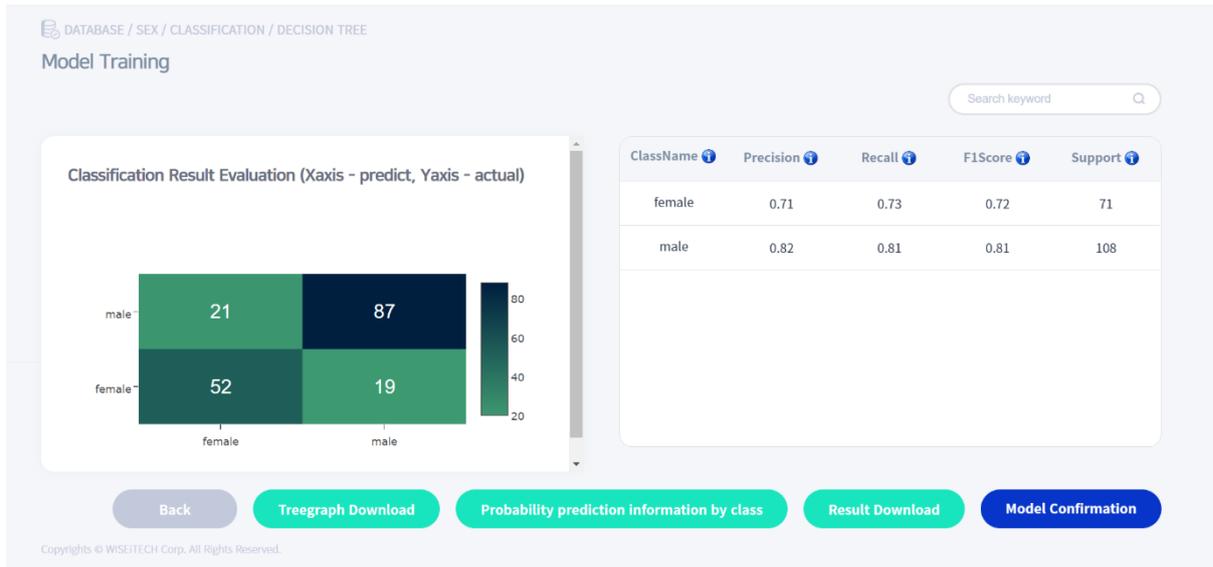
Cluster is divided into K, which is the number the user selected. It can be checked with graph of model execution. Clustering is the unsupervised learning, therefore detailed validation metrics of the model execution result doesn't show up



- This is a screen of Clustering Model Training.
- 1) You can check the Clustering Result by visualization: colour is dividing cluster.
- 2) If you want to download this prediction result, click Result Download. You can download a CSV file.
- 3) With clicking Model Confirmation, model save screen will show up. Enter the model name, and click Save, and Model would be saved.

### 3.7.2. Classification

Classification prediction model shows the actual value and the predictive value of training/validation result by visualizing with Confusion Matrix. And it demonstrates its details: Precision, Recall, F1Score, and Support.



- This is a screen of Classification model execution.
- 1) It visualizes the result of model execution with classification matrix
- 2) It shows Classification Model evaluation metrics with graph. Detailed metrics are as follows.
- 3) In case of Decision Tree algorithm, you can download Treegraph. In case of Classification model, you can download Probability prediction information by class.

\*\* ex: Breast cancer prediction – assuming positive/negative classification model

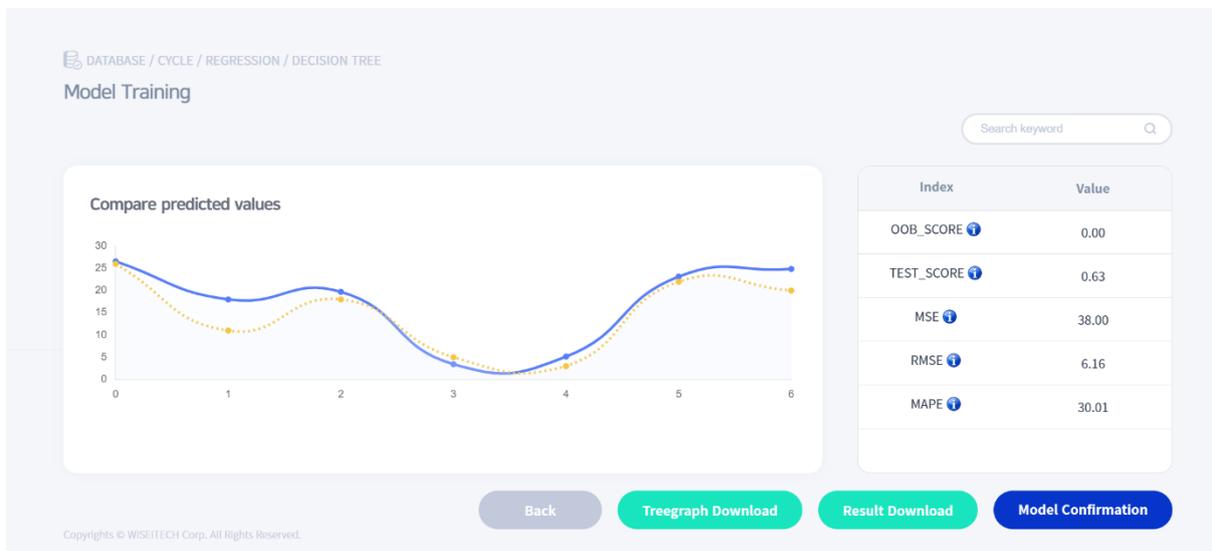
		Actual result	
		positive	negative
Predictive result	positive	80	5
	negative	10	5

Evaluation metrics	description
Accuracy	Percentage of the actual results to match total data (Breast Cancer prediction accuracy = $(80 + 5) / 100 = 85\%$ )
Precision	Percentage of the actual results to match the result predicted positive. (Breast Cancer precision = $80 / (80 + 10) = 88\%$ )
Recall	Percentage of the actual results being positive that match the predicted results.

	(Breast Cancer recall = $80 / (80 + 5) = 94\%$ )
F1Score	Harmonic mean of Precision and Recall
Support	Predicted value's total

### 3.7.3. Regression

Regression prediction model visualize the trend line between actual and predictive values about training and validation results. Also, it shows MSE, RMSE, and MAPE, which is detailed validation metrics about predictive results.

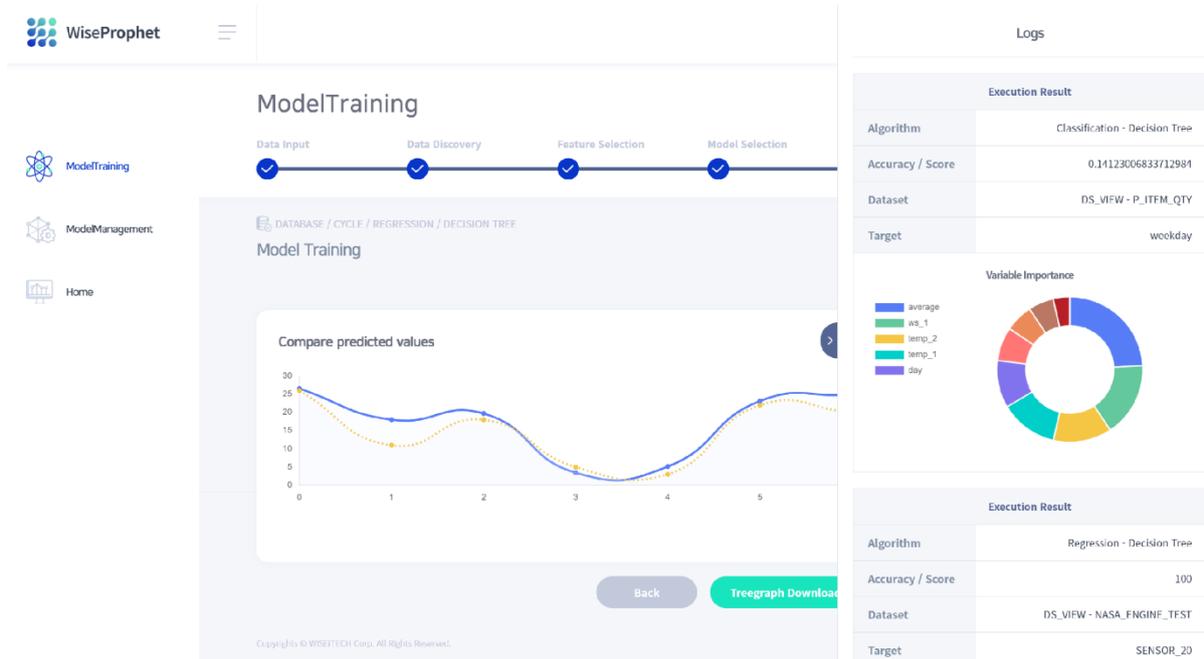


- This is a screen of Regression model execution.
- 1) You can check model execution results with the trend line between actual and predictive values.
- 2) There are detailed validation metrics about predictive results. You can see the description below or check by mouse over the right icon of each validation item.

Validation metrics	Description
RMSE	Square root of predictive and actual values.
MAPE	Percentage of distance difference between predictive and actual values.

### 3.7.4. Model Logs

You can check the data analysis logs in executing model training at the Model Logs page.



- This is a screen of Model Logs.
- 1) A pie chart is the list and proportion of variables which have affected the model prediction.
- 2) Clustering is displayed as a chart in the form of radiation. Also, It is the list and proportion of variables which have affected each clustering set.
- 3) At the model logs, you can check the algorithm types, Precision/score, Data set, and a Target. So you can compare and analyze the results of data analysis easily. In case of clustering, Precision/score means Silhouette score.
- 4) If you chose the optimization at the Model Selection, you can see a list of the optimal parameters.

# Chapter 4. Model Management

This chapter describes how to use WiseProphet for Model Management.

## 4.1. Screen Layout

The Model Management screen consists of 1) menu 2) a list of Predictive models 3) Model Management. You can see a list of Predictive model which was generated at the phase of Model Training, manage model: execution, deletion, entering and predicting the new data, and visualizing the predictive result.

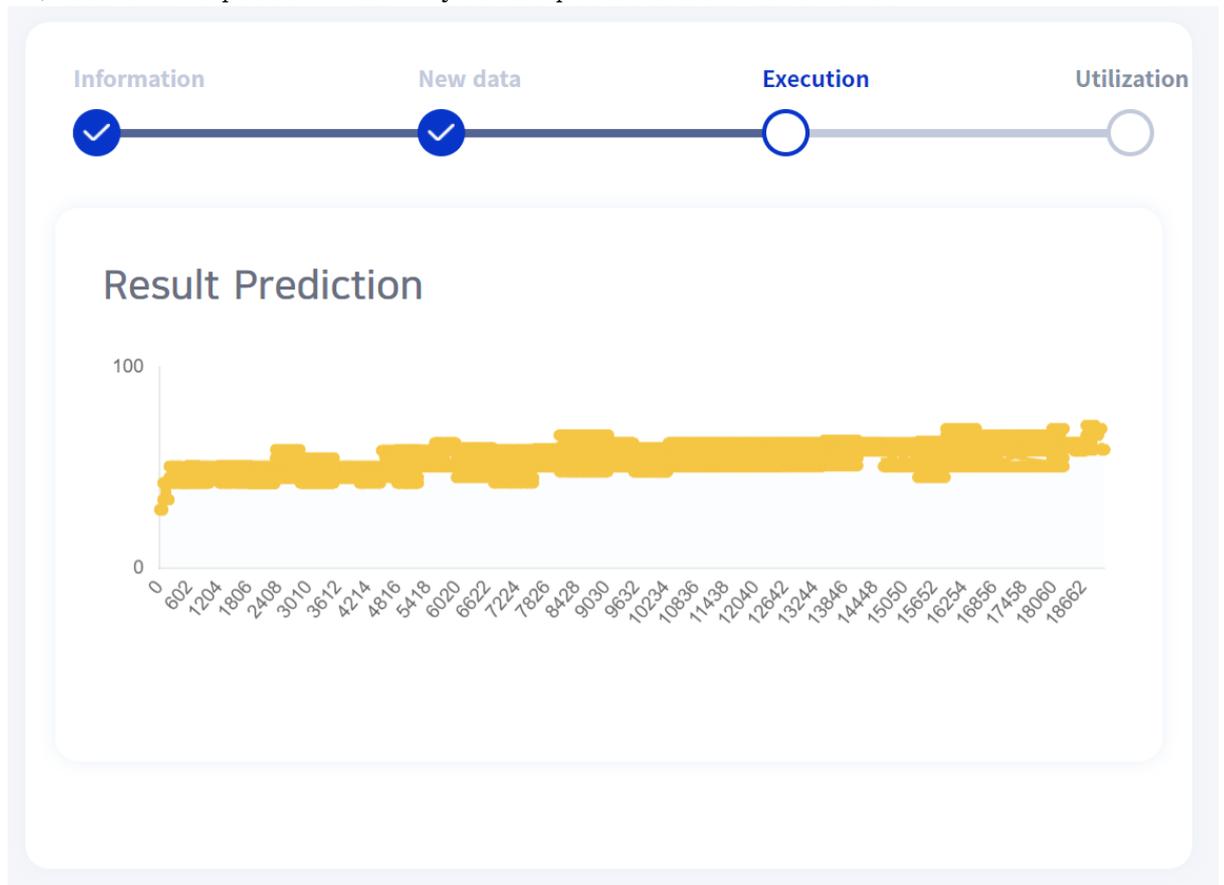
## 4.2. Model Management

The screenshot shows the 'Model Management' interface. On the left, a sidebar contains navigation icons for 'ModelTraining', 'ModelManagement', and 'Home'. The main area is titled 'Model Management' and includes a search bar. Below the search bar is a table of predictive models. The table has columns for 'Model Name', 'Status', 'Data...', and 'Edit'. Two models are listed: 'N1' with status 'waiting' and dataset 'Heart', and 'æ' with status 'waiting' and dataset 'JoeB...'. To the right of the table is a detailed information panel for the selected model 'N1'. The panel includes a progress bar and a table with the following data:

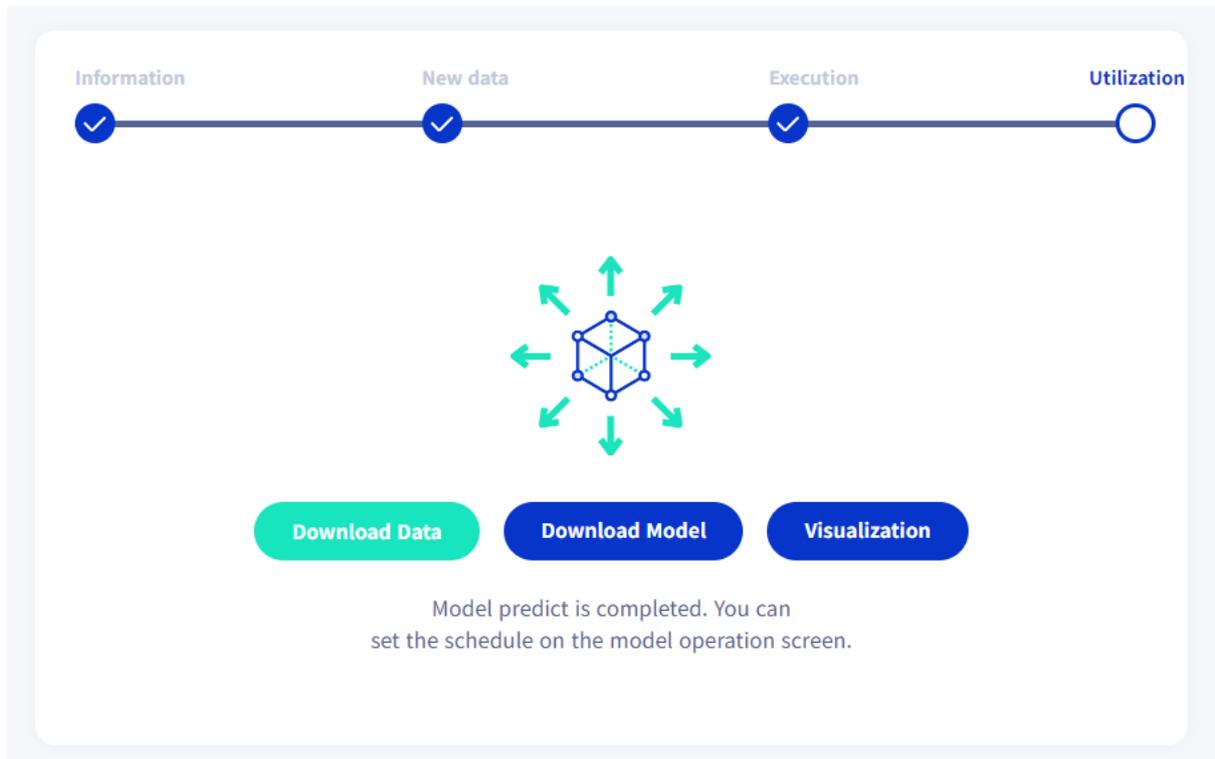
Information	New data	Execution	Utilization
Model Name	N1	Status	
Model Type	Regression	Algorithm	decisiontree
Verification Type	Train-Validation-Holdout	Verification Ratio	20
Dataset Type	DS_VIEW	Dataset Name	Heart
Prediction Rate	91	Target	age

- This is a screen of Model Management.
- 1) It is the same list as Model Training's.
- 2) It shows a list of Predictive Model generated at the phase of Model Training and each one's name, type, and Dataset name.

- 3) Click the execution button in the list of Predictive Model, and you can load values set at the phase of the corresponding Model Generation and modify them by changing existing ones.
- 4) ③ Information of the Predictive Model displays the information of the corresponding model: Model Type, Verification Type and Ratio, Dataset Type, Target, and so on.
- 5) At ③, you can upload another data by clicking New data, and delete unnecessary one by clicking the delete button in the list of Predictive Model. At this time, a variable list of upload data should match Model Training's.
- 6) If the data upload is successful, the Predictive Model will run.
- 7) The above process shows you the prediction result as below.



- 8) Click Utilization after Execution has completed, and you can see a screen as below. You can download the predictive result of uploaded data or model, visualize it.



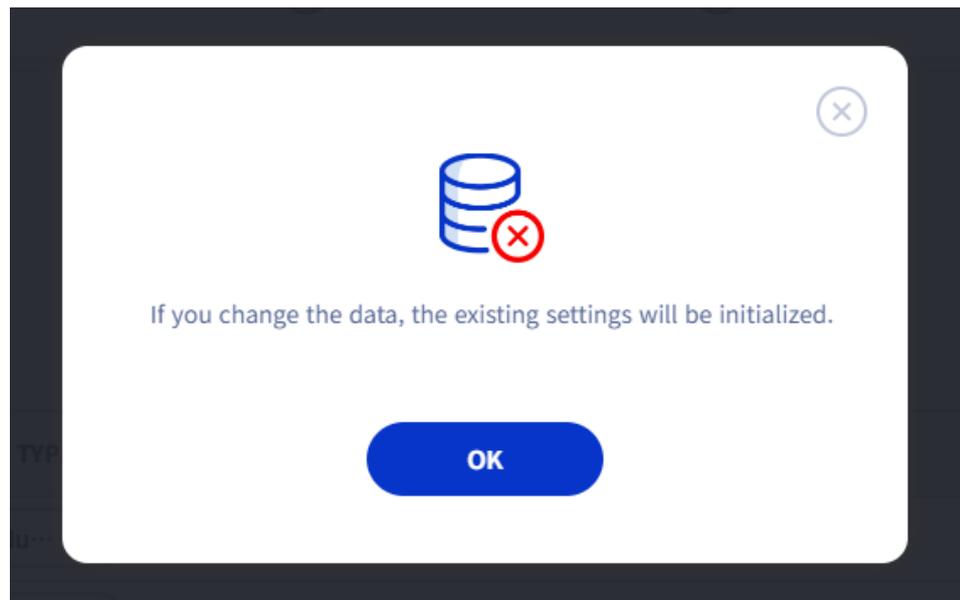
- 9) You can see a screen below by clicking Visualization. If you select a variable you want to visualize, the corresponding one will be added into Columns. If you want to add into Rows or Filter, you can drag and drop it. You can download its graph by clicking the right upper button, and export its table below the graph as an excel file by clicking the excel button.

### 4.3. Model Modification



	Model Name	Status	Data...	Edit
	N1	waiting	Heart	 
	조	waiting	JoeB...	 

- At the Model Management list, you can modify model with  button.
- 1) You are going straight to the Data Discovery phase, you can modify models just like the previous training's.
- 2) Data Cleansing is first converted to stored value, the unset Missing Value is converted to 0 automatically.
- 3) If column type of a variable is modified, the saved settings are initialized.



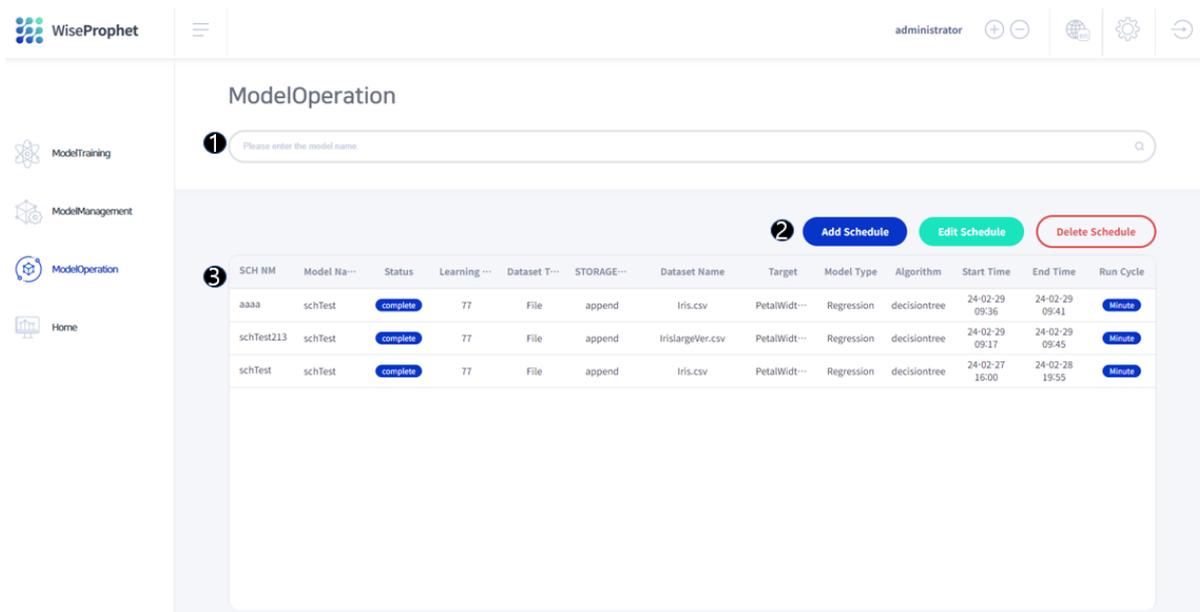
- 4) The value set at saving the model is the default in the Data Initialization.



# Chapter 5. Model Operation

In this chapter, it describes how to use WiseProphet for model operation.

## 5.1. Screen Layout



The screen for operating the model with scheduled tasks consists of 1) schedule search, 2) schedule management, and 3) schedule list.

- 1) Search by entering the schedule name.
- 2) Able to add, modify, or delete schedules.
- 3) Users can view the model, status, execution frequency, algorithm, etc., for currently registered schedules.

## 5.2. Settings and Functions

### 5.2.1. Add Schedule

**Batch schedule** ✕

SCH NM: \*

Start Time: \*

End Time: \*

Model Name: \*

Dataset Type:

Dataset Name:

Excution Unit:

Run Cycle: \*

Setting for storage location

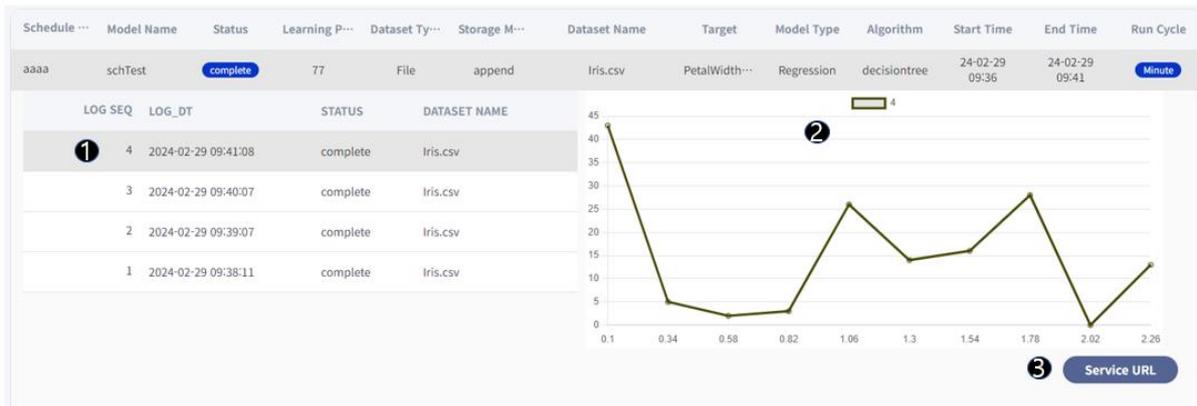
STORAGE FILETYPE: \*

STORAGE NAME: \*

STORAGE MODE: \*

Using the model generated from model training, schedules can be created. Start and end times for the schedule can be set, and the execution unit can also be configured. After filling in all the values, clicking the 'Save' button will promptly register the schedule.

## 5.2.2. Run and verify the schedules



When a schedule is added, it is registered according to the specified start time, and the model runs at the set execution intervals. Clicking on the schedule reveals: 1) Logs, and clicking on these logs displays, 2) Visualizations of the results, 3) URL is provided to access the results in JSON format.

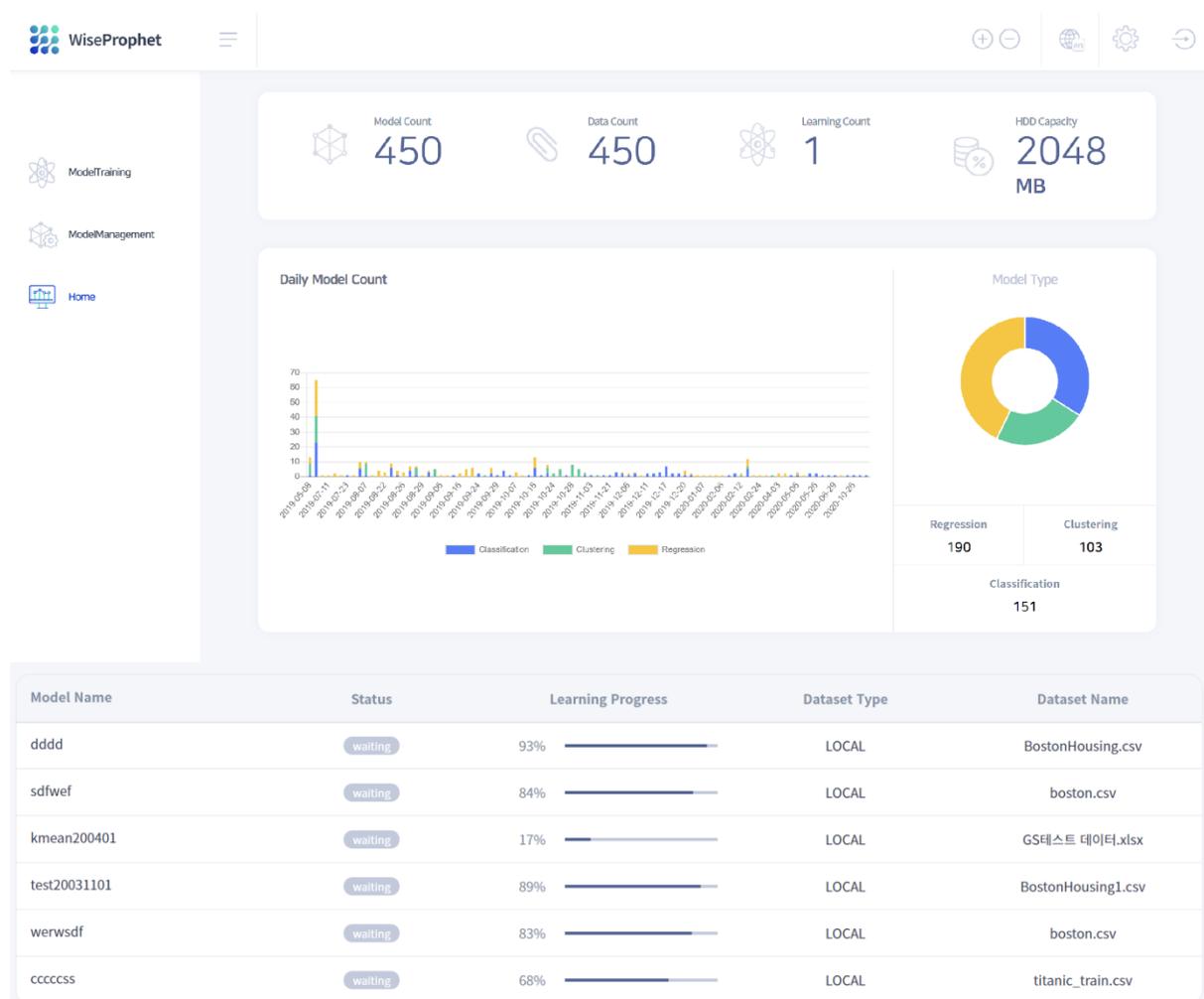
# Chapter 6. Monitoring

This chapter describes how to use WiseProphet for model monitoring.

## 6.1. Screen layout

A page of Home consists of the model a user generated and the user information like model information, data, HDD Capacity, and so on. You can check Daily Model Count and Model Type.

## 6.2. the information of user and model



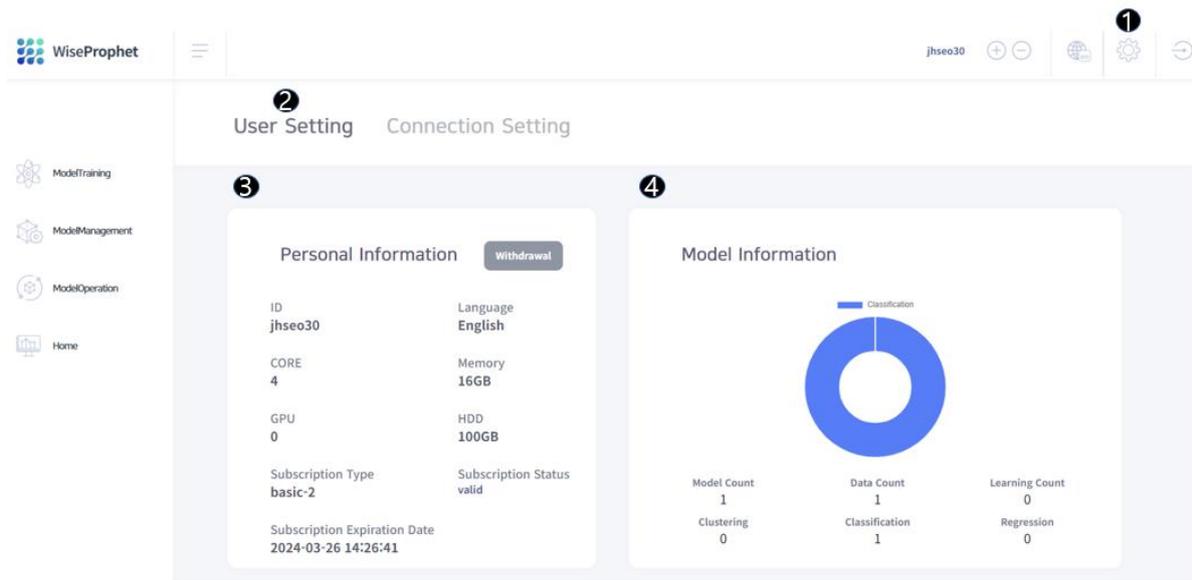
- This is a screen of Monitoring.

- 1) Model Count is the total number of models which you have saved until now. Data Count is the total number of data you have used. HDD Capacity means the data upload allowance.
- 2) The graph of Model Type displays the ratio of Regression, Clustering, and Classification. The one of Daily Model Count shows the number of model saved by day and type.
- 3) You can check the list you saved, composed of Status, Learning Progress, Dataset Type, and Dataset Name.

# Chapter 7. Setting

This chapter describes how to use WiseProphet for user setting.

## 7.1. User Setting



- 1) Clicking on the respective button navigates to the settings screen.
- 2) It displays the screen for user settings.
- 3) It shows the user's server information such as CPU, memory, etc., and allows for membership withdrawal.
- 4) It presents an overview of the models created by the user.

## 7.2. Connection Setting

Connect Name	DB Name	Connect Type	IP	User	Port	DB Type	Owner (Oracle)	Description	Registration Date
wise_demo	wise_demo	db	15.164.216.49	root	3306	mysql		데모	2024-02-23 17:01:54

- 1) It displays the screen for connection settings.
- 2) Users can add database (DB) connection information to be used for tasks such as model training.
- 3) It shows the list of added DB connections.
- 4) Users can modify or delete the DB connection information.

## Chapter 8. FAQ

Q. A case that an error occurs when I upload a file

A. In this case, the file you uploaded is unsupported, or it is an encoding problem. You can only upload csv files. Excel file extension should be in csv format. You should set the encoding as UTF-8.

Q. At the phase of Data Exploration, an error message says “Data should use at least 2 variables.”

A. When at least 2 variables were used, you could generate the analysis model by calculating the effect by variable.

Q. At the phase of Model Management, the model does not run after inputting new data.

A. It does not match the column of data used in generating model.

Q. An error message says “Calling url was invalid or server shut down”

A. You need to refresh the page or run server again.

Q. An error message says “The categorical variable whose unique value is more than 100 cannot be used.”

A. In the case of Categorical variable, column increases by number of unique values because dummy variable is created with One-Hot Encoding. Too many unique values result in too many columns and features for model generation. Because of poor model performance, it is excluded.