



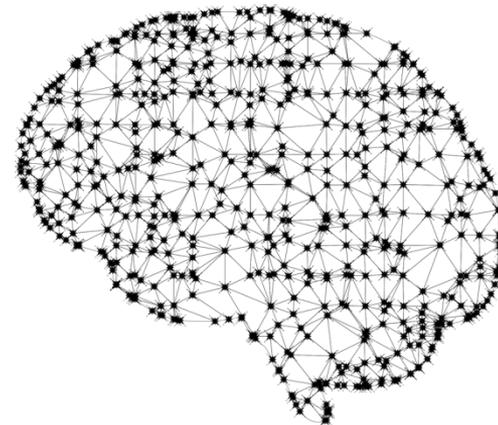
Past and Future of Our AI Making Use of Data Governance: How to Make Process/Product Innovations

Ryo Kiguchi, Shogo Miyazawa,
Masakazu Fujiwara, Yoshitake Kitanishi
SHIONOGI & CO., LTD.



Agenda

- What is AI?
- Importance of Data Governance in using AI (Past)
 - AI Making Use of Data Governance
 - Case : AI SAS Programmer (Ver.1.0)
- What is “True” Data Governance in AI ? (Now)
 - AI Making Use of “True” Data Governance
 - Case1 : AI SAS Programmer (Ver.2.0)
 - Case2 : AI application which semi-automatically implement QC
- For "Product Innovation" making use of Data Governance (Future)



What is AI?

The Japanese Society for Artificial Intelligence Ethical Guidelines

Preamble

Artificial Intelligence (“AI”) research focuses on the realization of AI, which is the enabling of computers to possess intelligence and become capable of learning and acting autonomously. AI will assume a significant role in the future of mankind in a wide range of areas, such as Industry, Medicine, Education, Culture, Economics, Politics, Government, etc. However, it is undeniable that AI technologies can become detrimental to human society or conflict with public interests due to abuse or misuse.



In considering the definition of artificial intelligence, it is important to understand the difference in position.

What is AI?

- Scientific potion
 - ✓ Artificial General Intelligence (AGI)
 - ✓ A position to make a machine with human intelligence itself
- Engineering potion
 - ✓ Narrow AI
 - ✓ A position to let a machine do what human beings do with intelligence



What is AI?

The “AI technology” we define

- ✓ Point : AI is a composite technology that combines various knowledge



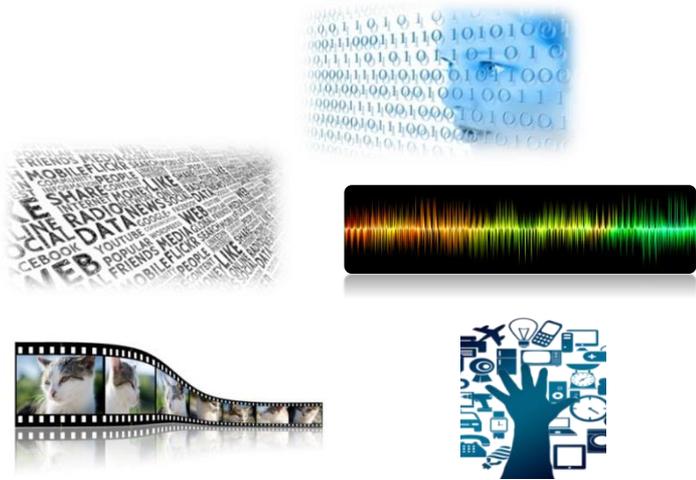
We define the technology used in this series of processes (Recognition / Learning / Action) as “AI technology”

Importance of Data Governance in using AI (Past)

Importance of Data Governance in using AI (Past)



- Data used in AI has various data structures such as structured data and unstructured data, and there are various kinds of data.



Structured data
Type of data that **can** be stored in the database

Clinical trial data



FAERS
FDA

JADER
PMDA

Unstructured data
Type of data that **cannot** be stored in the database

Package Inserts



シオノギ製薬
CRESTOR

PubMed



Twitter



- The Recognition, Learning, and Action of AI technology are different depending on the data format.
- In any data format, **Data governance** to collect, manage and archive data for innovation is important.
 - ✓ In anticipation of the future, it is important to do data governance.
- If achieve this data governance, it is possible to smoothly execute from “Recognition” to “Action”, and the acquisition of AI also is achieved.



Our Data governance strategy

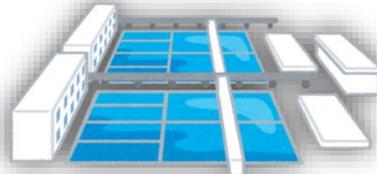
When try to manage data on the downstream, it will cost several times.

We have to realize quality improvement on the **upstream**.

Pour internal and external data



Data Lake



Data Reservoir



Extraction, reconstruction, re-accumulation from multiple in-house systems



Data Warehouse (DWH)



Data Mart

Level to aim!

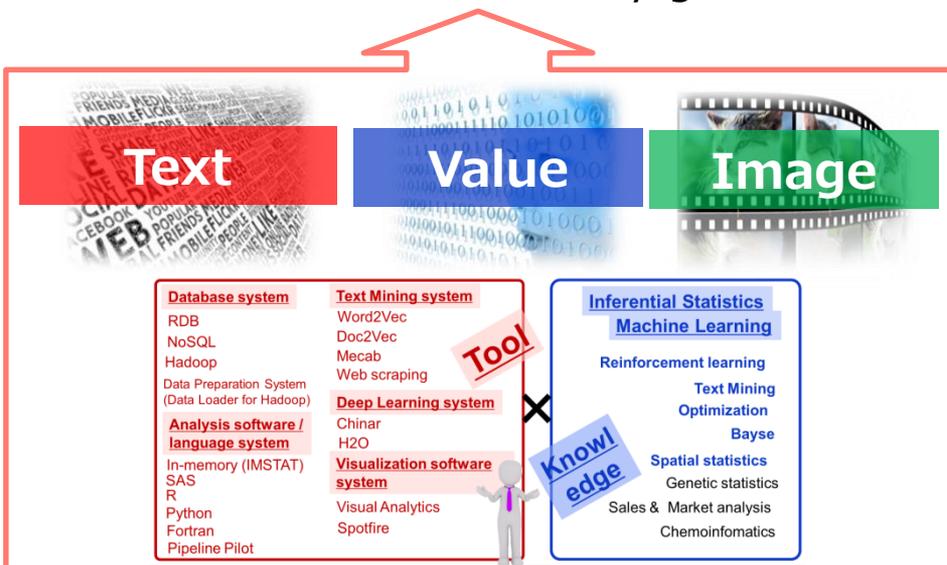
Store internal and external data with **high quality**.

Partially extract from DWH for specific purposes.

Innovation example by Data Governance AI SAS Programmer (Ver.1.0)

Semi-automate clinical trial analysis

- Semi-automatically generates SAS programs for analysis materials



SAS THE POWER TO KNOW.

NEWS RELEASE

報道関係各位

2016年4月7日
SAS Institute Japan 株式会社

AI project is jointly developed with SAS Institute Japan
We made a press release in April 2016

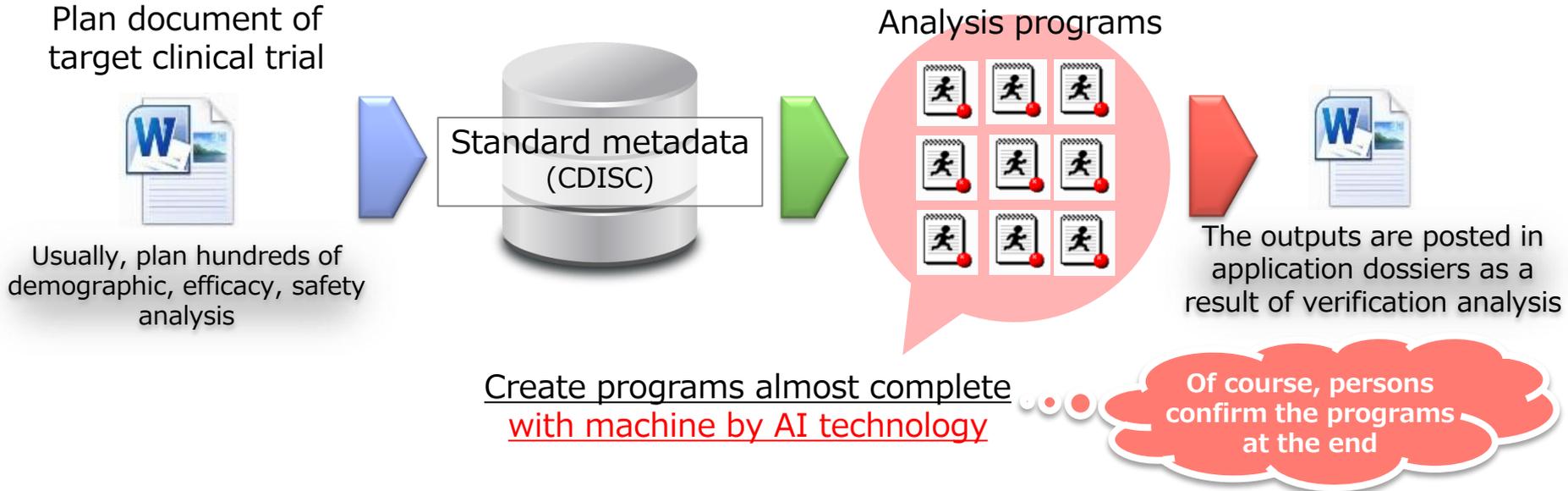
～AI技術による解析プログラム開発と解析ドキュメンテーション業務のセミオートメーション化～

～AI技術による臨床試験解析のセミオートメーション化に着手～

アナリティクスのリーディングカンパニーである SAS Institute Japan 株式会社(本社:東京都港区、代表取締役社長:堀田徹哉、以下 SAS)は、塩野義製薬株式会社(本社:大阪府中央区、代表取締役社長:手代木 功、以下 塩野義製薬)が、SAS の Hadoop 対応製品の一つである「SAS® In-Memory Statistics for Hadoop」の機械学習エンジンを活用し、臨床開発業務で使用される SAS プログラムとその関連文書を自動生成する人工知能(AI)アプリケーションの開発に着手したことを発表しました。

Outline of AI SAS Programmer

> Flow of Validation Analysis of Clinical Trial

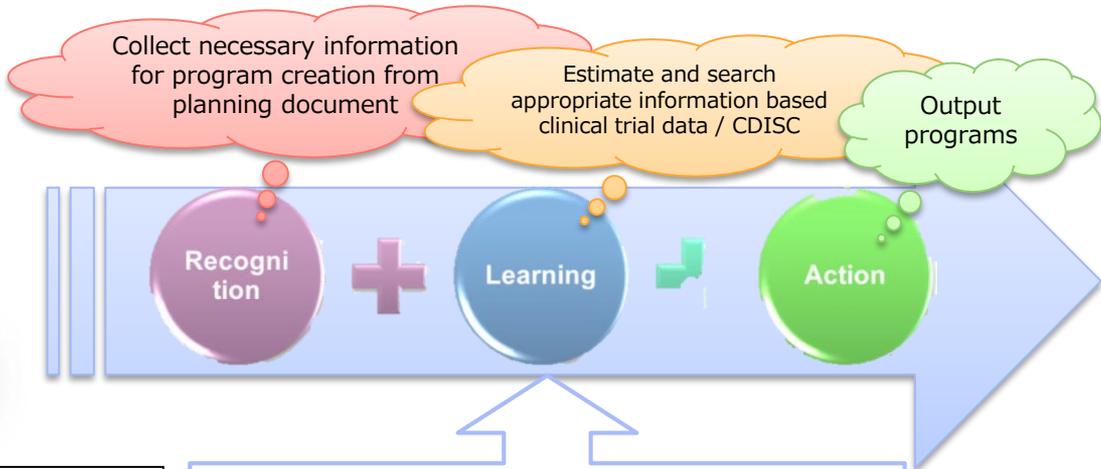


System outline and Governance strategy of AI SAS Programmer

Planning document for the target clinical trial



It's usually planned to create hundreds of programs



Semi-automatically generated analysis programs

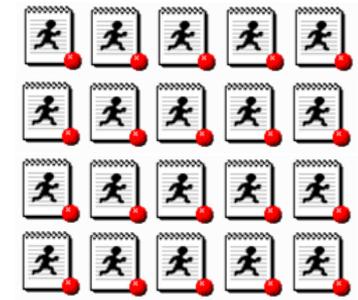
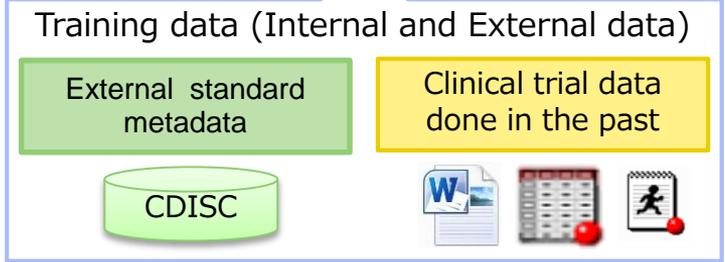


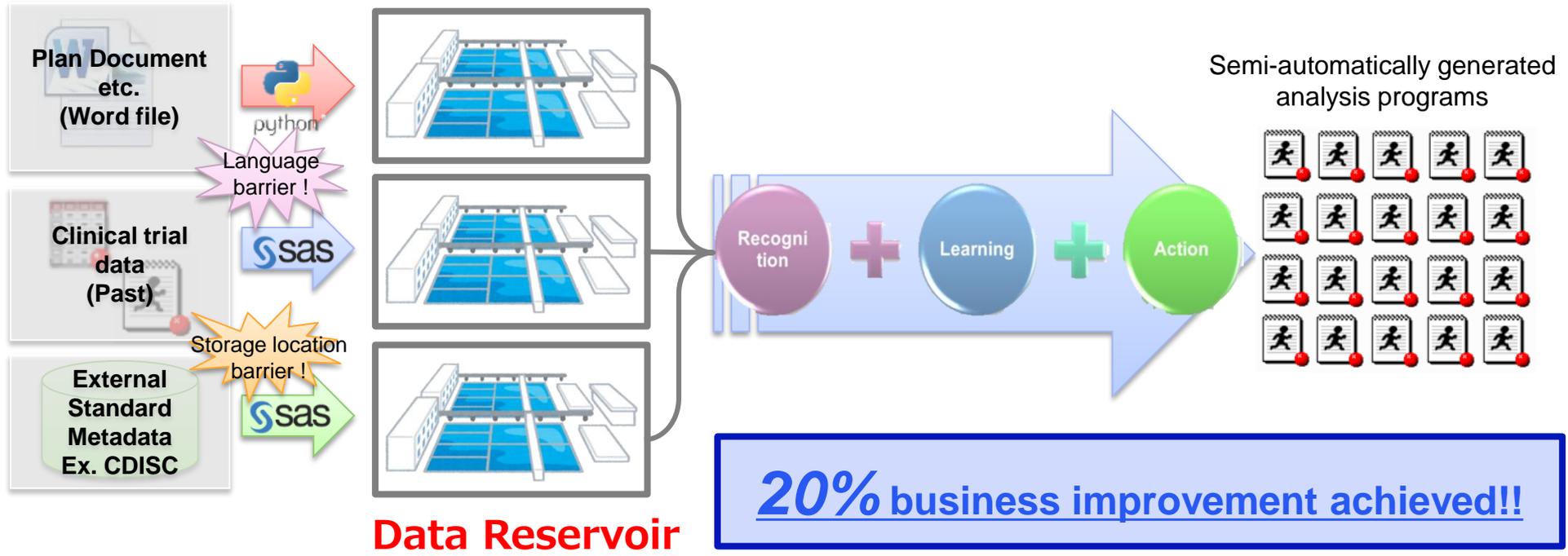
Table 14.1.1.1.1... Patient Disposition¹
All Randomized Patients¹

	S***** ¹	Placebo ¹
	N=** ¹	N=** ¹
	n(%) ¹	n(%) ¹
Completed ¹	** (***) ¹	** (***) ¹
Withdrawn ¹	** (***) ¹	** (***) ¹
Reason for withdrawal ¹	** (***) ¹	** (***) ¹
ex. Adverse event ¹	** (***) ¹	** (***) ¹
ex. Lack of efficacy ¹	** (***) ¹	** (***) ¹
ex. Recovery ¹	** (***) ¹	** (***) ¹
ex. Withdrawal by subject ¹	** (***) ¹	** (***) ¹
ex. Lost to follow-up ¹	** (***) ¹	** (***) ¹
ex. Other ¹	** (***) ¹	** (***) ¹

Plan document



Data Governance strategy in AI SAS Programmer (Ver.1.0)



What is “True” Data Governance in AI ?(Now)

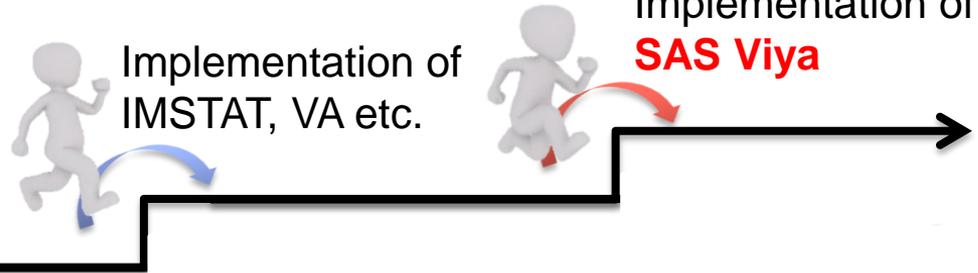
Transition and evolution of analysis environment

Points of analysis system’s construction

- ✓ Consider the contents of data science strategy, progress situation, and system tailored to the times.
- ✓ For data scientists, the environment that can be analyzed efficiently is necessary, because data size becomes huge and the analysis method becomes complicated.
- ✓ Clarify the roles of multiple systems, and consider the combination, optimization.

Transition of analysis environment of Biostatistics Center

In 2011,
began to develop the analysis
environment of BigData



What is “True” Data Governance in AI ?(Now)

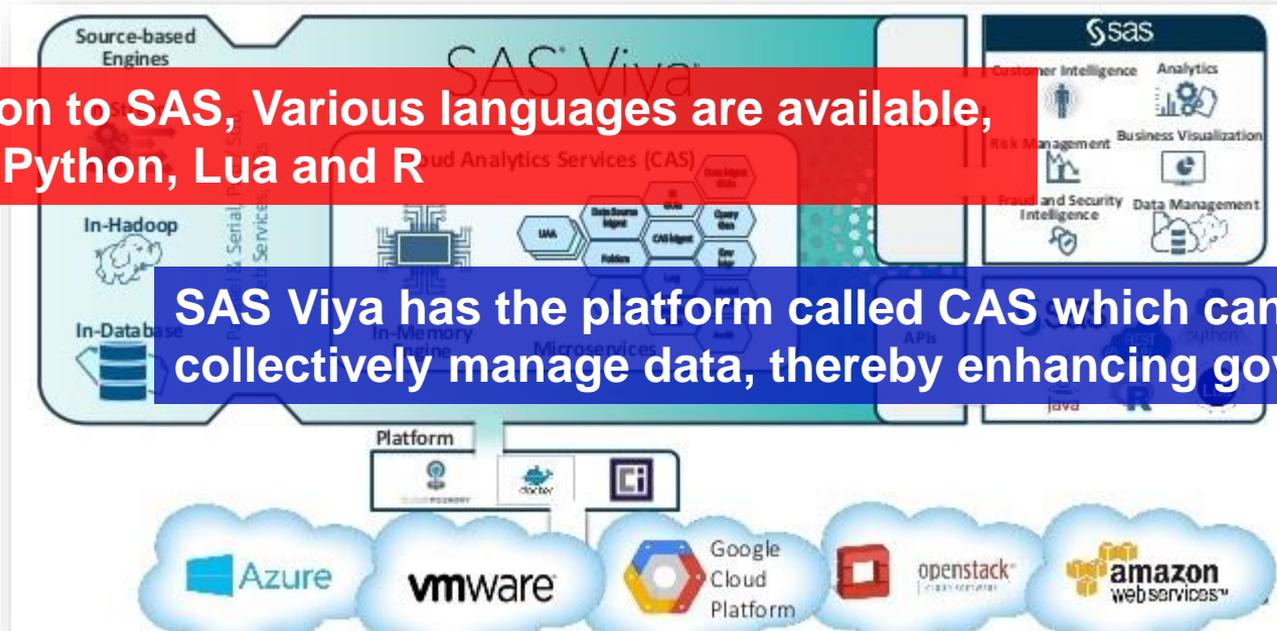


Data Governance Strategy by SAS Viya

- ✓ The characteristics of SAS Viya which we focused on

In addition to SAS, Various languages are available, such as Python, Lua and R

SAS Viya has the platform called CAS which can collectively manage data, thereby enhancing governance.



Data Governance Strategy by SAS Viya

✓ The characteristics of SAS Viya which we focused on

- In addition to SAS, Various languages are available, such as Python, Lua and R

Pros

Cons

SAS

- Good at data handling
- Validation is guaranteed

- In order to implement the latest analysis method, it is necessary to make programs
- Cannot handle Word or PDF

R

- Library is fulfilling (It is possible to easily implement the latest analysis method)

- Not good at data handling
- Validation is not guaranteed

Python

- Can handle Word or PDF
- Can implement the latest method using packages etc

- Not good at data handling
- Processing time is slow

Lua

- Good at data governance because it is possible to set variables independent of data format
- Processing time is fast

- Debug is difficult
- It is difficult to solve dynamic errors

Data Governance Strategy by SAS Viya

- ✓ The characteristics of SAS Viya which we focused on
 - SAS Viya has the platform called CAS which can collectively manage data, thereby enhancing governance.

- Since analysis engines are CAS in common regardless of languages, analysis results are always consistent.
- Data can be easily abstracted from SAS Viya’s platform as necessary, and it is easy to analyze with appropriate programming languages.
- Data driven programming languages selection can be performed smoothly.



What is “True” Data Governance in AI ?(Now)

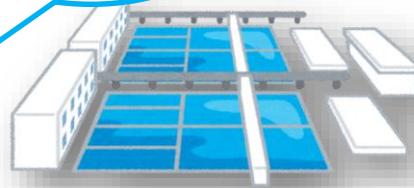
Data Governance Strategy by SAS Viya

✓ Summary of data governance strategy

Data driven programming
languages selection



Deep Learning
CNN
Machine Learning
Text Mining
Propensity Score
etc.



Data Reservoir



**Data driven
Innovation**



What is “True” Data Governance in AI ?(Now)

“True” Data Governance = One step forward of data governance



Why?

The usage scene of AI is ...



What is “True” Data Governance in AI ?(Now)

“True” Data Governance = One step forward of data governance

Deliverable Governance

Innovation deliverables

- Model
- Program
- Macro etc.

Deep Learning
Machine Learning
Text Mining
Hidden Markov Model
Propensity Score
etc.

Data driven
Innovation

There are some sort of governance, so people should feel like it!

Example of “some sort of governance”

- ✓ CDISC
- ✓ MedDRA
- ✓ Deliverable like “Package Inserts”
- ✓ Internal standard template
- ✓ Model
- ✓ Program
- ✓ Macro etc.

- ✓ Data with governance as deliverables or data with some sort of regularities is likely to cause innovation.
- ✓ By using the innovation deliverables(subject to governance), if it is similar to the governance system of past innovation, can make use of it by extending, applying, integrating and eliminating options in existence when want to innovate with data having new “some sort of governance”.

What is “True” Data Governance in AI ?(Now)



Data Governance strategy in AI SAS Programmer (Ver.2.0)

New analysis in AI SAS Programmer: CNN (Convolutional Neural Network)

What's CNN ?

- CNN is Neural Network (NN) with many layers.
- CNN is one of the powerful tools in the field of image analyses.

```
%matplotlib inline
import sys
sess = svml.CAS("sasviya-address", 123, "sasword")
from dply.images import ImageTable
my_images = ImageTable.load_images(sess, path="img_path")
my_images.resize(width=224)

from dply.splitting import two_way_split
tr_img, te_img = two_way_split(my_images, test_rate=20, seed=123)
tr_img.as_patches(width=200, height=200, step_size=24, output_width=224, output_height=224)

from dply import Model, Sequential
from dply.layers import *
from dply.applications import *

model1 = Sequential(sess, model_table="CNN sample program")
model1.add(InputLayer(3, 224, 224, offsets=tr_img.channel_means))
model1.add(Conv2d(8, 7))
model1.add(Pooling(2))
model1.add(Conv2d(8, 7))
model1.add(Pooling(2))
model1.add(Dense(10))
model1.add(OutputLayer(act="softmax", n=2))
```

CNN
by SAS Viya's Python interface

```
proc cas ;
loadactionset "image";

image.loadImages / casout=(name="image" replace=1)
path="/var/viya_data/image"
recurse=TRUE decode=TRUE labels=;

imageFunctions={functionOptions={functionType= RESIZE ,height=224,width=224}};

shuffle / casout=(name="resized_shuffled", replace=1)
table={name="resized_image"};

deepLearn.buildModel /
modelTable={name="SAS_TEST", replace=TRUE} type="CNN";

deepLearn.addLayer /
layer={type="INPUT" nChannels=3 width=224 height=224}
modelTable={name="SAS_TEST"} name="data";

deepLearn.addLayer /
layer={type="CONV" nFilters=8 width=7 height=7}
modelTable={name="SAS_TEST"} srdLayers={data};

deepLearn.addLayer /
layer={type="POOL" width=2 height=2}
modelTable={name="SAS_TEST"} name="pool1" srdLayers={conv1};

deepLearn.addLayer /
layer={type="CONV" nFilters=8 width=7 height=7}
modelTable={name="SAS_TEST"} name="conv2" srdLayers={pool1};

deepLearn.addLayer /
layer={type="POOL" width=2 height=2}
modelTable={name="SAS_TEST"} name="pool2" srdLayers={conv2};

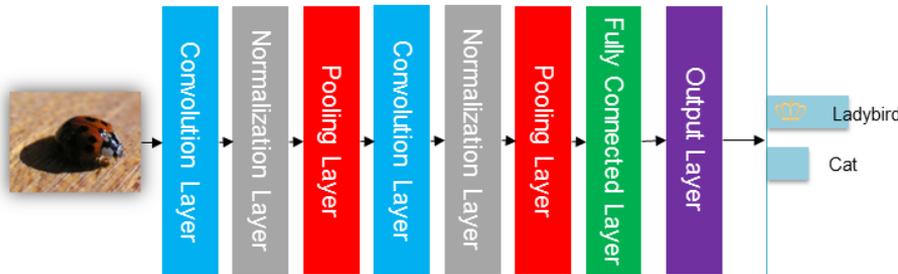
deepLearn.addLayer /
layer={type="OUTPUT" act="softmax" n=2}
modelTable={name="SAS_TEST"} name="outlayer" srdLayers={pool2};

deepLearn.modelInfo / modelTable={name="SAS_TEST"};

deepLearn.dlTrain /
inputs={name="image_"}
modelTable={name="SAS_TEST"}
modelWeights={name="SAS_TEST_W" replace=TRUE}
table={name="resized_shuffled"}
optimizer={maxEpochs=1 minBatchSize=128};

quit;
```

CNN
by SAS Viya's SAS studio interface



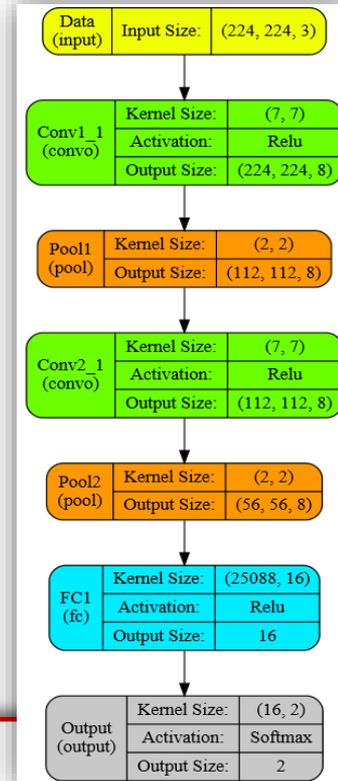
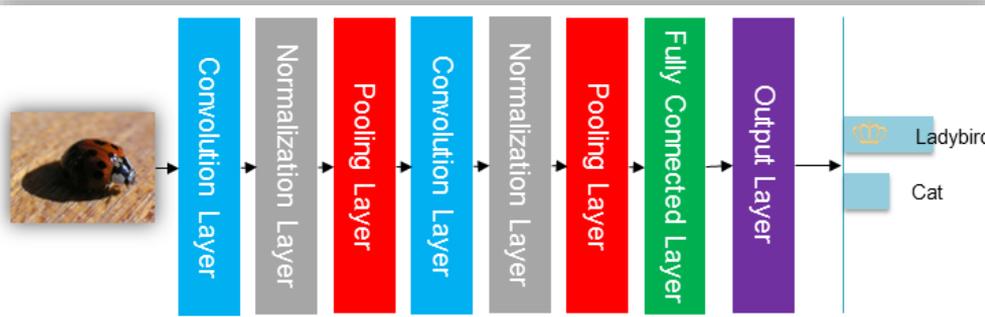
What is “True” Data Governance in AI ?(Now)

Data Governance strategy in AI SAS Programmer (Ver.2.0)

New analysis in AI SAS Programmer: CNN (Convolutional Neural Network)

CNN's modeling program by SAS Viya's Python interface

```
model1 = Sequential(sess, model_table='CNN sample program')
model1.add(InputLayer(3, 224, 224, offsets=tr_img.channel_means))
model1.add(Conv2d(8, 7))
model1.add(Pooling(2))
model1.add(Conv2d(8, 7))
model1.add(Pooling(2))
model1.add(Dense(16))
model1.add(OutputLayer(act='softmax', n=2))
```



What is “True” Data Governance in AI ?(Now)

Data Governance strategy in AI SAS Programmer (Ver.2.0)

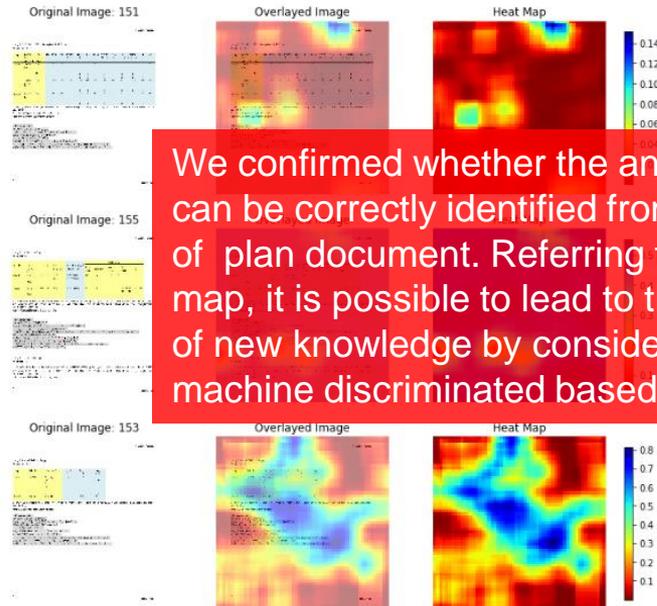
New analysis in AI SAS Programmer: CNN (Convolutional Neural Network)

Table 14.1.1.1... Patient Disposition^a
All Randomized Patients^a

	S***** ^a	Placebo ^a
	N=*** ^a	N=** ^a
	n (%) ^a	n (%) ^a
Completed ^a	** (***) ^a	** (***) ^a
Withdrawn ^a		** (***) ^a
Reason for withdrawal ^a		** (***) ^a
ex. Ineligibility ^a	** (***) ^a	** (***) ^a
ex. Adverse event ^a	** (***) ^a	** (***) ^a
ex. Lack of efficacy ^a	** (***) ^a	** (***) ^a
ex. Recovery ^a	** (***) ^a	** (***) ^a
ex. Withdrawal by subject ^a	** (***) ^a	** (***) ^a
ex. Lost to follow-up ^a	** (***) ^a	** (***) ^a
ex. Other ^a	** (***) ^a	** (***) ^a

Plan document

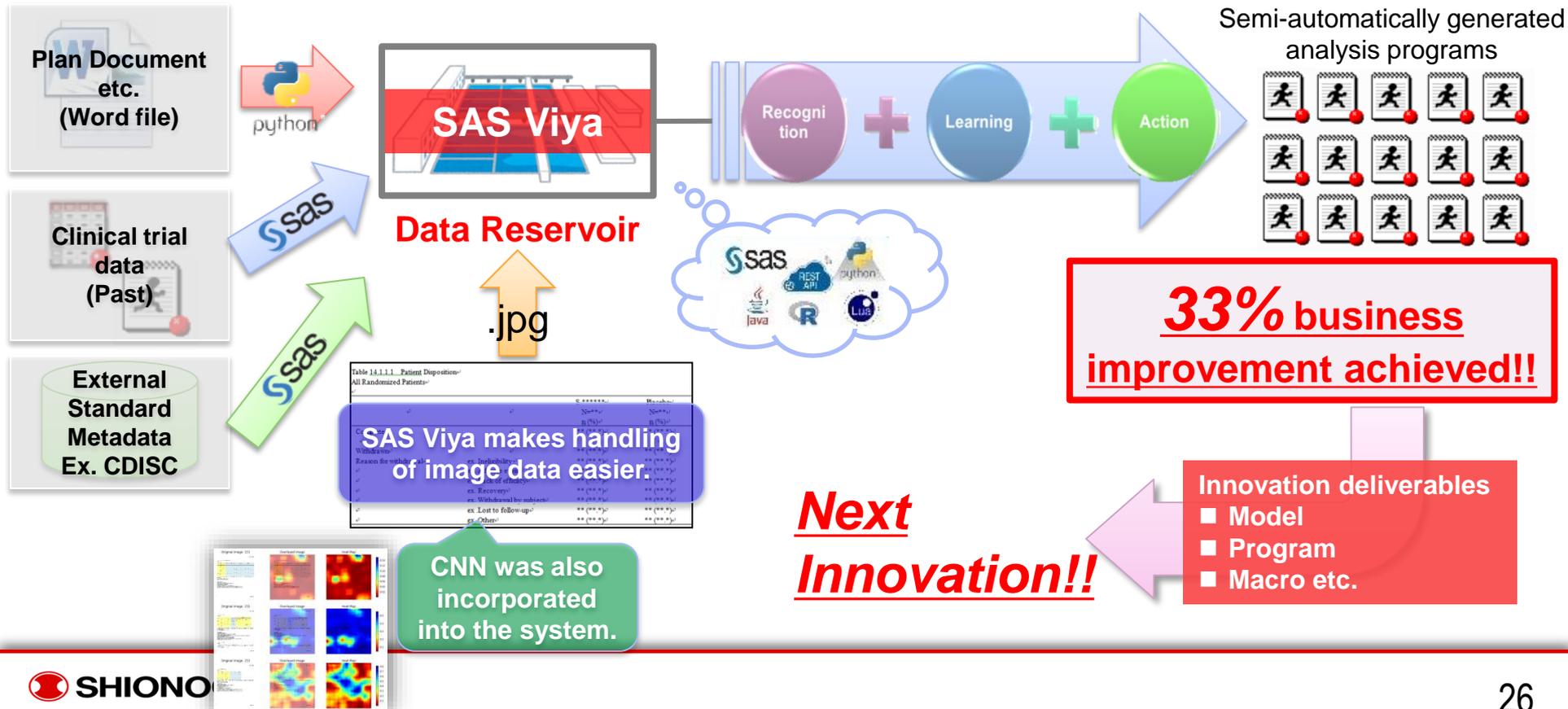
Convert sentences and tables into jpg files (image data) and store them in SAS Viya



We confirmed whether the analysis content can be correctly identified from the images of plan document. Referring to this heat map, it is possible to lead to the discovery of new knowledge by considering where the machine discriminated based on.

What is “True” Data Governance in AI ?(Now)

Data Governance strategy in AI SAS Programmer (Ver.2.0)



What is “True” Data Governance in AI ?(Now)



AI application which semi-automatically implement QC

Since it is used for approval documents, it is necessary to ensure the quality of the analysis result. For that purpose, we will implement QC (Quality Control)

Importance of Data Governance in using AI (Past)

Outline of AI SAS Programmer
> **Flow of Validation Analysis of Clinical Trial**

Plan document of target clinical trial

Standard metadata (CDISC)

Analysis programs

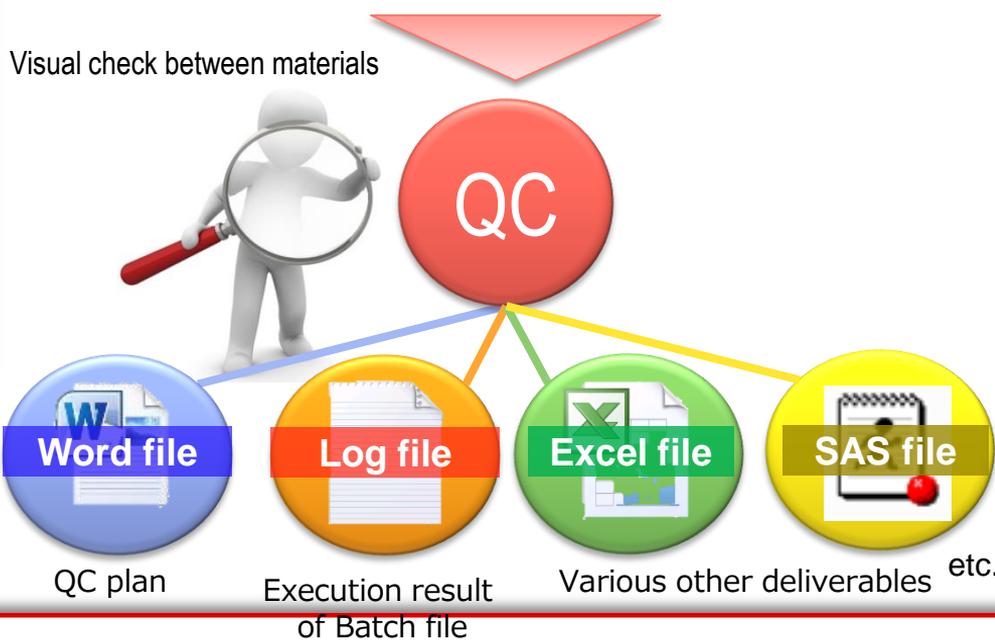
Usually, plan hundreds of demographic, efficacy, safety analysis

Create programs almost complete with machine by AI technology

The outputs are posted in application dossiers as a result of verification analysis

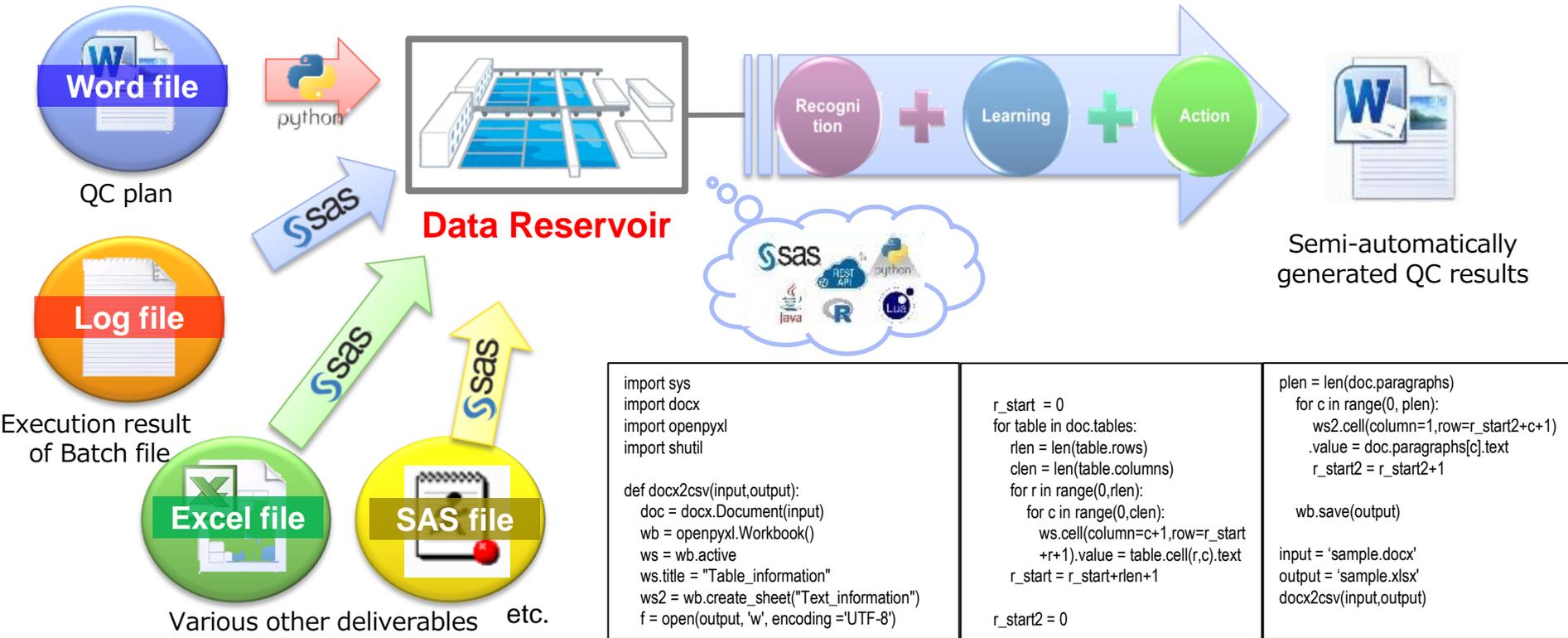
Of course, persons confirm the programs at the end

SHIONOGI 12



What is “True” Data Governance in AI ?(Now)

AI application which semi-automatically implement QC



```
import sys
import docx
import openpyxl
import shutil

def docx2csv(input,output):
    doc = docx.Document(input)
    wb = openpyxl.Workbook()
    ws = wb.active
    ws.title = "Table_information"
    ws2 = wb.create_sheet("Text_information")
    f = open(output, 'w', encoding='UTF-8')
```

```
r_start = 0
for table in doc.tables:
    rlen = len(table.rows)
    clen = len(table.columns)
    for r in range(0,rlen):
        for c in range(0,clen):
            ws.cell(column=c+1,row=r_start
                    +r+1).value = table.cell(r,c).text
            r_start = r_start+rlen+1

r_start2 = 0
```

```
plen = len(doc.paragraphs)
for c in range(0, plen):
    ws2.cell(column=1,row=r_start2+c+1)
    .value = doc.paragraphs[c].text
    r_start2 = r_start2+1

wb.save(output)

input = 'sample.docx'
output = 'sample.xlsx'
docx2csv(input,output)
```

For "Product Innovation"
making use of Data Governance (Future)

Key Word is Data Driven !

➤ Data driven type business improvement (Past / Now)

- Based on internal and external data, achieve business improvement from diversified viewpoints by making use of the latest technology.

➤ Data driven drug development (Now / Future)

- ✓ Perform the operation of clinical trial plan, strategy planning, and efficient analysis
- ✓ Based on internal and external data, create new hypotheses by data analysis at the time of planning or after the trial as the purpose that effectively development and maximize the product value.



Product value maximization making use of data governance

Product Innovation!!

Conclusion



- Data Governance is important in acquiring AI technology.
- In this presentation, introduced the data governance strategy at our company, and introduced the following process innovation case
 - ✓ AI SAS Programmer
 - ✓ QC application
- For the future, based on the Data Governance strategy, we will focus on the following.
 - ✓ Maximize product value (Product Innovation) using AI technology



- Yoshitake Kitanishi (2018). “Data Scientist and SAS Viya ~Speed up the cycle of Hypothesis \Leftrightarrow Verification~.” SAS FORUM JAPAN 2018.
- Ryo Kiguchi (2018). “Way of Life as Data Scientist from Viewpoint of Awareness Reform by SAS Viya ~How do we live?~.” SAS user general meeting2018.
- Medium, “Convolutional Neural Networks (CNN, or ConvNets)”, 2017
[<https://medium.com/@phidaouss/convolutional-neural-networks-cnn-or-convnets-d7c688b0a207>]
- Ryo Kiguchi (2018). “Process Innovation Making Use of CNN by SAS Viya ~ How does the Machine cognize Analytic TLF (Table, Listing, Figure)? ~.” SAS user general meeting2018.