

# AI Platform Trends: Challenges or Services?



Jeongcheol Lee ( J. C. Lee )

jclee@kisti.re.kr



SOMETHING  
USEFUL  
KNOWLEDGE



4.8 / 5

Job Score

4.2 / 5

Job Satisfaction

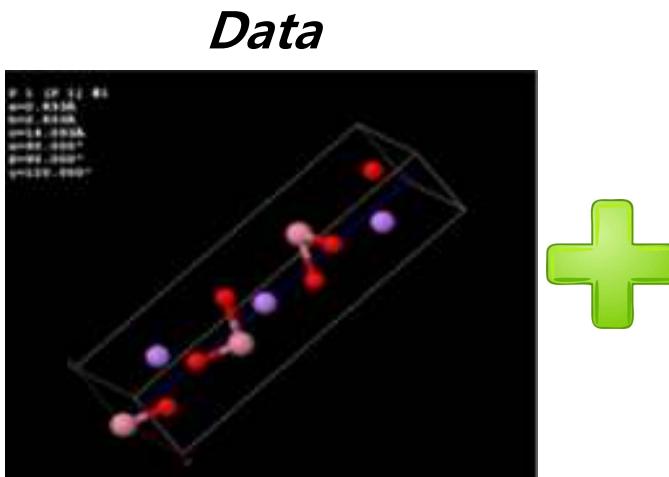
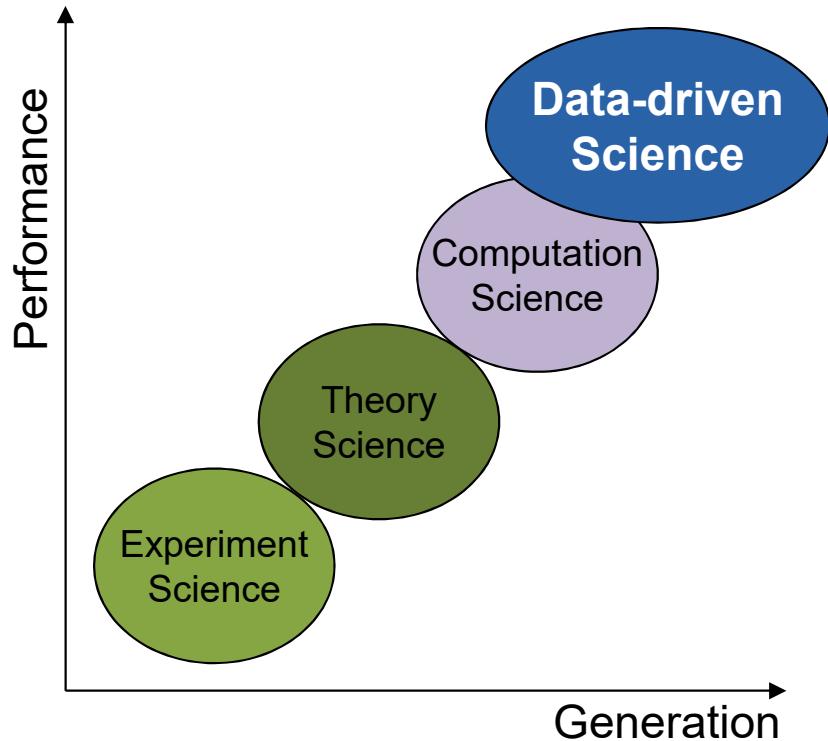
\$110,000

Median Base Salary

4,524

Job Openings

[View Jobs](#)

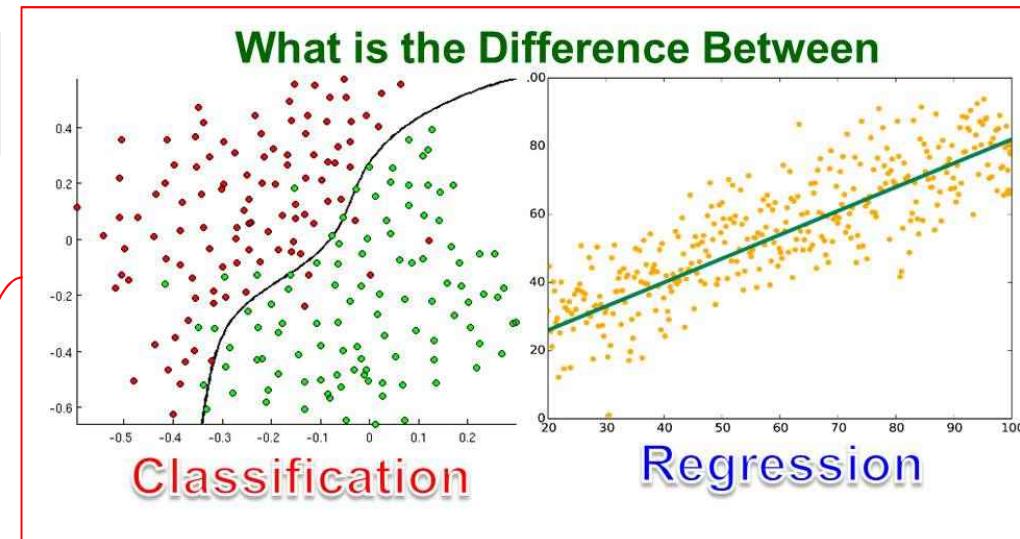
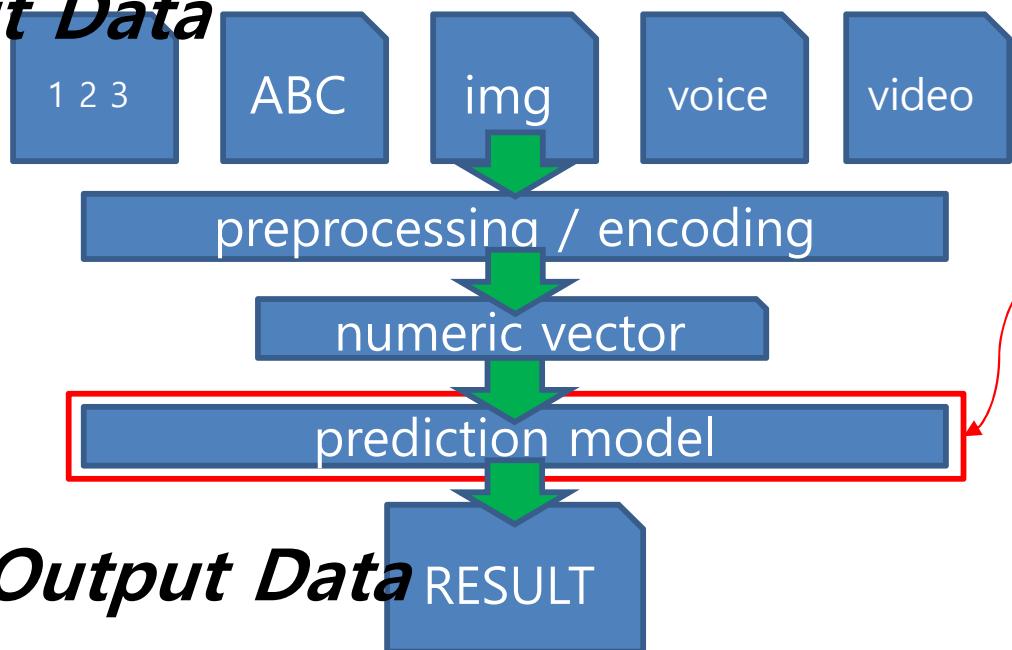


## ➤ Benefits

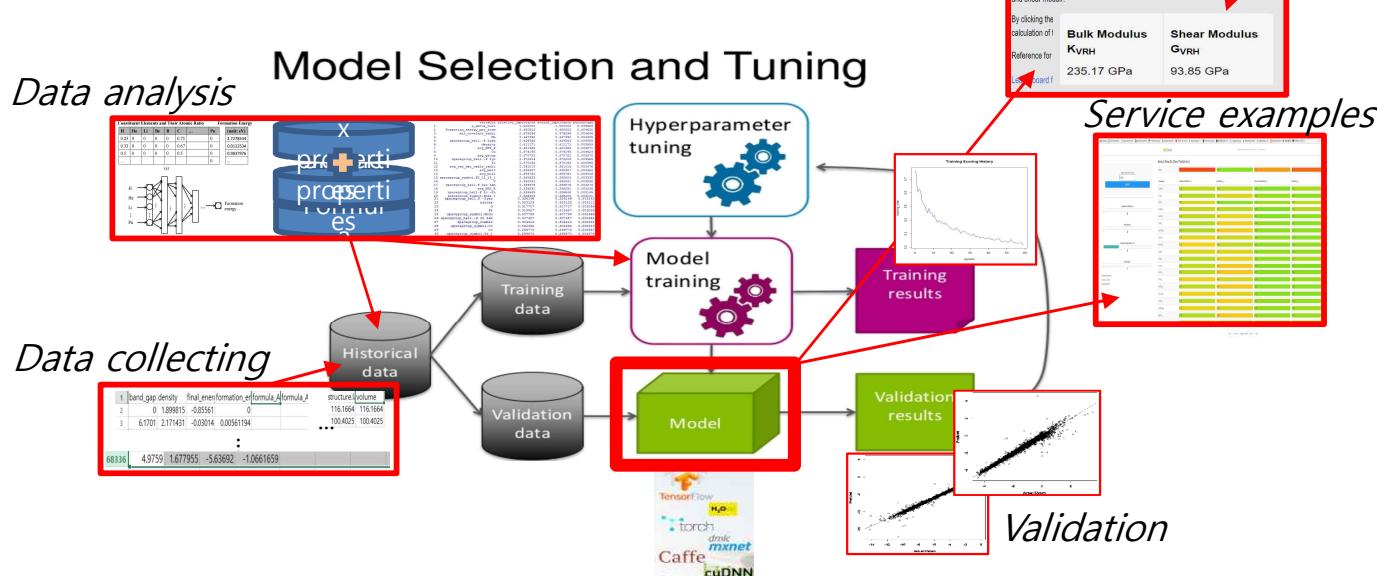
- Eliminating redundant calculations that cost much time and expenses
- Mining a meaningful knowledge based on computation data

# Fundamental Purpose

## *Input Data*



## *Output Data* RESULT



# Amazon SageMaker

기계 학습 모델을 대규모로 구축, 교육 및 배포

[Amazon SageMaker 시작하기](#)



Amazon SageMaker는 개발자 및 데이터 과학자가 어떠한 규모의 기계 학습 모델이든 쉽고 빠르게 구축, 교육 및 배포할 수 있도록 지원하는 완전관리형 플랫폼입니다. Amazon SageMaker는 기계 학습을 사용하려는 개발자가 일반적으로 빠른 속도를 내지 못하게 하는 모든 장벽을 제거합니다.

[Google Cloud](#)

Why Google Products Solutions Pricing Security Documentation Customers Partners Support Marketplace

AI & Machine Learning Products

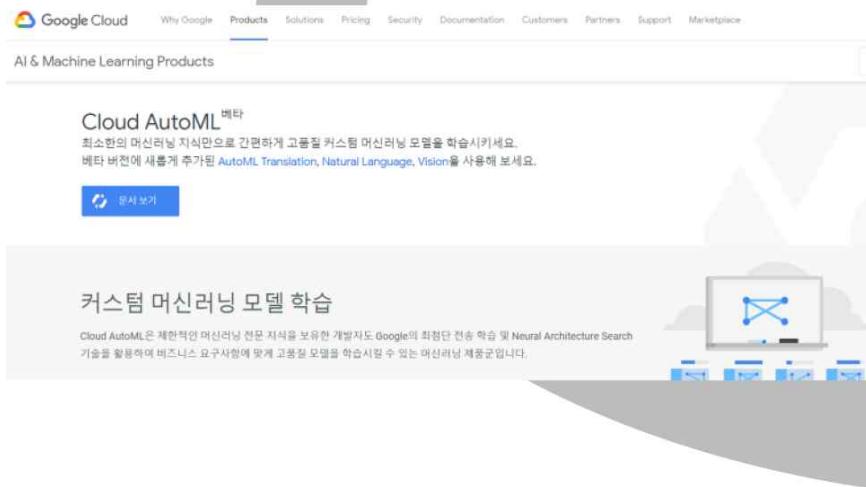
**Cloud AutoML** 베타

최소한의 머신러닝 지식으로 간편하게 고급화 커스텀 머신러닝 모델을 학습시키세요.  
비타 버전에 새롭게 추가된 AutoML Translation, Natural Language, Vision을 사용해 보세요.

[문서 보기](#)

**커스텀 머신러닝 모델 학습**

Cloud AutoML은 제한적인 머신러닝 전문 지식을 보유한 개발자도 Google의 최첨단 전송 학습 및 Neural Architecture Search 기술을 활용하여 베즈스 요구사항에 맞게 고급화 모델을 학습시킬 수 있는 머신러닝 제품군입니다.




New!

Azure Machine Learning service **PREVIEW**

Try it today!

[Sign In](#)

No Azure subscription? No credit card? No problem! Choose anonymous Guest Access, or sign in with your work or school account, or a Microsoft account.

Not an Azure ML user? [Sign up here](#)

Pricing & FAQ

By using this free version, you agree to be bound by the Microsoft Azure Mobility Terms of Use.

kaggle Search kaggle

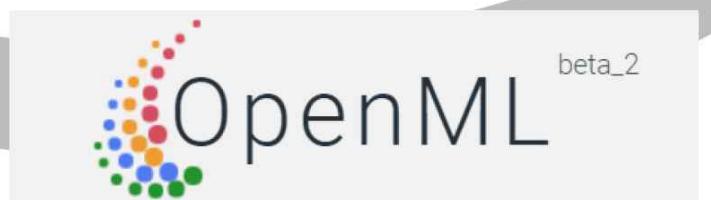
Competitions Datasets Kernels Discussion Learn ... Sign In

Competitions

General InClass Sort by Grouped All Categories Search competitions

14 Active Competitions

	<b>Two Sigma: Using News to Predict Stock Movements</b> Use news analytics to predict stock price performance Featured · 2 months to go · news agencies, time series, finance, money	\$100,000 1,153 teams
	<b>Airbus Ship Detection Challenge</b> Find ships on satellite images as quickly as possible Featured · 19 days to go · image data, object detection, object segmentation	\$60,000 464 teams





*"You can make a model by yourself, we can help."*

# Service-based

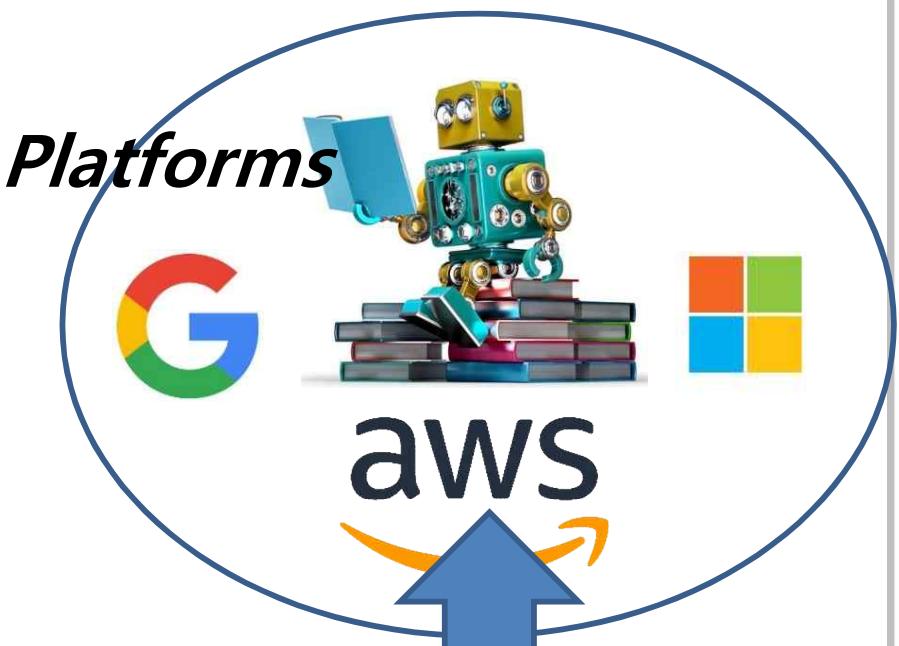
vs.

# Challenge-based

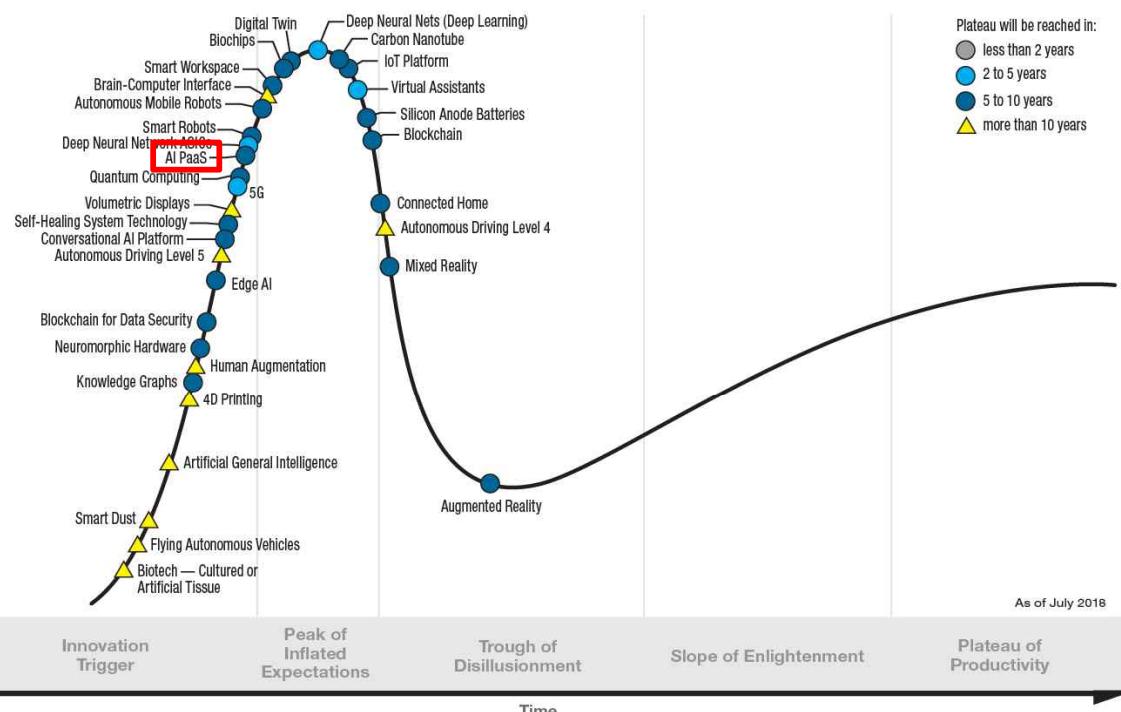
*"Prize money to make a model by experts, we can help."*

- Machine Learning as a Service (MLaaS)

= AI Platform as a Service !



Hype Cycle for Emerging Technologies, 2018



[gartner.com/SmarterWithGartner](http://gartner.com/SmarterWithGartner)

Source: Gartner (August 2018)  
© 2018 Gartner, Inc. and/or its affiliates. All rights reserved.

Gartner®





# Artificial Intelligence as a Service (AlaaS) Market 2019 Application, Solutions, Developments Status, Technology & Analysis, Segmentation, Trends, Business Opportunities 2023

Published: Apr 18, 2019 8:54 a.m. ET



[MARKETSANDMARKETS](#)

[KNOWLEDGE STORE](#)

[MARKET RESEARCH REPORTS](#)

[PRACTICES](#)

[CONSULTING](#)

[EVENTS](#)

[360 QUADRANTS](#)

## Artificial Intelligence as a Service Market

[HOME](#) > [PRESS RELEASES](#) > Artificial Intelligence as a Service Market worth 10.88 Billion USD by 2023

### Artificial Intelligence as a Service Market worth 10.88 Billion USD by 2023

[VIEW FULL TABLE OF CONTENTS](#)

[METHODOLOGY](#)

[DOWNLOAD PDF](#)

The report "Artificial Intelligence as a Service Market by Service Type (Software Tools and Services), Technology (Machine Learning and Deep Learning, and Natural Language Processing), Organization Size, Vertical, and Region - Global Forecast 2023". The AI as a service market size is expected to grow from USD 1.52 Billion in 2018 to USD 10.88 Billion by 2023, at a Compound Annual Growth Rate (CAGR) of 48.2% during the forecast period.



Machine Learning as a Service Market

[Machine Learning as a Service \(MLaaS\) Market Size - Segmented by Application \(Advertising & Marketing, Predictive Maintenance, Automated Network Management\), Organization Size \(SMB and Large Enterprises\), End User, and Region - Growth, Trends and Forecasts \(2019 - 2024\)](#)



The global MLaaS market is expected to register a CAGR of about 43.46% during 2018-2023 (the forecast period), to reach a value of USD 8.315 billion, by 2023, from USD 0.932 billion, as of 2017. The scope of this report is limited to the solutions that are offered by the major market players including service providers. The regions, considered in the scope of this report, include North America, Europe, and 'others'. The study highlights the various applications of machine learning, such as predictive maintenance, risk analytics, and fraud detection, among various others.



## ▪ Amazon

- the largest provider of cloud-based services and provides a platform for machine learning platform services such as 'Amazon ML' and 'Amazon SageMaker'
- Amazon ML provides automated solutions for data classification, prediction, and clustering
- Amazon SageMaker provides an automated machine learning model builder and parameter optimization tools

## ▪ Microsoft

- Azure ML Studio provides a GUI-based interface for developers to create, test, and deploy machine learning algorithms
- Should learn how to use the MS Azure interface including workflows
- New platform allows user to do model engineering without providing built-in method.



## ▪ Google

- Google Prediction API was deprecated after April 2018, instead of, provides more powerful service AutoML
- AutoML enables people without data science expertise to learn high-quality custom data models in an automated fashion
- AutoML Vision will be available as the first product to implement a custom image recognition model
- Long-term delivery of services for all other major areas of AI

## ▪ IBM

- provides collaborative development environment that supports both GUI interface and code level editable notebook.
- collaborates with H2O driverless AI to provide very simple and fully-automated solutions for making ML models.
- Watson ML provides various model learning and distribution interfaces

*Has the ML expert group (including data scientist) no longer needed?*



- **Kaggle**      [www.kaggle.com](http://www.kaggle.com)

- acquired by Google in 2017
- creates a model to compete with data and challenges presented by business, government agencies, organizations, research institutes, or individuals.
- establishing a reward system in which winners or prize money are delivered to winners
- the best way to follow up the recent techniques (know-hows) in an interested field

- **OpenML**      [www.openml.org](http://www.openml.org)

- Open platform for sharing, reconfiguring, and discussing machine learning results among data scientists
- Dynamic allocation of tasks such as idea formation, data collection, data mining, and reinterpretation of results

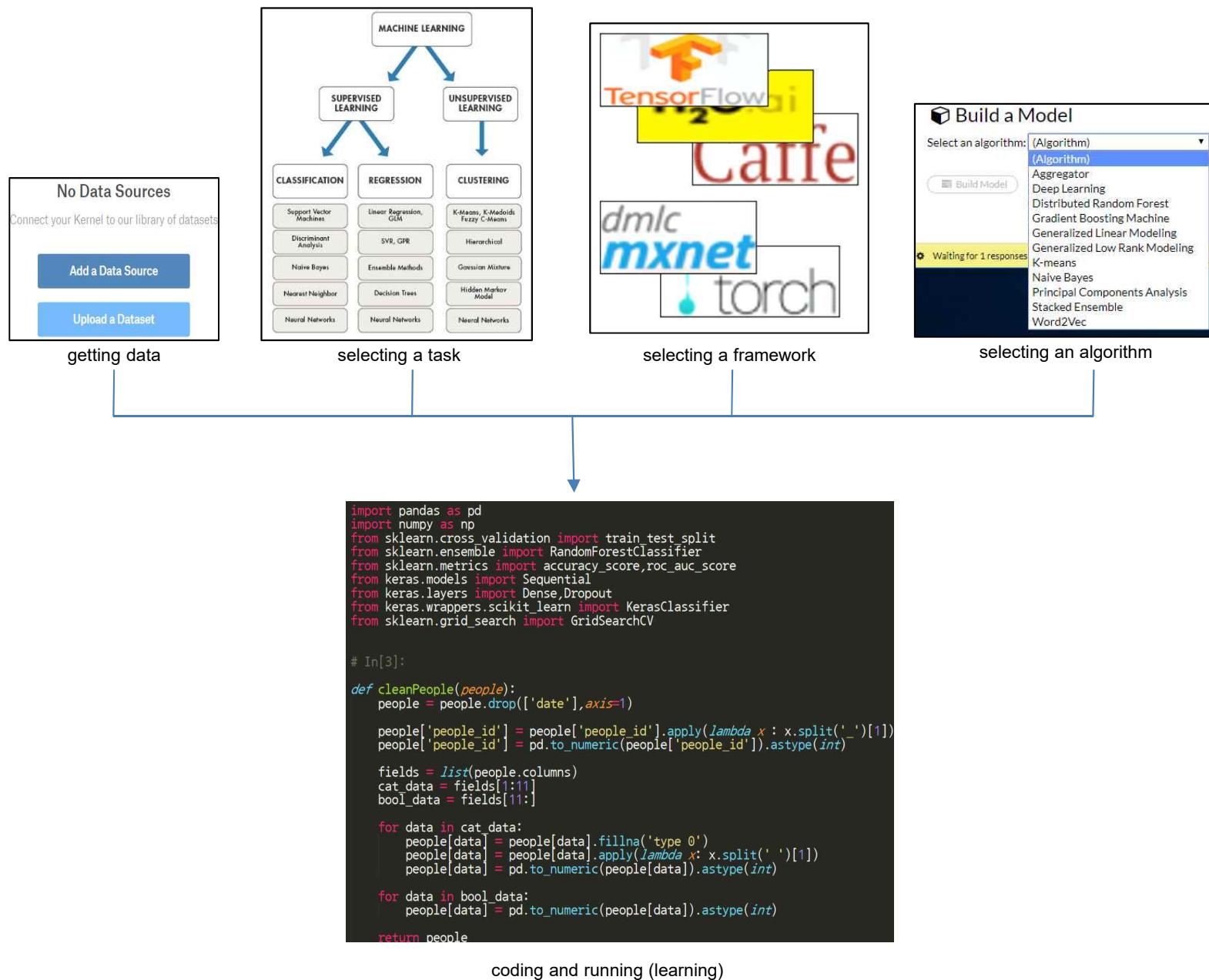
- **others:** DrivenData, Innocentive, ... , at cetera.

*I know there are many of know-hows, but how can I use them?*

The diagram features three cloud-like shapes above the chart. The first shape, labeled "Service-based", contains "Google Cloud AI Platform" and "MS Azure ML Studio". The second shape, labeled "Challenge-based", contains "Kaggle" and "Open ML". The third shape, labeled "DO ALL!", is empty.

	Google Cloud AI Platform	MS Azure ML Studio	Kaggle	Open ML	KISTI Data Platform
GUI based ML Modeling	✓	✓	✗	✗	✓
ML Data Preprocessing	✓	✓	✗	✗	✓
Deploy as a Web Service	✓	✓	✗	✗	✓
ML Programming Environment	✓	✓	✓	✓	✓
ML Program Sharing	✗	✗	✓	✓	✓
ML Data Sharing	✗	✗	✓	✓	✓
Specialized	Analysis		Sharing		Easy Analysis + Sharing
Purpose	Detailed Analysis + Deployment (from individual intermediate to expert)		Competition (experts group)		Education (from beginners to experts)

*Now, the KISTI Data Platform is available at the PRAGMA platform !*  
<https://www.edison.re.kr/web/pragma>



Home About Data Community Tool News Sample

Sign-in

ML Data Submission Edit Scheme GUI based parameter setting Preview Code

GUI based parameter setting

Please select your Kernel for Machine Learning R Python

Example for R

```
[ 3] > # we can use the print() function
> print("Hello World")
[1] "Hello World"
> # we can also represent in the output
> print("Hello World", quote = FALSE)
[1] Hello World
> # If there are more than 1 line, we can concatenate using paste0()
> print(paste0("How", "are", "you"))
[1] "How are you?"
```

Which task? Regression Classification

Preview for Regression Fit non-linear least squares

Fit non-linear least squares

```
## simulate some data
set.seed(20160227)
x<-rnorm(100)
y<-rnorm(100)
yc<-c(rnorm(1,10,20)*x+rnorm(1,0,10*x))+rnorm(10,0,1)
## fit simple models via find good starting values for the
parameters even if it throw a warning
m<-nls(y~q*x/(b*x))
```

Algorithm? Multiple Linear Regression Support Vector Regression Random Forest Boost Tree Local Regression

Figure 1. A Random Forest is built one tree at a time.

A Random Forest is a collection of decision trees. Each tree gets a "vote" in classifying. There are two components of randomness involved in the building of a Random Forest. First, at the creation of each tree, a random subsample of the total data set is selected to grow the tree. Second, at the creation of each tree, a random subset of the total features is selected to split on. One of all genes is chosen as a "splitting variable". The splitter variable attempts to separate patients in one class (e.g., Response) from those in the other class (e.g., Non-Response). The tree is grown with additional splits.

Parameter?

Option	Value	Description
none	none	no option
crossvalidation	Dropout	crossvalidation in r
caret	crossvalidation	caret in r

crossvalidation : validation options : value

Visualization History score Validation Scattering

Preview for Validation Scattering

Input Output

Desktop Title Index

Output

Desktop Title Index

Do Generate Code Cancel

C:\WR code generator\123\_gui\_parameters.json - Sublime Text (UNREGISTERED)

File Edit Selection Find Goto Tools Project Preferences Help

1 {"kernel": "R",  
2 "task": "Regression",  
3 "algorithm": "MLR",  
4 "hparams": {"cv": 2, "kernel": "linear", "ntrees": 100, "epochs": 3, "hidden": "10,10"},  
5 "perf\_eval": ["mae", "mape", "rmse", "rmsle", "r2"],  
6 "analysis": ["Variable Importance", "leveragePlots", "qqPlot", "Distribution of Std. Residuals"]  
7 "whole\_columns\_index\_and\_name": {"0": "thickness", "1": "Umach", "2": "AOA", "3": "RE", "4": "C1"},  
8 "input\_columns\_index\_and\_name": {"0": "thickness", "1": "Umach", "2": "AOA", "3": "RE"},  
9 "output\_columns\_index\_and\_name": {"4": "C1"},  
10 "datatype\_of\_columns": {"0": "Numeric", "1": "Numeric", "2": "Numeric", "3": "Numeric", "4": "Numeric"}  
11 "testing\_frame\_rate": 0.2,  
12 "testing\_frame\_extract\_method": "random",  
13 "ml\_file\_path": "C:/@WORK@/4. Airfoil/R\_Script/",  
14 "ml\_file\_name": "kflow\_data.csv",  
15 "number\_of\_rows": 7210  
16 }

## Data Loading

```
all_data = read.csv("C:/@WORK@/4. Airfoil/R_Script/kflow_data.csv", sep=",", header=TRUE)
```

## Discover the head of your data

```
head(all_data)
```

## Pre-processing - 1. Input Missing values

## There are many options in order to deal with a missing value such as:

## - A constant value that has meaning within the domain, such as 0, distinct from all other values.  
## - A value from another randomly selected record.  
## - A mean, median or mode value for the column.  
## - A value estimated by another predictive model.

## We are going to replace a missing value to 0. The following is the Simple R function.

```
na.zero <- function (x) {
  x[is.na(x)] <- 0
  return(x)
}
```

## Filling 0 for missing values

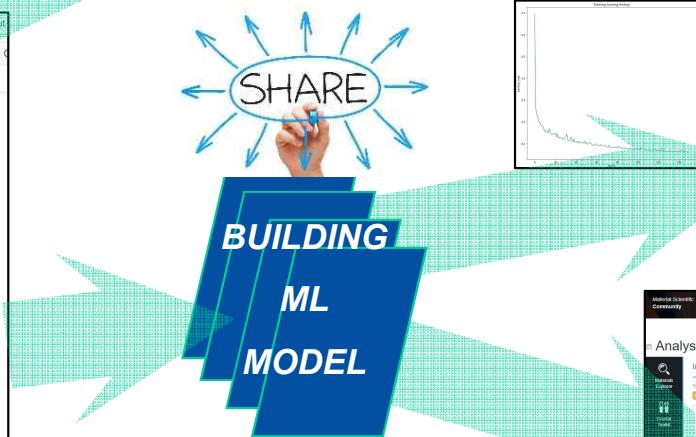
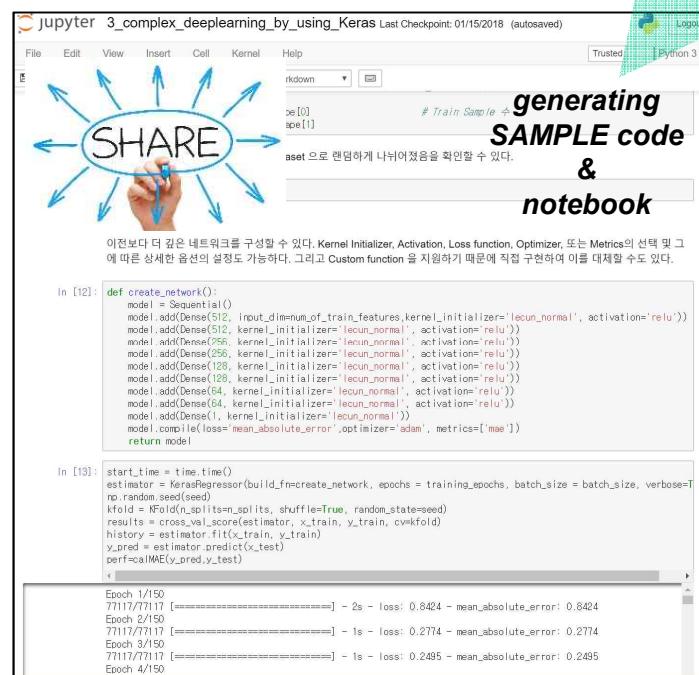
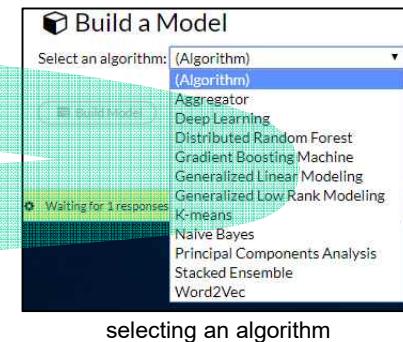
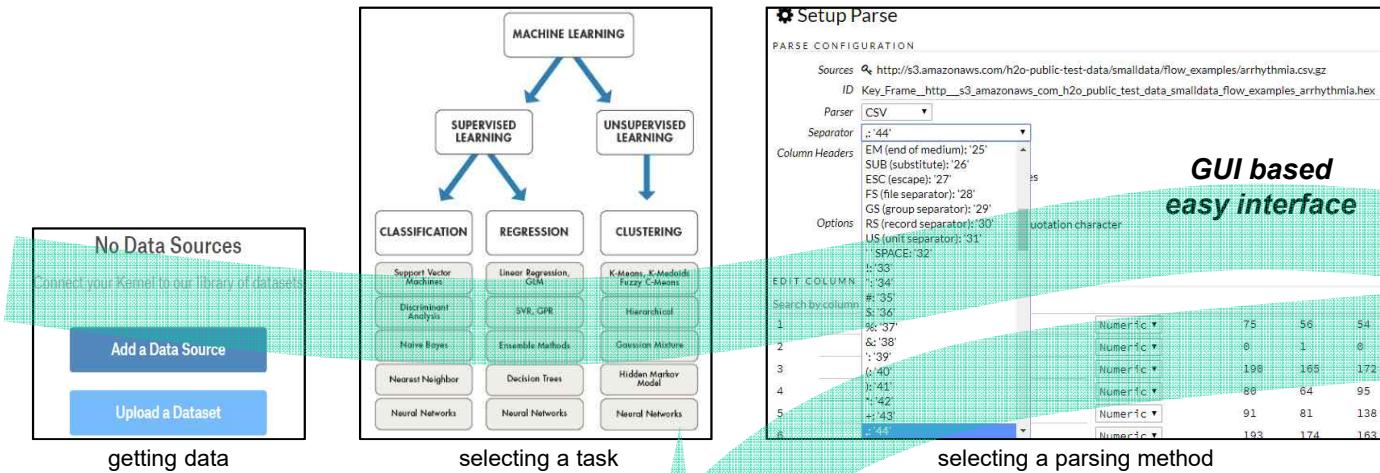
```
all_data <- na.zero(all_data)
```

## Pre-processing - 2. Vectorization

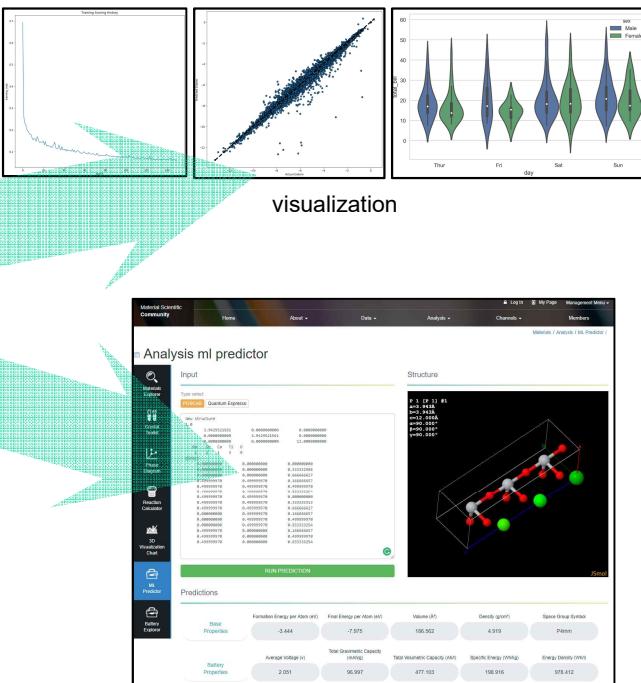
## Numeric data only. There is nothing to be vectorized.

## Train and Test Dataset Split

```
set.seed(123)
all_data = all_data[sample(nrow(all_data)), ]
train_data = all_data[1:5768, ]
test_data = all_data[5769:nrow(all_data), ]
```



# *RUN on the platform*



**1 Please upload your data (CSV file only)**

Data should have column names in the first row.

Title	My Dataset 01
Access level	Public (Visible)
License	CC BY
Data file	<a href="#">[CSV]</a> <small>or when uploading a CSV file, specify the first line as a header (column name).</small>
COD4M estimate	Auto

COPYRIGHT(C)2015 NIDM. KISTI ALL RIGHTS RESERVED. | Privacy Policy | Email: kya@kistia.go.kr

**2 Please check the data type and category**

Select this category to true if the data type is a string type or false if the data type is a number type.

ID	Title	Datatype	Category	Example Data
0	Sepal Length	Numeric	False	5.1
1	Sepal Width	Numeric	False	3.5
2	Petal Length	Numeric	False	1.4
3	Petal Width	Numeric	False	0.2
4	Species	String	True	Iris-setosa

COPYRIGHT(C)2015 NIDM. KISTI ALL RIGHTS RESERVED. | Privacy Policy | Email: kya@kistia.go.kr

**3-1 Select a target(Y)**

**3-2 Select feature lists (X)**

**3-3 Select Train / Test ratio**

Train: 80% Test: 20%

COPYRIGHT(C)2015 NIDM. KISTI ALL RIGHTS RESERVED. | Privacy Policy | Email: kya@kistia.go.kr

**4-1 Select a programming language**

Python

**4-2 Select a task**

Regression Classification

**4-3 Select an algorithm**

And Select optional parameters

**4-4 Select what to visualize**

AP Correlation Matrix, Violin Plot

**4-5 Select what to analyze**

MICE, RIDGE, PERT, LASSO

COPYRIGHT(C)2015 NIDM. KISTI ALL RIGHTS RESERVED. | Privacy Policy | Email: kya@kistia.go.kr

**4-3 Select an algorithm**

And Select optional parameters

**4-4 Select what to visualize**

AP Correlation Matrix, Violin Plot

**4-5 Select what to analyze**

MICE, RIDGE, PERT, LASSO

COPYRIGHT(C)2015 NIDM. KISTI ALL RIGHTS RESERVED. | Privacy Policy | Email: kya@kistia.go.kr

**5 Preview Code**

```

In [0]: ## Data Loading
all_data = read.csv("./input.csv", sep=",", header=TRUE)
## Discover the head of your data
head(all_data)

In [0]: ## Pre-processing - 1. Impute Missing values
## There are many options in order to deal with a missing value such as:
## - a value from another record in the dataset, such as 0, distinct from all other values
## - a value from another randomly selected record
## - a mean, median or mode for the column
## - a value from another predictor variable
## We are going to replace a missing value to 0. The following is the Simple R function
na_impute = function(x) {
  x[is.na(x)] <- 0
  return(x)
}

## Filling 0 for missing values
all_data <- na_impute(all_data)

In [0]: ## Pre-processsing - 2. Vectorization
## Numeric data only. There is nothing to be vectorized.

In [0]: ## Train and Test Dataset Split
all_data = all_data[sample(nrow(all_data), 1)] # data shuffling
train_size_rate = 0.47
split_index = nrow(all_data) * train_size_rate
print(split_index)
train_data = all_data[1:split_index]
test_data = all_data[(split_index + 1):nrow(all_data),]

In [0]: ## Generalized Boosted Regression (Decision Tree)
library(gbm)
model = gbm(formula = Sepal.Length ~ Petal.Length, data = train_data, n.trees = 100, cv.fold = 5)
## total number of trees to fit, k-fold cross-validation, custom options
#OPTIONAL NAME = (optional value)
#x: A symbolic description of the model to be fit. It may include an offset term (e.g., ~ offset(n))
#x: Data Frame
#n.trees = 100 // The total number of trees to fit. This is equivalent to the number of iterations and the number of basis functions in the additive expansion.
#cv.fold // Number of cross-validation folds to perform. "cv.error" shows an estimate of generalization error
#You can find arguments details on R cran repository at https://cran.r-project.org/web/packages/gbm/gbm.pdf

In [0]: ## results
print(model)

```

COPYRIGHT(C)2015 NIDM. KISTI ALL RIGHTS RESERVED. | Privacy Policy | Email: kya@kistia.go.kr



EDISON

Artificial  
Intelligence

Home

MI Data

My Page

Log Out My Account



1 Submit Data

2 Check Type

3 Select Label

4 Select Algorithm

5 Preview Code

## 1 Please upload your data (CSV file only)

Data file should have column names at the first raw.



Title	kflow data		
Access level	<input checked="" type="radio"/> Public <input type="radio"/> Private	License	CC BY
Community	Artificial Intelligence	CSV delimitate	Auto
Keyword	add a tag		
Data File	<input type="button" value="파일선택"/> 선택된 파일 없음 ※ When uploading a CSV file, specify the first line as a header (column name).		



- **Simple and Easy Machine Learning Web Platform (KISTI Data Platform)**
- **Community-based Data Sharing**
  - Datasets, collections, and interactive notebooks
- **A comfortable place for machine learning beginners to learn field experts' know-how**
- **User-Friendly Designed GUI Interfaces to Automatically Generate Programming code with User's Own Data**



THANK YOU

