

# Knowledge-Based AI Framework for Smart Factory

Ver. 1.3

User's Manual

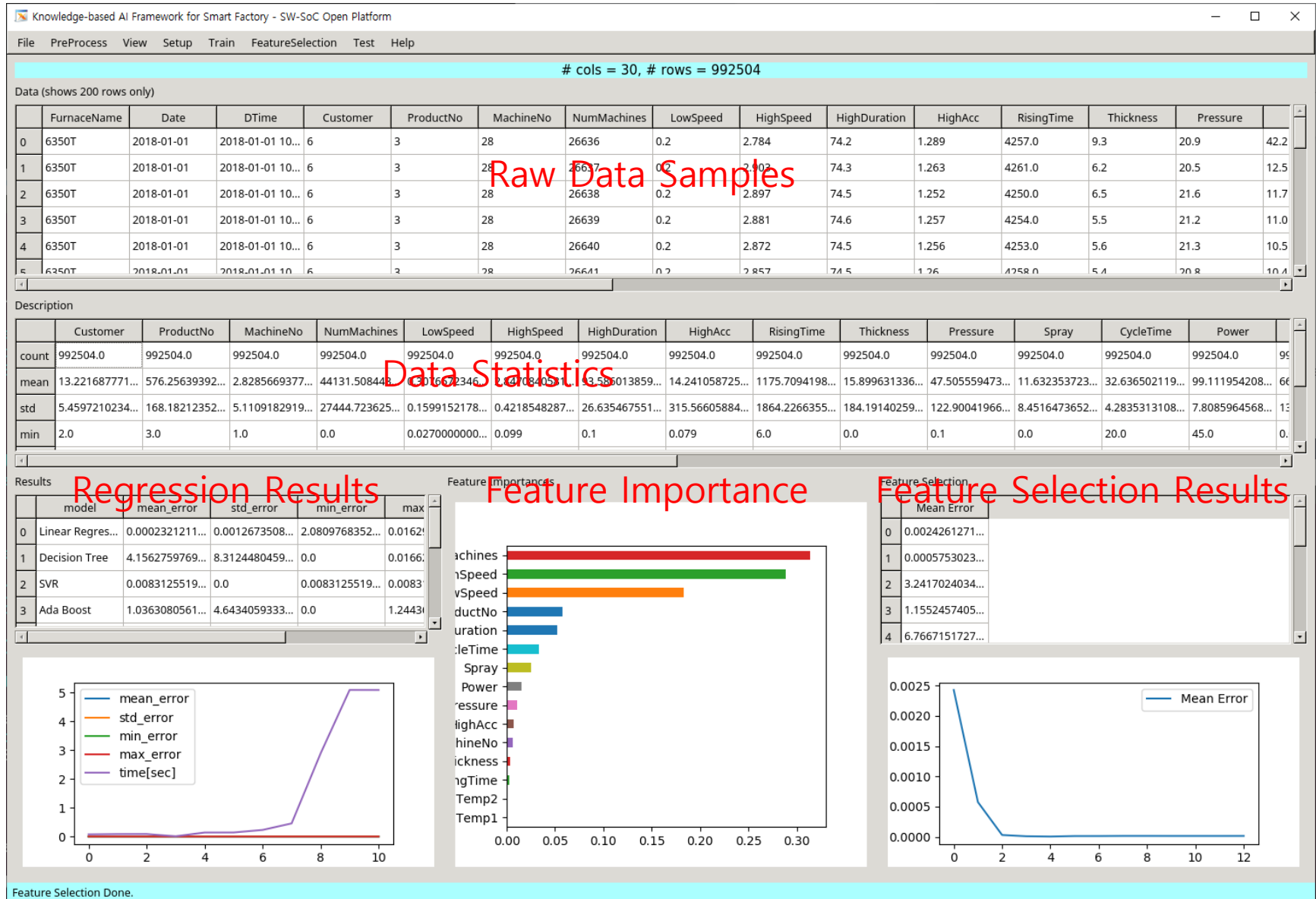
2019. 9. 30

서울SW-SoC융합R&BD센터  
김원종

제4차 산업혁명을 선도하는  
ICT Innovator



# Knowledge-Based AI Framework



# Major Functions



❖ Forecasting will be supported from ver. 1.5

- ❖ Supports **Data View** in various Graphs
- ❖ Supports 11 Regression, 10 Classification, Forecasting Models
- ❖ Supports Multi-output Regressions
- ❖ Supports Feature Selection
- ❖ Supports Feature Extraction
- ❖ Support Hyper Parameter Optimization (AutoML) – Alpha version

No	Regression	Classification
1	Linear Regression	AdaBoost
2	LASSO	Decision Tree
3	Elastic Net	Gaussian Process
4	Ada Boost	Linear SVM
5	SVR	Naive Bayes
6	Gradient Boosting	Nearest Neighbors
7	KNN	Neural Net MLP
8	Random Forest	Random Forest
9	Extra Trees	RBF SVM
10	Decision Tree	QDA
11	XGB	

❖ Forecasting models is being developed.

- ❖ LASSO: Least Absolute Shrinkage and Selection Operator
- ❖ QDA: Quadratic Discriminant Analysis

# Release Notes



- ❖ Ver. 1.0
  - First release with 11 Regression Models
  - Used PyQt ver. 4
- ❖ Ver. 1.1
  - Used PyQt ver. 5
- ❖ Ver. 1.2
  - Added 10 Classification Models
  - Added Classification Data: iris, wine, breast cancer -> make <iris | wine | br>
- ❖ Ver. 1.2.1
  - Added Scaling before Regression/Classification
- ❖ Ver. 1.3
  - Added Hyper Parameter Optimization (AutoML) – alpha version
- ❖ Future Releases
  - Forecasting model: Moving average, ARIMA
  - Export Best Model to Deploy

# Requirements



## ❖ Python3, PyQt5

- ❖ Libraries: numpy, pandas, matplotlib, scipy, sklearn, xgboost, tpot

```
# PyQt5 for python3  
sudo apt install python-qwt5-qt4    # QT Widget 5 for PyQt4  
sudo apt install python3-tk tk-dev  
sudo apt install qt5-default pyqt5-dev pyqt5-dev-tools qttools5-dev-tools  
# Libraries for python3  
pip3 install numpy, pandas, matplotlib, scipy sklearn xgboost tpot
```

## ❖ Not recommended!

```
# PyQt4 for python2  
sudo apt install python-qt4 qt4-default qt4-dev-tools pyqt4-dev-tools  
# PyQt4 for python3  
sudo apt install python3-pyqt4  
sudo apt install qt4-designer    # QT-Designer for PyQt4  
Pip install numpy, pandas, matplotlib, scipy sklearn xgboost tpot
```

# Basic Setup



- ❖ Configuration File <Not in Demo version, Only for Development Mode>
  - config\_gui\_pm.py
    - ✓ PYQT\_VER = 5      # Python Qt Version 5 사용
    - ✓ PYQT\_VER = 4      # Python Qt Version 4 사용
- ❖ Environment Setup
  - make req    # Python Qt version 4/5 and Python Libraries



## ❖ Command line Usage – use pm.py for Demo Version

➤ `python3 pm.py [-d <data file>] [-i <input columns file>] [-o <output columns file>]`

## ❖ Pre-setup scripts in Makefile

➤ `make gui` # Start Empty GUI  
✓ `python3 gui_pm.py`

### ➤ **make guid**

✓ # Start GUI with Regression Model Data file set, Input/Output Columns loaded  
✓ `python3 gui_pm.py -i Data/SeA/inputs.csv -o Data/SeA/outputs.csv -d Data/SeA/params_merged_Toshiba_2018.csv`

### ➤ **make guid5**

✓ # Start GUI with Regression Model Data file, Input5/Output Columns loaded

### ➤ **make <iris | wine | br>**

✓ Start GUI with **classification model** Data file set, Input/Output Columns loaded  
✓ Iris, wine and breast cancer data

# File Formats



## ❖ Data file

- csv file or **Excel (xls,xlsx)** file
- columns: input features or output
- rows: instances

## ❖ Input

- csv file
- Input feature names

For make guid

ProductNo,MachineNo,NumMachines,LowSpeed,HighSpeed,HighDuration,HighAcc,RisingTime,Thickness,Pressure,Spray,CycleTime,Power,Temp1,Temp2

## ❖ Output

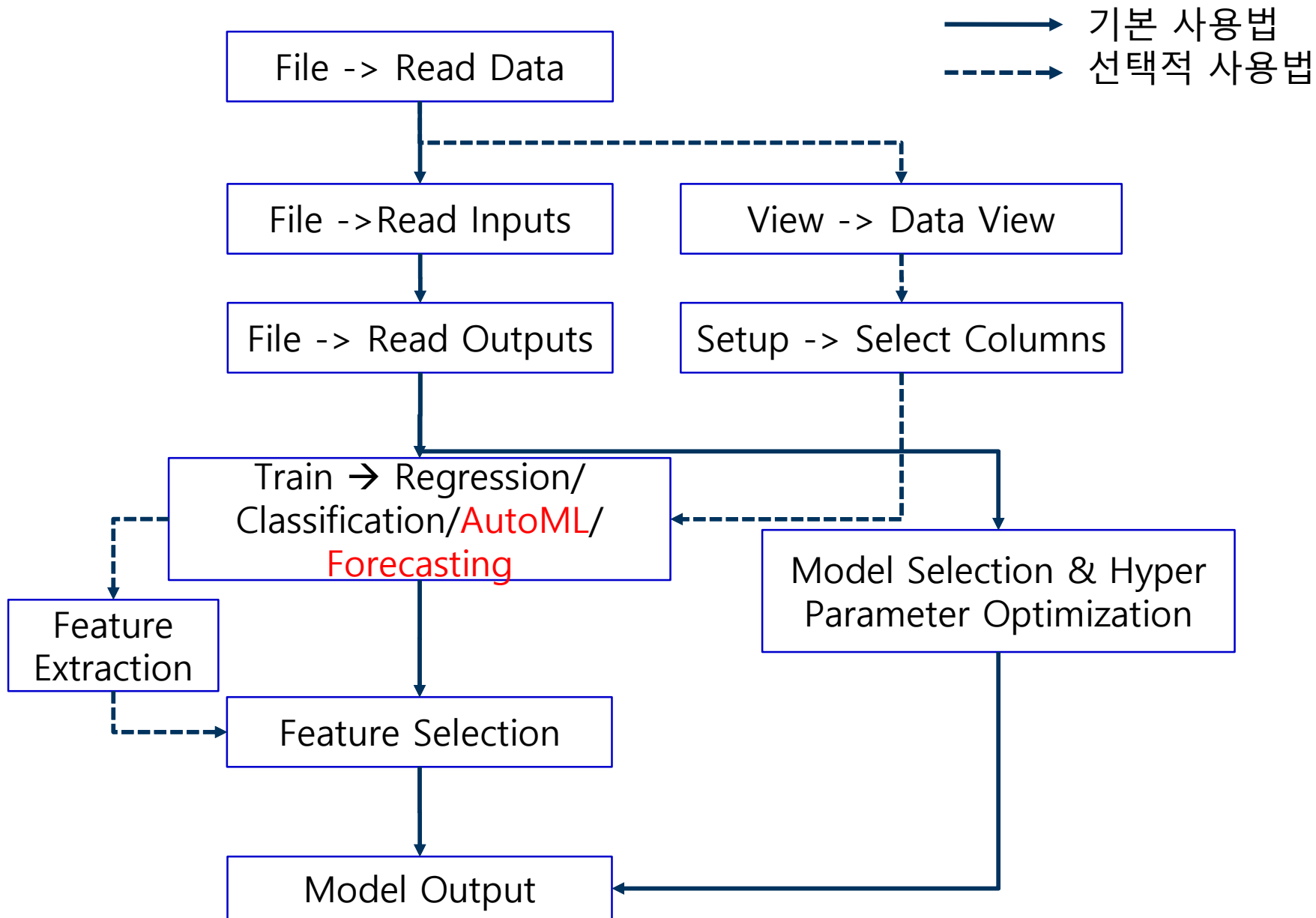
- csv file
- Output feature names

For make guid

"1","2","3","4","5","6","7","8","9","10","11"



# Basic Usage Flow



# Data View



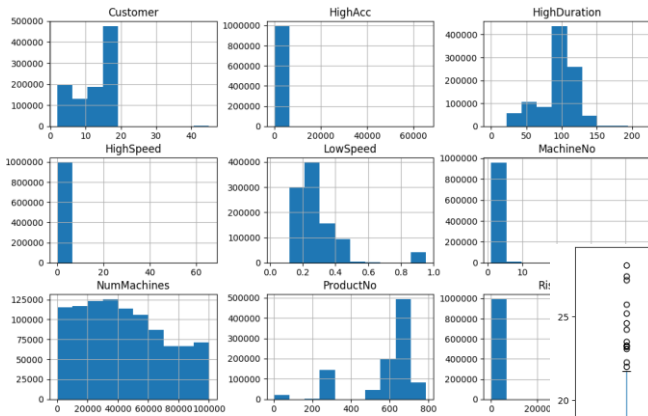
## ❖ Data View: View selected column data

- Select Plot Type: Data, Histogram, Density, box Correlation, Scatter
- Select/Unselect columns and click **Apply** to view data

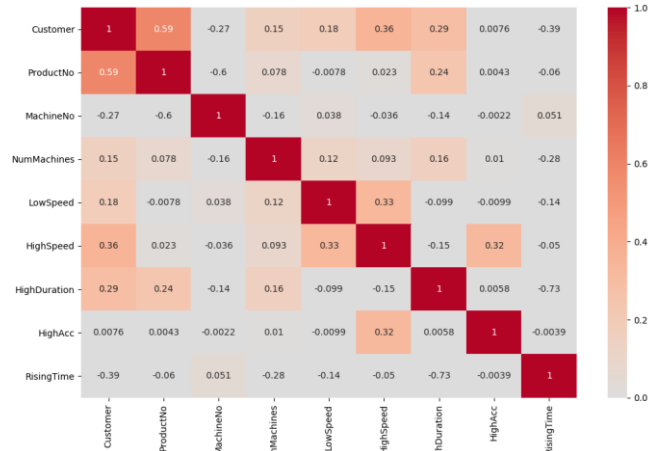


<Data View>

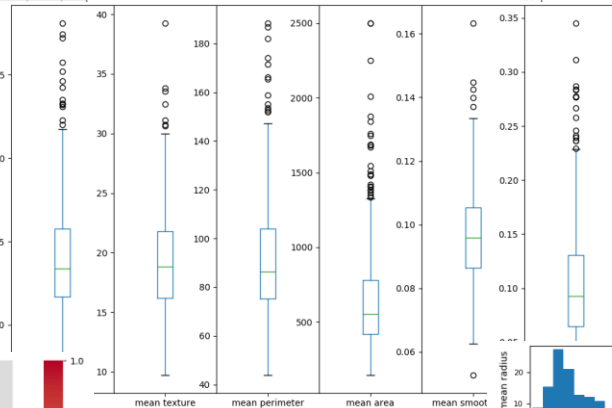
❖ Density, scatter plot may take time for big data



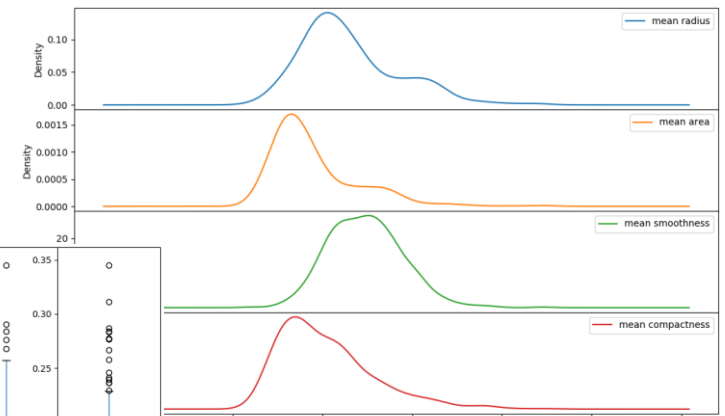
<Histogram>



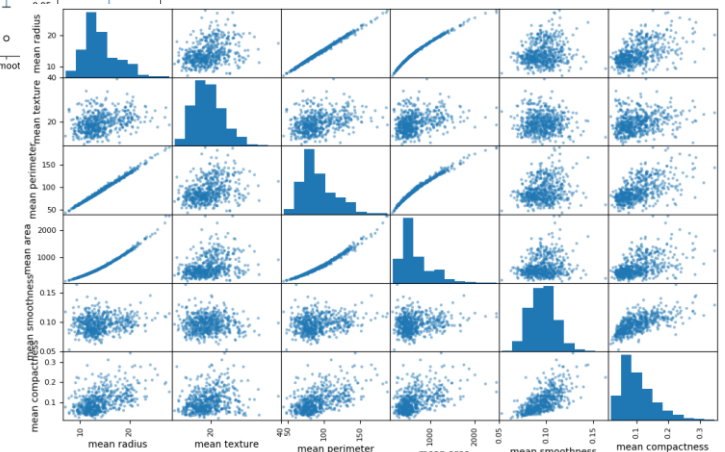
<Correlation>



<Box>



<Density>



<Scatter matrix>

# Setup -> Select Columns



- ❖ Choose each column as Ignore, **Input**, or Output
  - Input means use them as **input features**
  - Output means use them as **target**
  - Then, click OK

A screenshot of the 'Select Columns' dialog box. The dialog has a title bar 'Select Columns' and a close button. Below the title bar is a section 'Choose Columns for Input, Output, or Ignore'. It contains a list of columns with dropdown menus for each. The columns are: FurnaceName, Date, DTime, Customer, ProductNo, MachineNo, NumMachines, LowSpeed, HighSpeed, HighDuration, HighAcc, RisingTime, Thickness, Pressure, Spray, CycleTime, Power, Temp1, Temp2, and a list of numbers 1 through 11. The dropdowns for ProductNo, MachineNo, NumMachines, LowSpeed, HighSpeed, HighDuration, HighAcc, RisingTime, Thickness, Pressure, Spray, CycleTime, Power, Temp1, and Temp2 are set to 'Input'. The dropdowns for FurnaceName, Date, DTime, Customer, and the numbers 1 through 11 are set to 'Output'. At the bottom are 'Cancel' and 'OK' buttons.

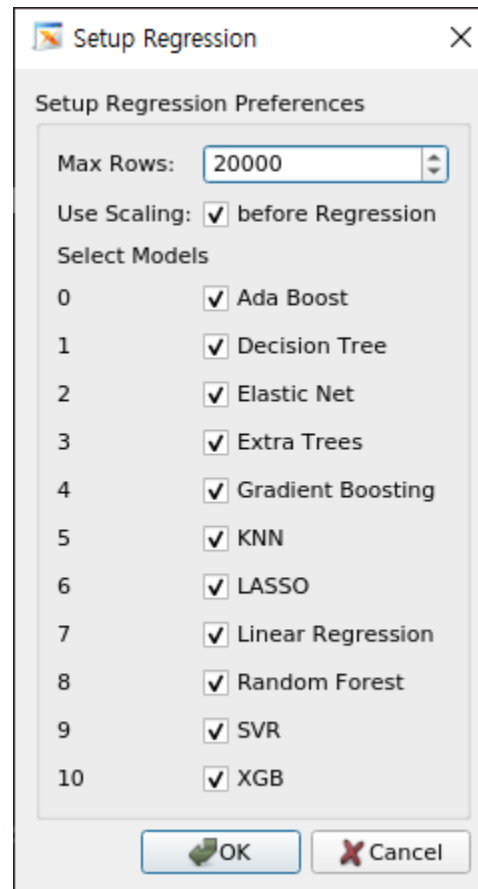
Column Name	Selected Option
FurnaceName	Ignore
Date	Ignore
DTime	Ignore
Customer	Ignore
ProductNo	Input
MachineNo	Input
NumMachines	Input
LowSpeed	Input
HighSpeed	Input
HighDuration	Input
HighAcc	Input
RisingTime	Input
Thickness	Input
Pressure	Input
Spray	Input
CycleTime	Input
Power	Input
Temp1	Input
Temp2	Input
1	Output
2	Output
3	Output
4	Output
5	Output
6	Output
7	Output
8	Output
9	Output
10	Output
11	Output

# Train -> Regression



## ❖ Options for Regression

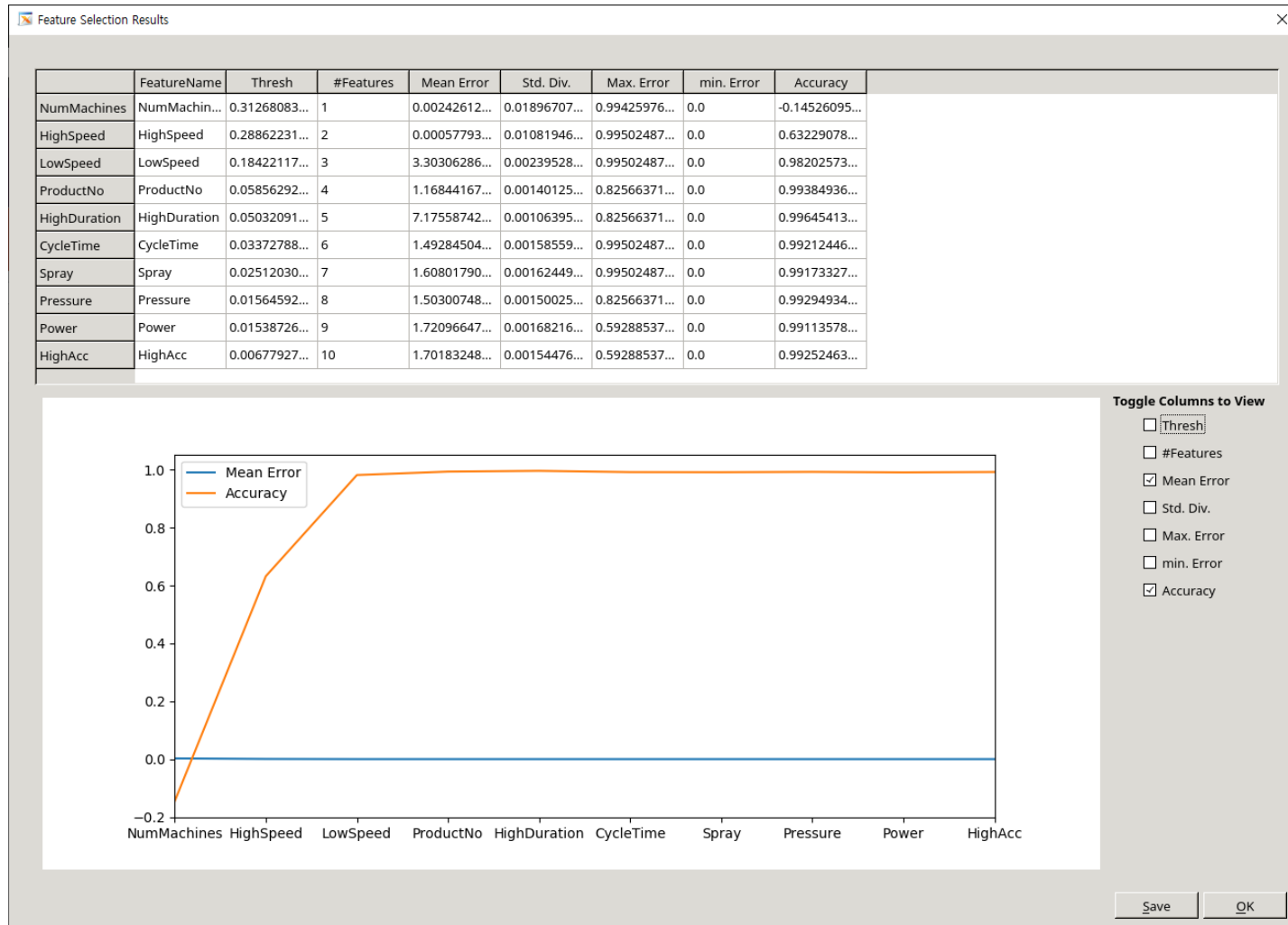
- Select Max Rows: 0 means all rows
- Check whether to use Scaling before Regression (New in ver. 1.3)
- Select/Unselect Models to test
- Then, click OK



# Train -> Regression



- ❖ Regression Results: Select columns to view output graphs
  - Usually **Mean error** and **accuracy** are **important**
  - **Save Results table to csv file**

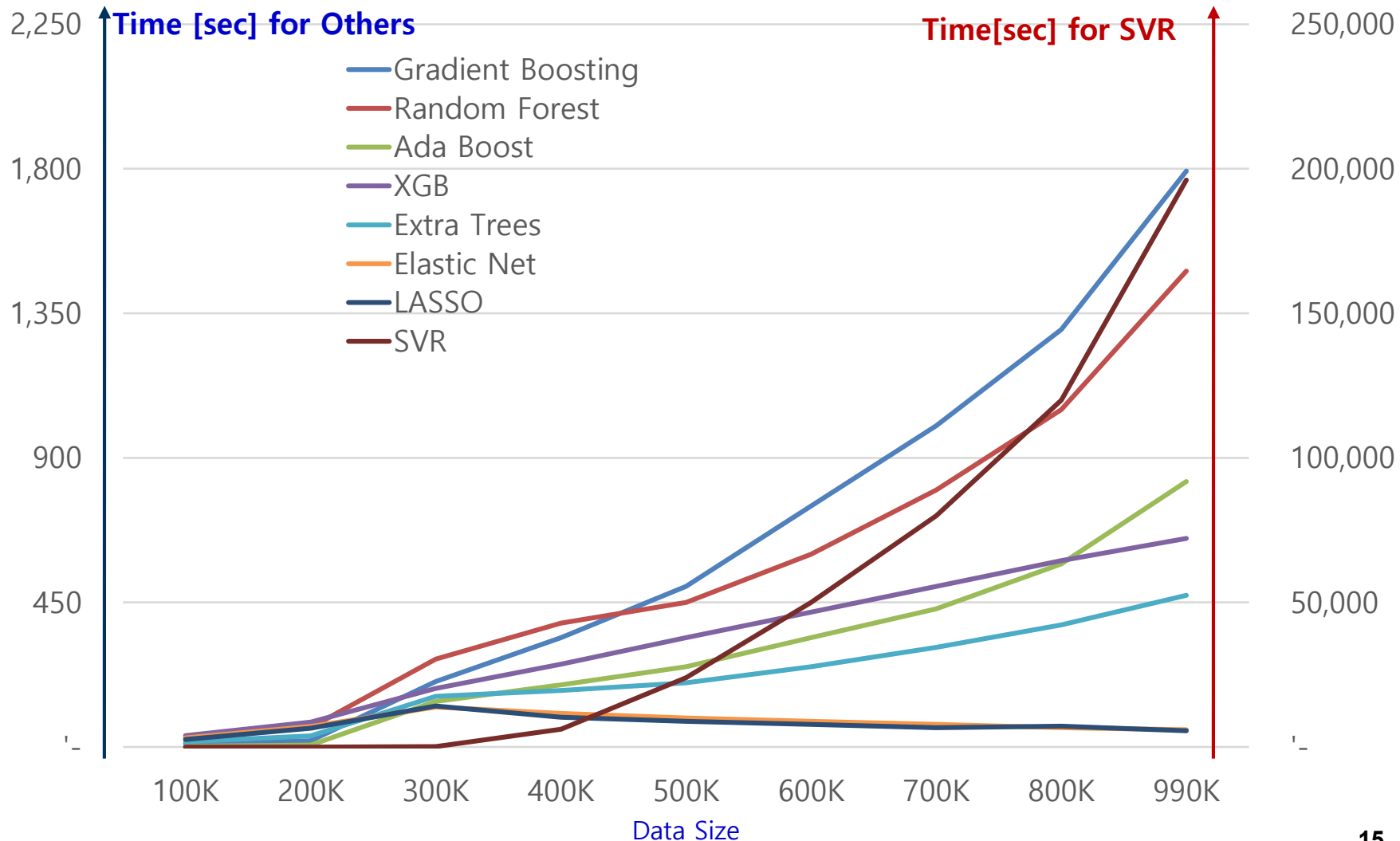


# Train -> Regression



❖ Regression Results: **Execution Time [sec]** (Linux ubuntu 16.04, 4 GHz)

➤ **SVR** is **not recommended** for data size **over 300K**



# Train -> Classification

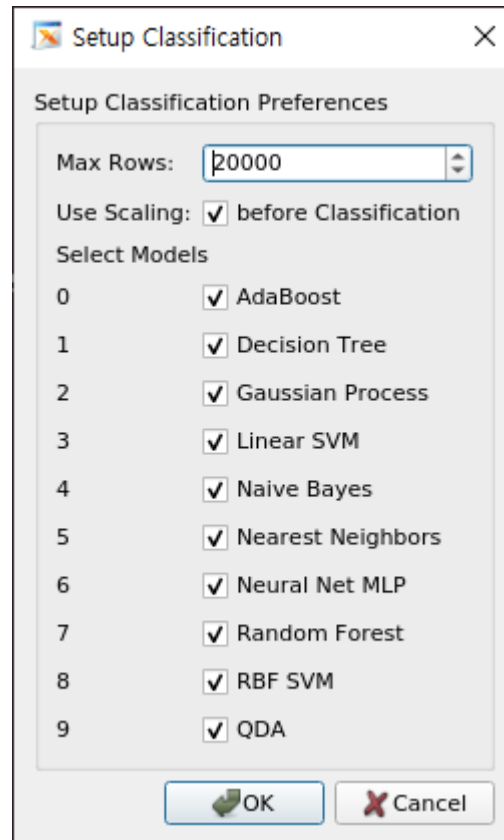


New in ver. 1.2

❖ \$ make br # for breast cancer data

❖ Options for Classification

- Select Max Rows: 0 means all rows
- Check whether to use Scaling before Classification (New in ver. 1.3)
- Select/Unselect Models to test
- Then, click OK

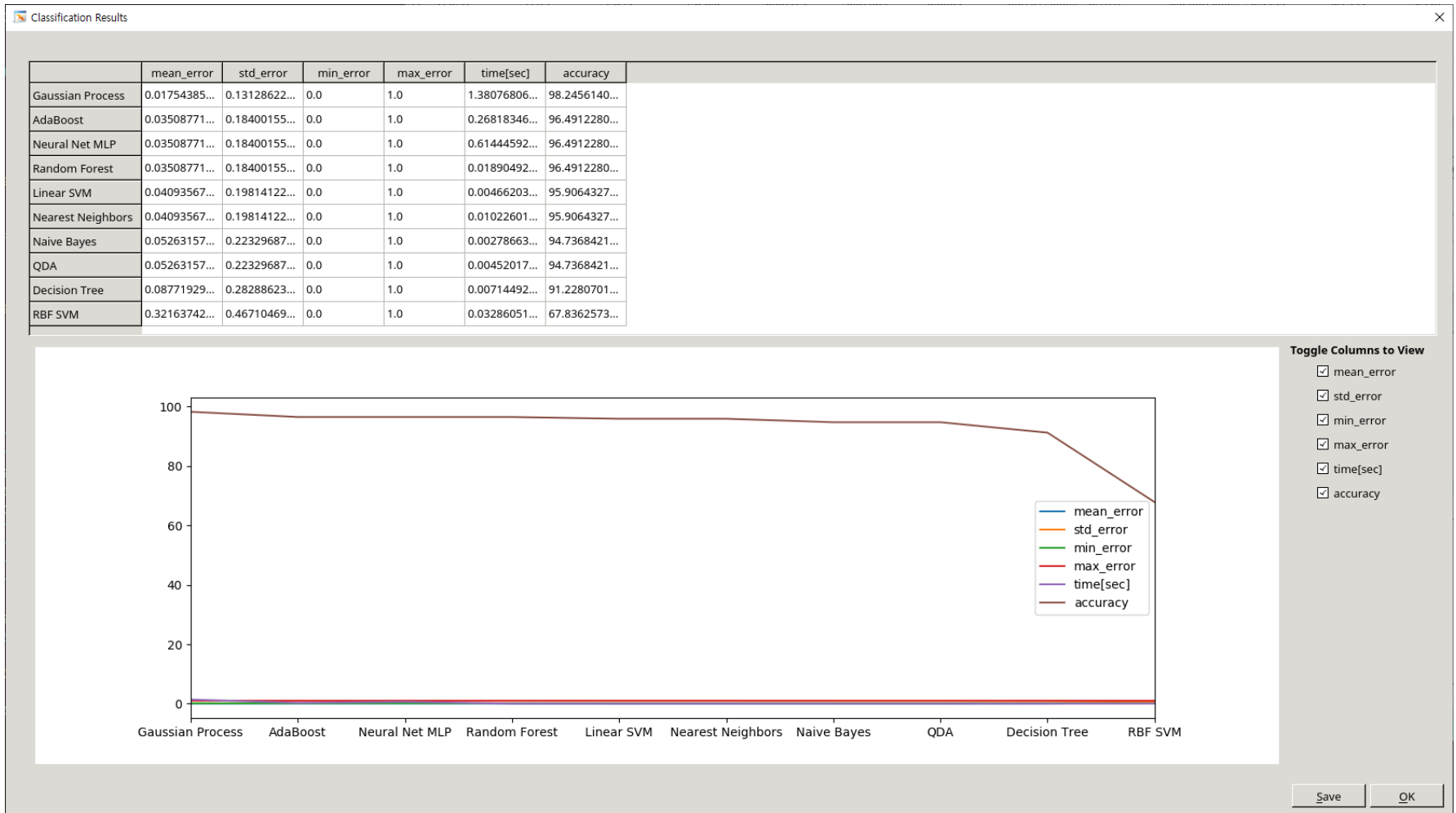




# Train -> Classification



- ❖ **Classification Results:** Select columns to view output graphs
  - Usually **accuracy** is **important** for classification
  - **Save Results table to csv file**

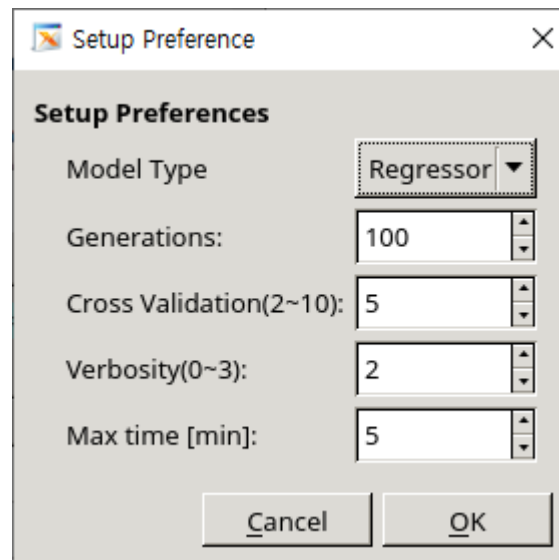


# Train -> AutoML

ETR

New in ver. 1.3  
Not complete yet!

- ❖ \$ make guid1
- ❖ Can be used to select model and optimize hyper parameters concurrently
- ❖ Options for AutoML
  - Select Regressor/Classifier
  - Select Generations: Number of iterations to the run pipeline optimization process. Must be a positive number.
  - Select Cross Validation: <2 means no cross validation
  - Select Max time [min]: How many minutes to use to optimize the pipeline. 0 means no limit.



# Train -> Forecasting



New in ver. 1.5

Not implemented yet!

- ❖ \$ make ss
- ❖ Forecasting Type
  - Daily, Monthly, Seasonal, Yearly
  - Select Starting Date to Estimate
  - Select steps for Forecasting

# Feature -> Feature Extraction



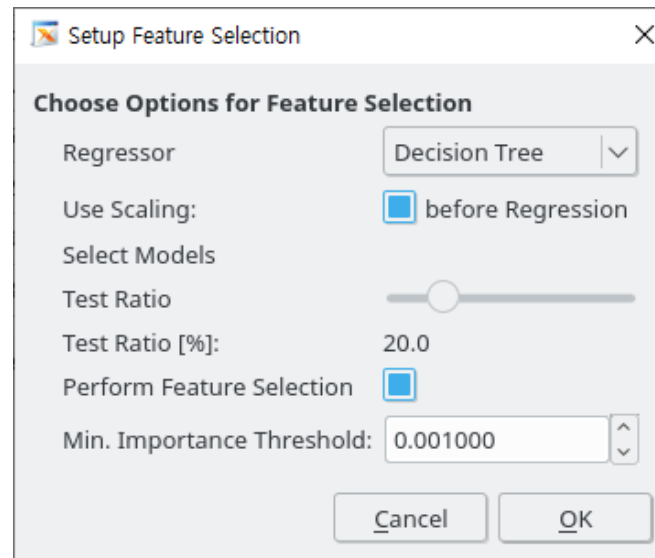
## ❖ Feature Extraction

- Can be used when input features are not enough to get good results
- Increase the number of features using given input features
- Performs Polynomial Feature Extraction
- Replace original features to  $x_1, x_2, \dots, x_n$
- Generate new features such as  $x_1^2, x_1 \cdot x_2$ , etc.

# Feature -> Feature Selection



- ❖ Can be used **to identify importance of input features** to get better results for the target
- ❖ Options for Feature Selection
  - Select Regressor model
  - Select if use **Scaling (new in Ver. 1.3)**
  - Select Test Ratio (%): Default is 20% for validation
  - Check/Uncheck Perform Feature Selection
    - ✓ **Uncheck** means that just show feature importance graph
  - Choose Minimum Importance Threshold value
  - Then, Click OK



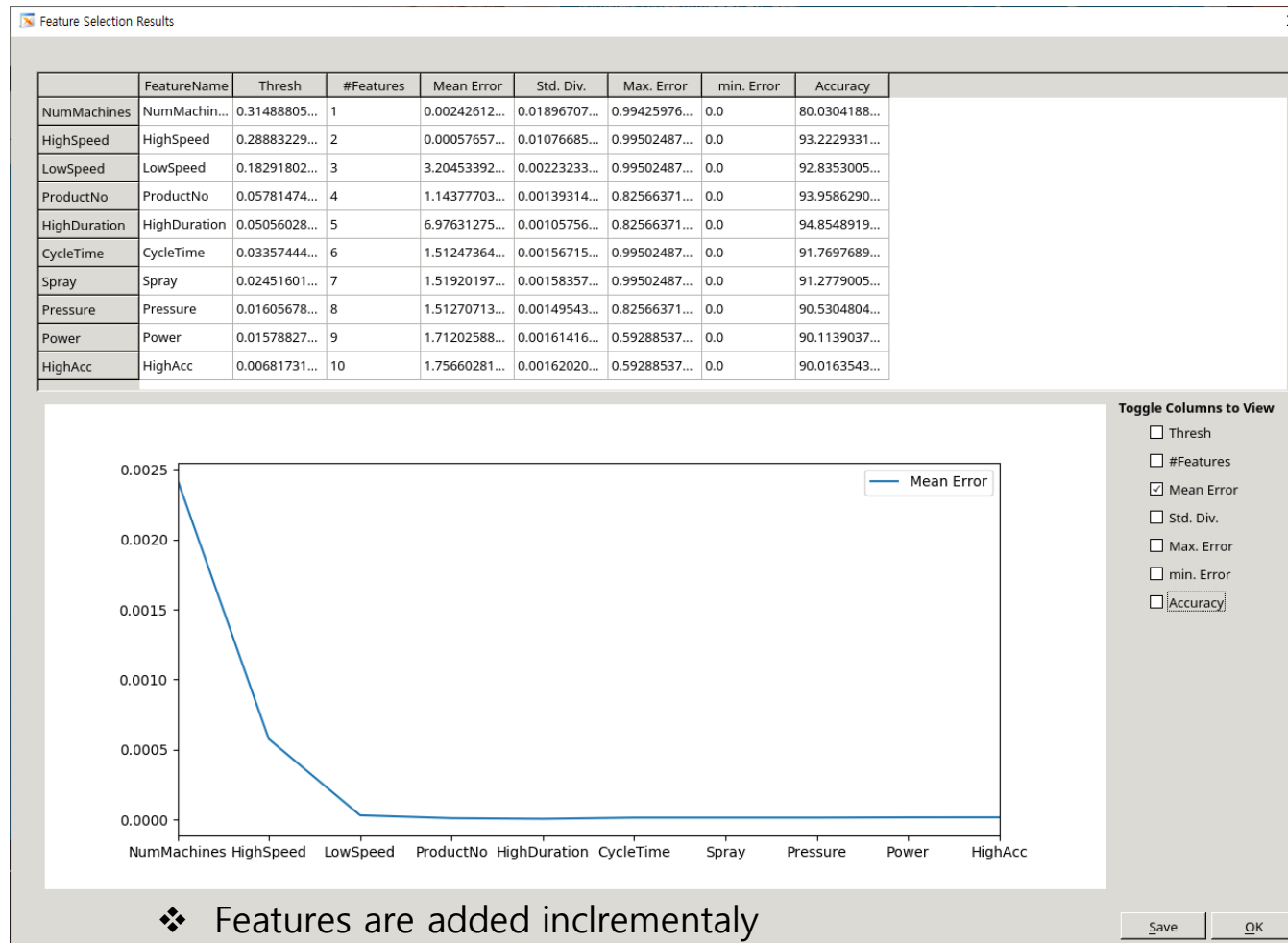
- ❖ **For classification data, Regression Models are used in feature selection!**  
(Future release will use Classification Models.)

# Feature -> Feature Selection



## ❖ Feature Selection Results

- When **Perform Feature Selection** is Checked
- Check/Uncheck columns to view
- **Save** result table to csv file, if needed



# Feature -> Feature Selection



- ❖ Feature Selection Results: Find the best number of input features
  - Using **only 5 features** (NumMachines, HighSpeed, LowSpeed, ProductNo, HighDuration) have best accuracy

