

빅데이터 분석 도구

R·Flow

2018. 5.

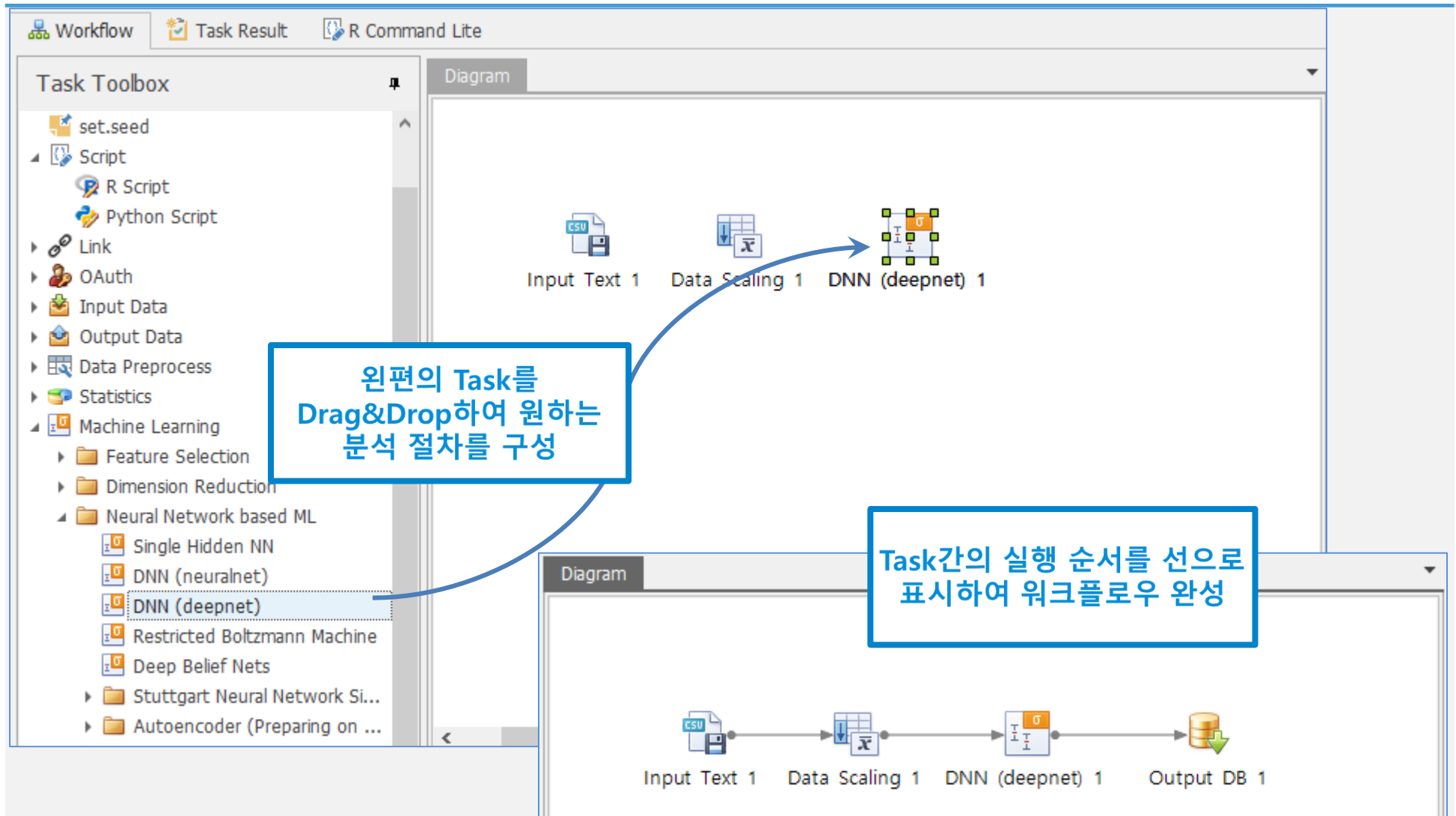


화수목

R-Flow 기본



R-Flow 기본(1/6) – 워크플로우 구성



※ Task는 주로 R Package로 구성된 모듈이나, R Script, Python Script도 사용할 수 있으며, TensorFlow, Keras, Spark도 지원

R-Flow 기본(2/6) – 옵션 설정

Task Information

Data Subset 2 | DNN (deepnet) 1 | DNN (deepnet) 2

Target Data: df.scale.1

Model(Object) Name: dnn.model.1

Merge Fitted Value

Merge Data: df.cnv

Output Name: dnn.result.1

☒ Enable

Variable Name	Data Type
donor	integer
age	integer
BMI	numeric

X: BMI, bchol, fchol, TG, HDL, LDL

Y: osteoporosis

Arguments

problem type: Classification

initW: NULL

initB: NULL

hidden: 10, activationfun: sigm

learningrate: 0.9, momentum: 0.5

learningrate_scale: 1, output: softmax

numepochs: 20, batchsize: 110

hidden_dropout: 0, visible_dropout: 0

set training: Data

오른편의 Task Information 창에서 Task에서 사용할 입출력 데이터와 Task 상의 옵션을 설정

Task Information 창의 Arguments를 클릭하면, 해당 Task Package의 설명이 제공됨

R-Flow Help

R Document | Task Document

nn.train {deepnet}

Training Neural Network

Description

Training single or mutple hidden layers neural network by BP

Usage

```
nn.train(x, y, initW = NULL, initB = NULL, hidden = c(10), activationfun = "sigm", learningrate = 0.8, momentum = 0.5, learningrate_scale = 1, output = "sigm", numepochs = 3, batchsize = 100, hidden_dropout = 0, visible_dropout = 0)
```

Arguments

x	matrix of x values for examples
y	vector or matrix of target values for examples
initW	initial weights. If missing chosen at random
initB	initial bias. If missing chosen at random
hidden	vector for number of units of hidden layers.Default is c(10).

※ 특정 Task가 아닌 R-Flow 전체 매뉴얼은 우측 상단 아이콘을 클릭하면 제공됨

R-Flow 기본(3/6) – 워크플로우 실행과 저장

The screenshot displays the R-Flow software interface. At the top, there is a menu bar with 'Home' and 'R Management Tools'. Below it is a toolbar with various icons. A blue dashed box highlights the 'Run All Task' and 'Run Selected Task' buttons. A blue arrow points from this box to a text box that says '워크플로우(전체 Task) 또는 선택한 Task만을 실행' (Execute workflow (all tasks) or only selected tasks). Another blue arrow points from the 'Run All Task' button to a text box that says '하단에서 처리 상태와 로그 확인' (Check processing status and logs at the bottom). A third blue arrow points from the 'Save Workflow' button to a text box that says '정의한 워크플로우를 저장' (Save the defined workflow). The 'Save Workflow' dialog box is open, showing fields for 'Workflow Name' (Deep Neural Network), 'Folder Name', 'Annotate', and 'Description'. The 'Task Procession Log' table is visible, showing a list of tasks and their execution times. The 'Task Console Log' is also visible, showing the output of the tasks.

Analysis Workflow x

Home R Management Tools

New [Icons] Run All Task Run Selected Task Batch Run Rename Task Delete Task, Link

정의한 워크플로우를 저장

Save Workflow

Workflow Name: Deep Neural Network

Folder Name:

Annotate:

Description:

Ok Cancel

Task Procession Log

Task Name	Start Time	End Time	Exec. Time	
Data Scaling by Scale In...	04:40:50	04:40:50	0:0:0:43	✓
Data Subset 1	04:40:50	04:40:50	0:0:0:74	✓
Data Subset 2	04:40:50	04:40:51	0:0:0:83	✓
Data Predict 1	04:40:51	04:40:51	0:0:0:19	✓
Data Predict 2	04:40:51	04:40:51	0:0:0:7	✓

Task Console Log

Clear

15 RMSE

16

17

18 The following object is masked from 'package:base':

19

20 Recall

21

22

워크플로우(전체 Task) 또는 선택한 Task만을 실행

하단에서 처리 상태와 로그 확인

R-Flow 기본(4/6) – 실행 결과

실행 결과 워크시트는 엑셀
다운로드하여 활용

Analysis Workflow

HomeR Management Tools

NewNewOpenSaveCloseRun All TaskRun Selected TaskBatch RunRename TaskDelete Task, LinkE.D.A

03. DNN(deepnet)

WorkflowTask ResultR Command Lite

D33

	A	B	C	D		H	I	J		
1		model name : dnn.model.1, package : deepnet, da								
2		problem typ	hidden	initW		momentum	learningrate_scale	output		
3		classificati	10	NULL	NULL		1	softmax		
4										
5		confusion matrix								
6			0	1						
7	0		51	21						
8	1		8	12						
9	2		11	20						
10										
11		overall statistics								
12		Accuracy	Kappa	95% CI Lower	95% CI Upper	No Information Rate	P-Value [Acc > NIR]	Mcneemar Test P-Value		
13		0.529801	0.270134	0.114	0.446982	0.307	0.611428	0.023		
14										
15		statistics by class								
16			Sensitivity	Specificity	Pos Pred Value	Neg Pred Value	Precision	Recall	F1	Prevalence
17		Class: 0	0.728571	0.429	0.679012	0.346	0.662337	0.662	0.728571	0.429
18		Class: 1	0.226415	0.094	0.857142	0.857	0.461538	0.462	0.226415	0.094
19		Class: 2	0.607142	0.857	0.747967	0.748	0.354166	0.667	0.607142	0.857

Task별 실행 결과를 워크시트에 표시

✓ 데이터 처리 Task는 작업 시 반복되는 기술통계와 차트를 기본으로 제공

✓ 모형 분석 Task는 주요 설정 옵션값과 검증성과지표를 자동으로 제공

R-Flow 기본(5/6) – 스케줄(서버 버전 only)

스케줄 등록

스케줄 목록

스케줄 명	시작일시	종료일시	상태
-------	------	------	----

스케줄 조회

조회일자 2016-12-28 ~ 2017-12-31

스케줄 명

상태 ☒ Waiting ☒ Running ☒ Done ☒ Error

워크플로우 명

조회

스케줄 목록

스케줄 명	자수	예약일시	시작일시	종료일시	상태	워크플로우 명
매월 말일 0시 0분 실행 스케줄	1	2016-12-31 00:00:00.0			Waiting	logistic model
매월 말일 0시 0분 실행 스케줄	2	2017-01-31 00:00:00.0			Waiting	logistic model
매월 말일 0시 0분 실행 스케줄	3	2017-02-28 00:00:00.0			Waiting	logistic model
매월 말일 0시 0분 실행 스케줄	4	2017-03-31 00:00:00.0			Waiting	logistic model
매월 말일 0시 0분 실행 스케줄	5	2017-04-30 00:00:00.0			Waiting	logistic model
매월 말일 0시 0분 실행 스케줄	6	2017-05-31 00:00:00.0			Waiting	logistic model
매월 말일 0시 0분 실행 스케줄	7	2017-06-30 00:00:00.0			Waiting	logistic model
매월 말일 0시 0분 실행 스케줄	8	2017-07-31 00:00:00.0			Waiting	logistic model
매월 말일 0시 0분 실행 스케줄	9	2017-08-31 00:00:00.0			Waiting	logistic model
매월 말일 0시 0분 실행 스케줄	10	2017-09-30 00:00:00.0			Waiting	logistic model

Schedule Information

스케줄 명

실행기간 2016-12-27 ~ 2016-

워크플로우 명

실행정보

Month

Week

☐ 1

☐ 월요일

☐ 2

☐ 화요일

☐ 3

☐ 수요일

☐ 4

☐ 목요일

☐ 5

☐ 금요일

☐ 6

☐ 토요일

☐ 7

☐ 일요일

☐ 8

☐ 9

☐ 10

☐ 11

☐ 12

User Name

Reg.DateTime

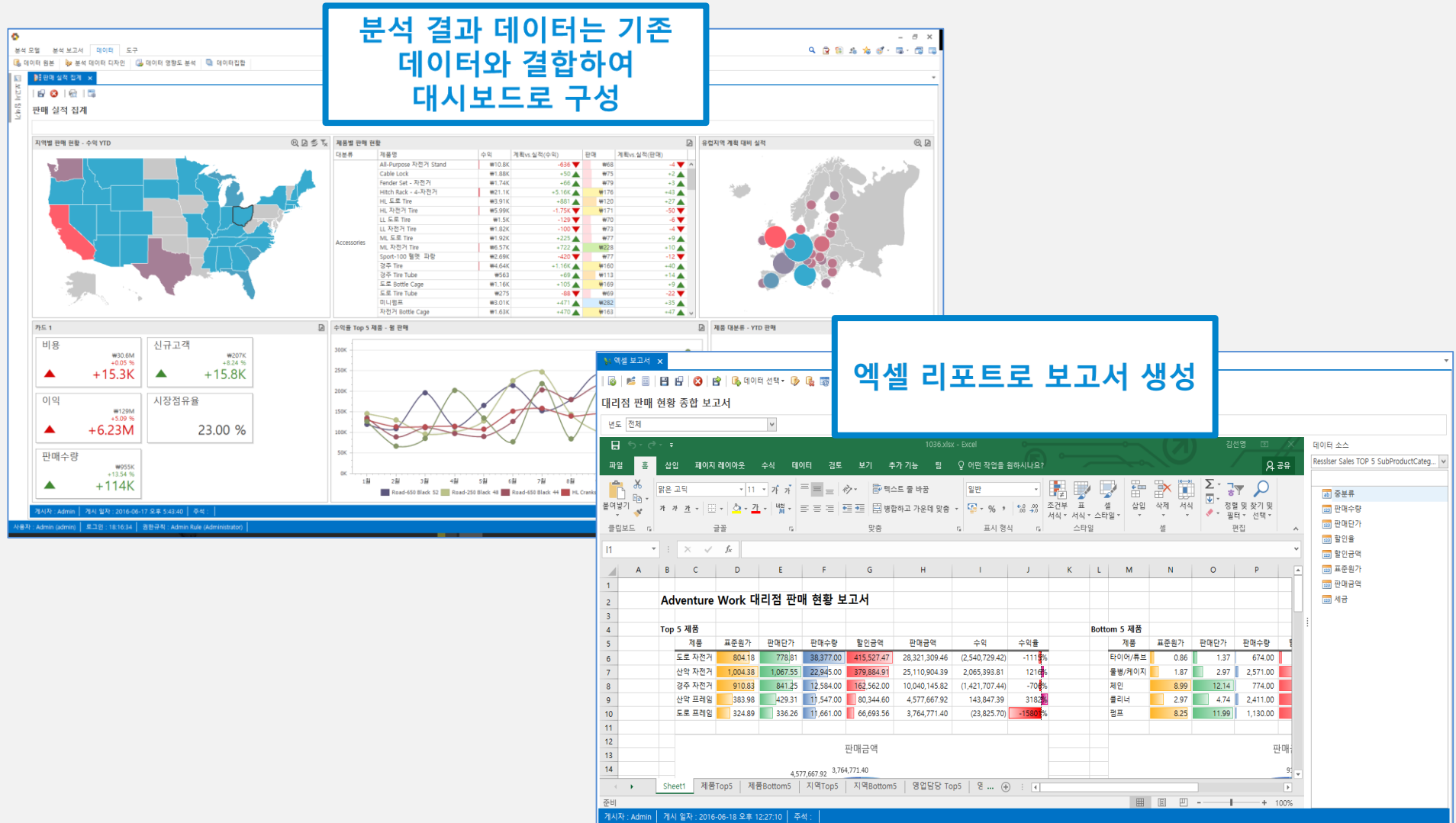
Annotation

Task 목록

태스크 명	시작일시	종료일시	상태	오류 메시지
-------	------	------	----	--------

주기적으로 실행하여야 할 워크플로우는
스케줄 설정으로 자동 실행됨

R-Flow 기본(6/6) – 리포팅(유료 버전 only)



R-Flow 무료 버전 다운로드 (software.naver.com에서 R-Flow 검색)

NAVER Software  **검색**

문서/사무 > 업무

내 PC의 OS와 브라우저 정보입니다. X

 **R-Flow**
인기도  | **업무** 127위 | 다운로드수 70 (최근 7일:7)



무료 다운로드

v2.0.0.7
★★★★★ 5 (4명)

이전인기버전 v2.0.0.5


업데이트 2017.12.11.

사용범위 프리 - 개인, 국내

 소프트웨어 사용범위 위반 시 법적 책임을 질 수 있습니다.

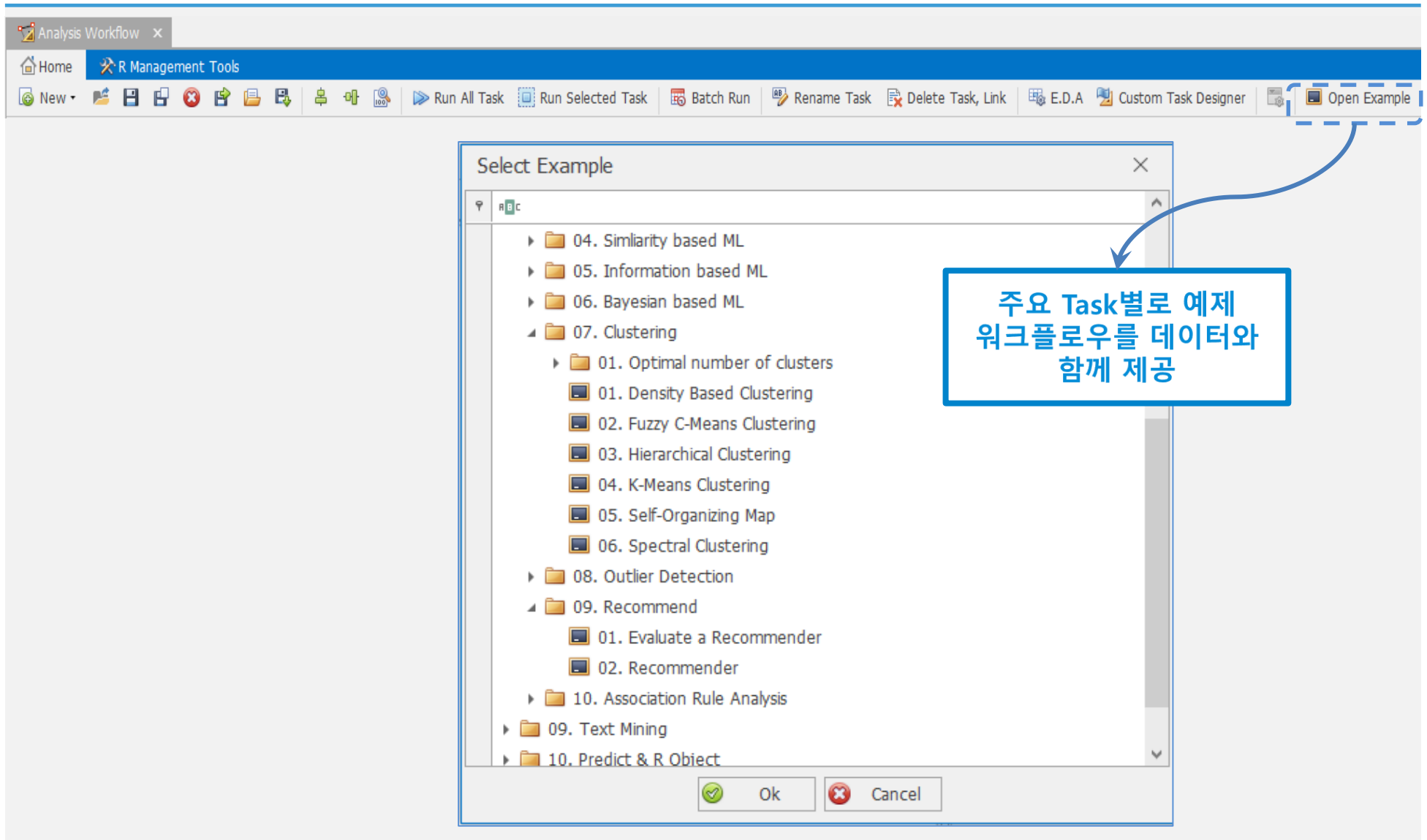
OS Windows

유한회사 화수목
개발사의 다른 S/W > 공식 웹사이트 >



Hwasum ok. Analytics 데이터전처리

R-Flow 예제

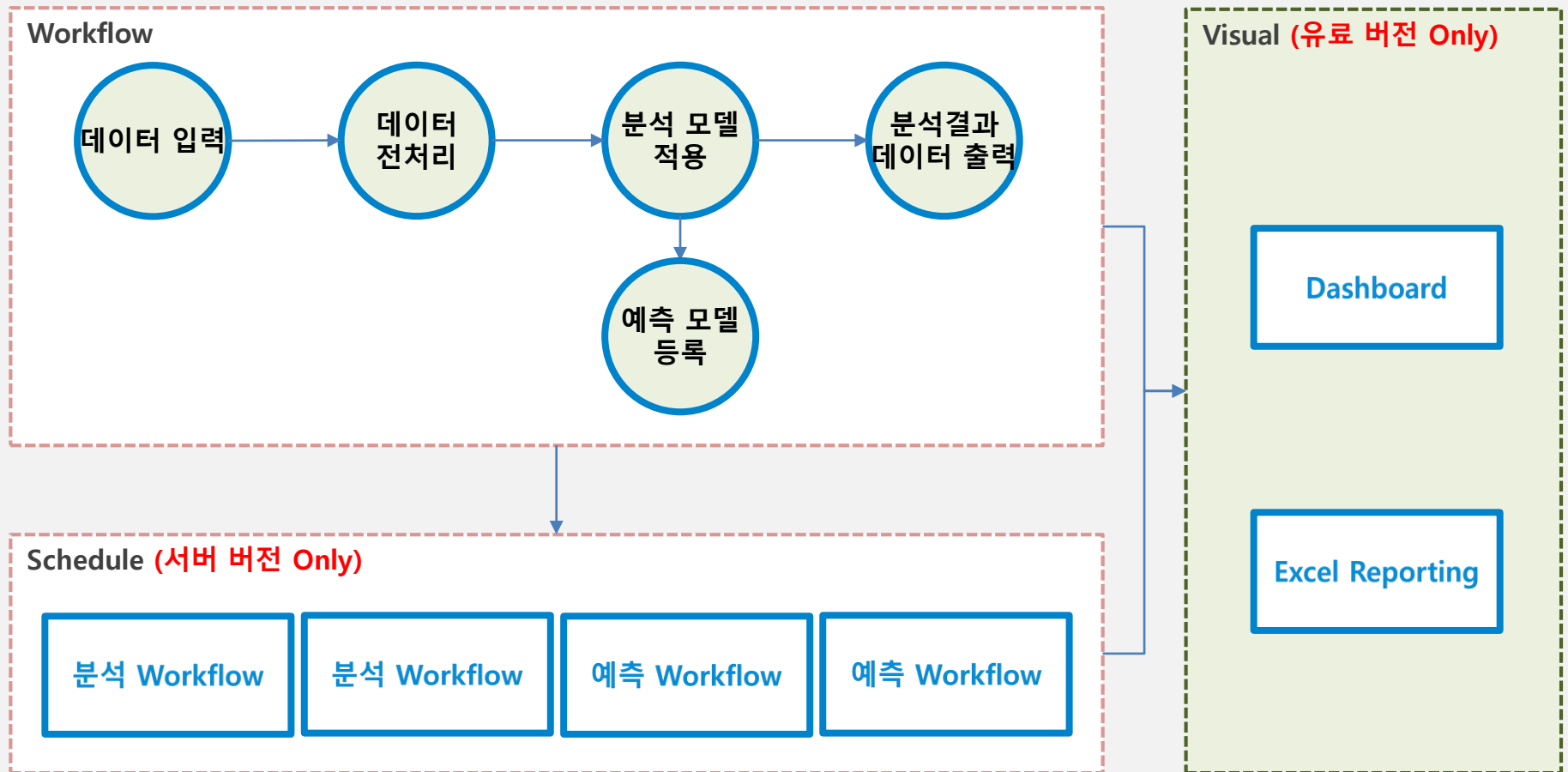




R-Flow

R-Flow 개요

▶▶▶ R-Flow는 **Workflow, Schedule, Visual**로 데이터를 분석하고 시각화 합니다



R-Flow 제품군



R-Flow Desktop

- 데이터 분석 기능을 PC에서 이용
- PC에서 Repository 관리

✓ 무료

- 네이버 소프트웨어 다운로드

✓ 유료

- 온라인 인증
- 연간 이용료



R-Flow Server

- 서버에서 Repository 관리
- 관리, 협업, 배포 기능

✓ On-Premise

- 파트너사를 통한 공급

✓ On-line

- 온라인으로 사용 인증
- 연간 이용료

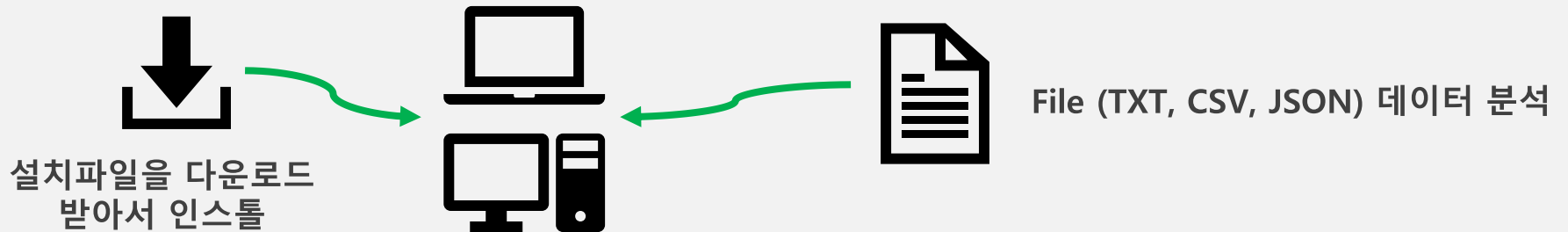
R-Flow 제품군 비교

기능		Windows PC		Linux Server
		무료	유료	
Workflow	분석 모델	O	O	O
	사용자 정의 Task	X	O	O
	입력/출력	File	File & DB	File & DB
	워크플로우 공유 (내보내기/불러오기)	X	O	O
Schedule	Workflow 스케줄 등록	X	X	O
Visual	Dashboard	X	O	O
	Excel Report	X	O	O
Repository		PC	PC	Server
사용자관리, 권한관리		X	X	O
개인/공용 모델 개인/공용 리포트		X	△ (템플릿 제공, 모델공유)	O

※ Dashboard/Excel Report input data : File(csv), DBMS(Oracle, MS SQL, MySQL, Maria)

R-Flow Desktop (무료)

▶▶ Windows PC에서 무료로 사용합니다



1 무료버전 제약사항

분석대상 데이터 입출력은
File(TXT, CSV, JSON)만 가능

DB접속, 스케줄 지원 안함

2 무료버전의 배포

배포사이트(네이버)에서 주기적으로
배포하며, 누구나 재배포 가능

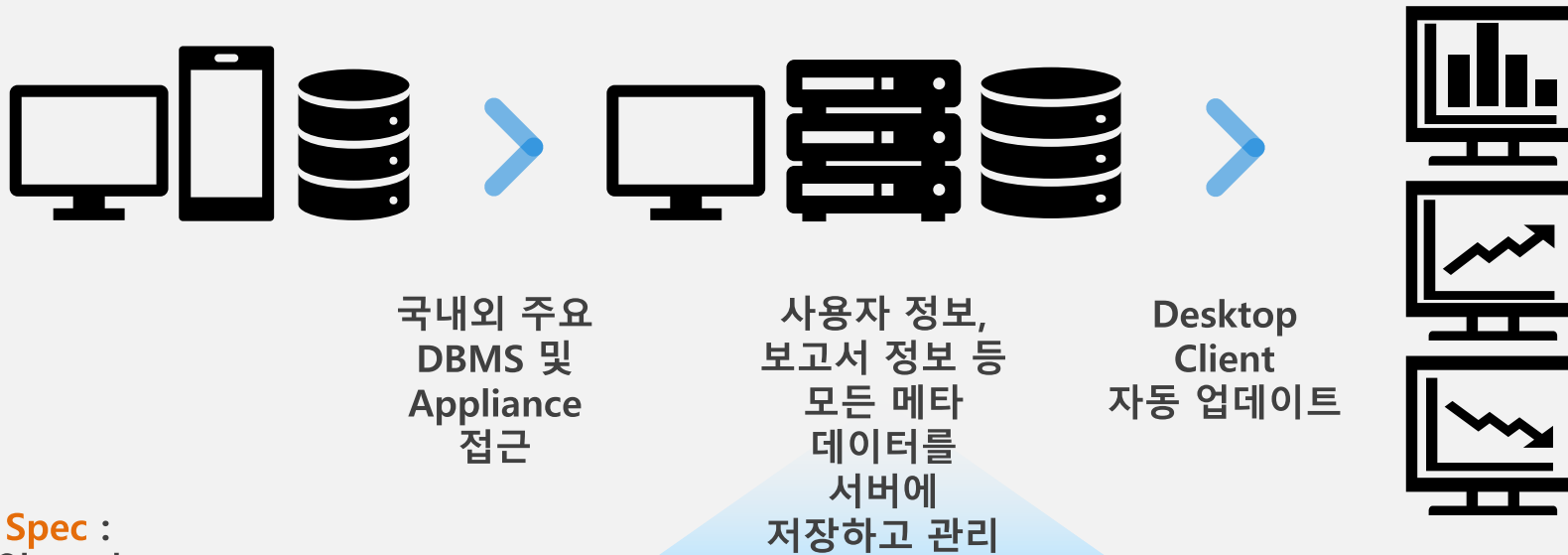
자동 업데이트가 안되며, 사용자가 배포
사이트(네이버)를 방문하여 최신 버전을
다운로드하여야 함

Client Spec :

Windows 7 over
R 3.4 (64bit Core)

※ R 외에 Python,
TensorFlow, Spark를
지원함에 따라 유료/무료
버전 모두 64 비트만 제공

R-Flow Server



국내외 주요
DBMS 및
Appliance
접근

사용자 정보,
보고서 정보 등
모든 메타
데이터를
서버에
저장하고 관리

Desktop
Client
자동 업데이트

Server Spec :

x86 계열 64비트

Linux (Centos 7.x over, Redhat 7.x over, Ubuntu)

16G Mem

3G HDD (프로그램과 R 설치용, 데이터 제외)

R 3.4

Java 1.7 over

Client Spec :

Windows 7 over

64비트

Office 2013 over

R 3.4 (64bit Core)

Analytics Application

WAS(내장형 또는 설치형)

내장형 Repository DBMS

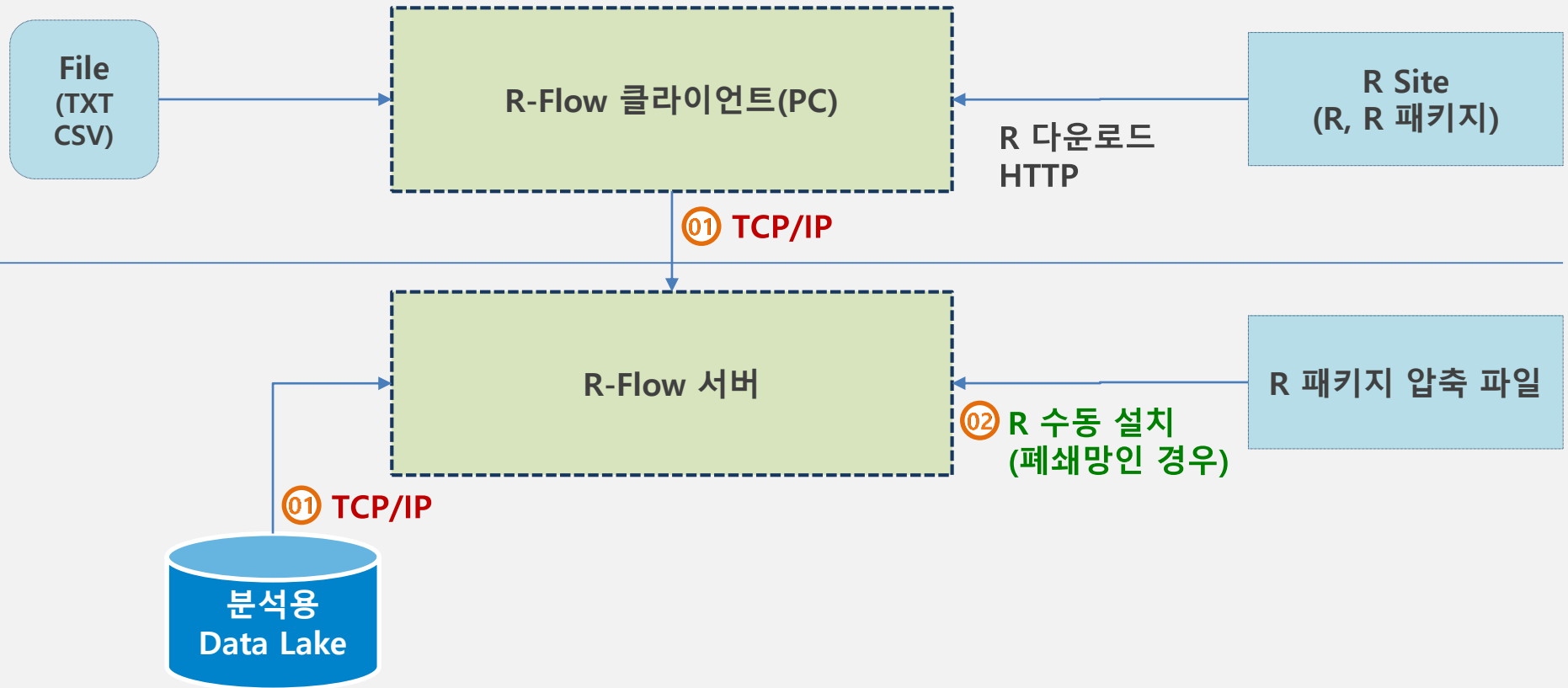
Desktop Client

- ▶ 관리자 기능
보고서 디자인과 배포 기능
로컬모드(PC에서 분석 수행)와
원격모드(서버에서 분석 수행)
지원

※ R과 WAS는 동일 서버에
설치되어야 함

※ 간단한 테스트 분석을 서버 연결
없이 PC내에서 하기 원하는 경우에
로컬모드로 분석

R-Flow Server 인터페이스



① R 자체의 통신방식(TCP/IP)
이를 고려하여, 원천 또는 DW에서 분석대상이 되는
데이터를 별도로 R 분석단(분석용 Data Lake)에
위치시키는 것을 권장

② 파트너사 작업

※ Port
8080: WAS 연결
6311: R Sever 연결

R-Flow와 R, Python, TensorFlow, Keras, Apach Spark

▶▶▶ R-Flow와 R

- ✓ R-Flow는 R이 설치되어 있어야 합니다
- ✓ 한글 환경에서는 R 또는 R Package 자체 특성에 따라 오류가 발생할 수 있으며, 이에 따라 루트 디렉토리에 R을 설치할 것을 권장합니다
- ✓ R 또는 R Package의 업데이트, 추가에 따라 R-Flow도 3~5개월 주기로 업그레이드 합니다. 데이터 구조가 바뀌는 R의 메이저 버전업이 있는 경우에는, 신규 버전의 R에서 R-Flow가 동작하지 않을 수 있습니다

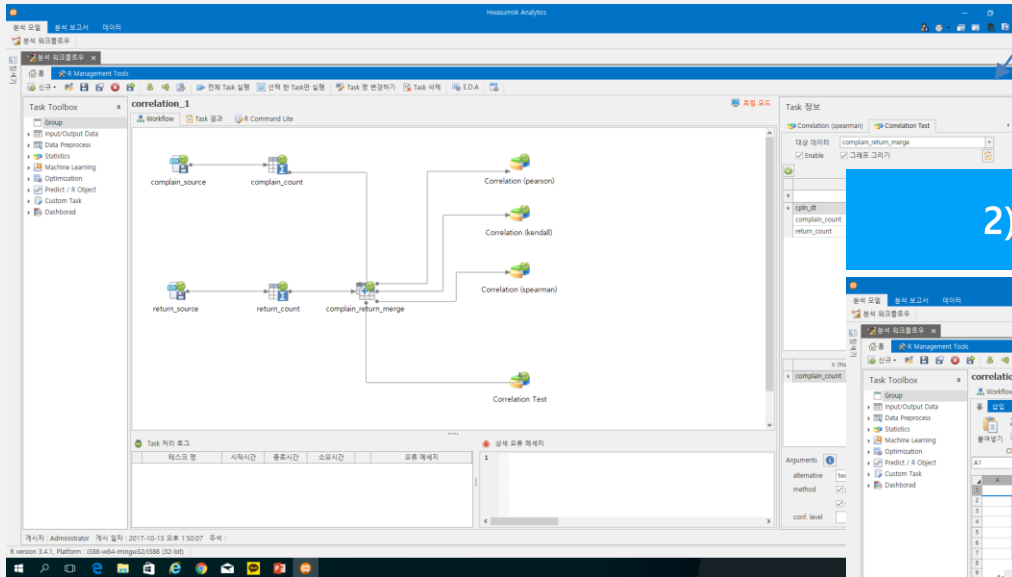
▶▶▶ R-Flow와 Python 등

- ✓ R-Flow에서 Python, TensorFlow, Keras, Spark 작업을 하려는 경우, R과 마찬가지로 해당 플랫폼이 설치되어 있어야 합니다
- ✓ Python, TensorFlow, Keras, Spark 버전업에 따라 데이터 구조 등이 바뀌는 경우에는 R의 메이저 버전업과 마찬가지로 R-Flow의 업그레이드를 기다려야 합니다

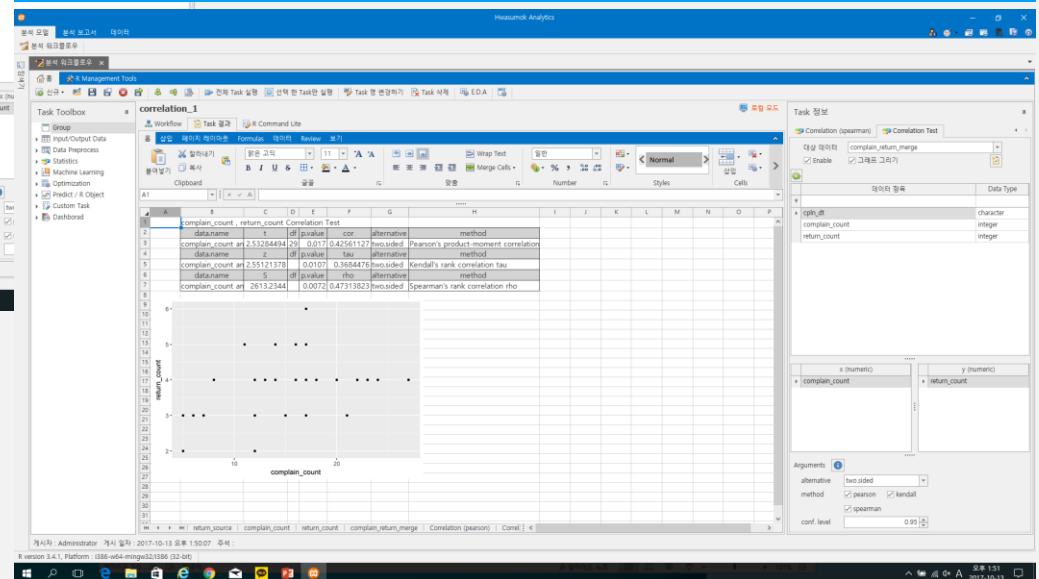
Workflow 작업

▶▶▶ **Workflow** 는 분석을 위한 입력 데이터, 전처리, 분석 모델, 출력 데이터를 정의합니다

1) Workflow를 디자인하고, Task 정보를 설정해주면



2) 데이터 정보와 분석 결과를 자동으로 산출



Workflow 특징

▶▶ 사용의 편의성

- ✓ **R Script 작성 없이 분석** – 아이콘만 선택하여 workflow를 구성하면 자동으로 R 분석
- ✓ **패키지를 찾을 필요 없이 분석** – R 패키지를 찾을 필요 없이, 아이콘으로 제공되는 30여종의 데이터 전처리 기능과 60여종의 통계/머신러닝/최적화 모형을 이용
- ✓ **단순 작업을 자동화** – 데이터 작업 시 반복되는 기술통계, 모형별 default 차트를 기본으로 제공

▶▶ 관리의 편의성

- ✓ **입력, 출력** – 다양한 소스(DB, File, API)로부터 입력받고, 다양한 방식(DB, File, Visual)으로 출력
- ✓ **분석 프로세스 재사용** – 데이터 입력부터 출력까지 모든 단계를 하나의 워크플로우로 정의만 하면, 등록하여 효율적으로 재사용하고 수정
- ✓ **예측 모델 등록** – 학습된 분석 모델은 예측 모델로 등록하여, 예측, 판별에 쉽게 적용

※ 회귀분석 모델 $y=ax+b$ 를 적용하여 $y=2x+1$ 을 얻었다면, 이 식을 예측 모델로 등록하고, 새로운 입력 데이터와 연결하여 예측 데이터를 출력하는 새로운 워크플로우를 정의하여 사용

Workflow – Input & Link

▶▶▶ Input

- ✔ **Text & Web Text & DBMS** – 다양한 유형의 데이터를 연동
- ✔ **Google & SNS** – Google Analytics, Big Query, Google Sheets, Twitter, Facebook, Instagram의 데이터를 수집하여 분석에 이용
- ✔ **Data Service** – Open API, Web page Scraping, Quandle API로 데이터 수집
- ✔ **Stats** – SAS, SPSS, STATA 데이터를 연동

▶▶▶ Link

- ✔ **연계 플랫폼** – Python, Tensorflow, Keras, Spark
- ✔ **1-Workflow, N-Platforms** – R로 전처리 후 Python이나 Tensorflow로 분석하고 결과를 R로 시각화하는 방식으로 응용(R과 Python은 상호 Object 연동도 지원)

Workflow – API 사용 예

The screenshot displays the R-Flow workflow editor. On the left, a 'Diagram' pane shows a workflow starting with 'OAuth for Twitter 1', which branches into 'Twitter Search 1', 'Twitter Retweet 1', and 'Twitter Trend (Preparing) 1'. On the right, the 'Task Information' panel is open for 'Quandl 1'. It includes fields for 'Output Name' (quandlout1), 'Quandl.api_key' (redacted), 'data Type' (table), 'database code' (MER), 'dataset code' (F1), and 'compnumber' (39102). Below these are date filters and checkboxes for 'Enable', 'Draw Graph', and 'Descriptive'. A red note states: '※ When there is a lot of data Draw Graph, Descriptive operation takes a long time.' A blue callout box points to the 'Quandl.api_key' field with the text: 'Quandle 회원가입 후 얻은 key 정보를 입력한 후, 설정값(대상 데이터, 기간)만 바뀌서 다양하게 수집 가능'. Another 'Task Information' panel is shown for 'Twitter Search 1', with fields for 'consumerKey', 'consumerSecret', 'accessToken', and 'accessTokenSecret', all of which are empty. A blue callout box points to this panel with the text: 'Twitter 회원가입 후 얻은 권한 정보를 한번 입력 후, 워크플로우를 저장하면 설정값만 바뀌서 다양하게 수집 가능'. A third panel for 'Twitter Retweet' is partially visible.

Task Information

Quandl 1

Output Name:

Quandl.api_key:

data Type: ☒ table ☐ time-series

database code: dataset code:

compnumber:

☐ time filtering

start_date: end_date:

☒ Enable ☐ Draw Graph ☒ Descriptive

※ When there is a lot of data Draw Graph, Descriptive operation takes a long time.

Diagram

OAuth for Twitter 1

Twitter Search 1

Twitter Retweet 1

Twitter Trend (Preparing) 1

Task Information

OAuth for Twitter 1

consumerKey:

consumerSecret:

accessToken:

accessTokenSecret:

☒ Enable

※ Perform OAuth authentication procedure for twitter connection. For more information, please refer to the following URL.

<https://dev.twitter.com/apps/>

Workflow – R to Python 예

The screenshot displays the R-Flow workflow editor. At the top, a workflow diagram shows the sequence: Python 1 → Input Text 1 → R to Python 1 → Python Script 1 → Python to R 1. The 'R to Python 1' task is selected, showing its configuration panel. This panel includes an 'Enable' checkbox, a table for mapping R objects to Python objects, and a 'Generate Parameter' button. The 'Python Script 1' task is also selected, showing its configuration panel with a 'Python Script' text area and a 'Generate Parameter' button. A 'Task Information' panel at the bottom shows the current task and its parameters.

Task Information (R to Python 1)

Enable ☒

R to Python

R Object Name	Python Object Name
df	r_df

Task Information (Python Script 1)

Enable ☒

Python Script

```
1 from sklearn import datasets
2 import pandas as pd
3 from sklearn.cluster import KMeans
4 import matplotlib.pyplot as plt
5 import seaborn as sns
6
7 model = KMeans(n_clusters=@k, algorithm='auto')
8 model.fit(r_df)
9 predict = pd.DataFrame(model.predict(r_df))
10 predict.columns=['predict']
```

Generate Parameter

Parameter Name	Param. Value
@k	3

Annotations:

- R에서 Python으로 전달할 Object를 정의** (Define the object to be passed from R to Python): Points to the 'r_df' entry in the 'Python Object Name' column of the 'R to Python' table.
- Script에서 사용하는 파라미터를 자동탐지** (Automatic detection of parameters used in the script): Points to the '@k' parameter in the Python script.
- 설정값을 변화시켜가며 실행할 수 있음** (Can be executed while changing the setting value): Points to the 'Generate Parameter' button.

Workflow – 데이터 전처리

Data Preprocess

Data Cleansing

- Missing Data
- Remove by Rule
- Remove by Correlation
- Remove Duplicate Data

Data Convert

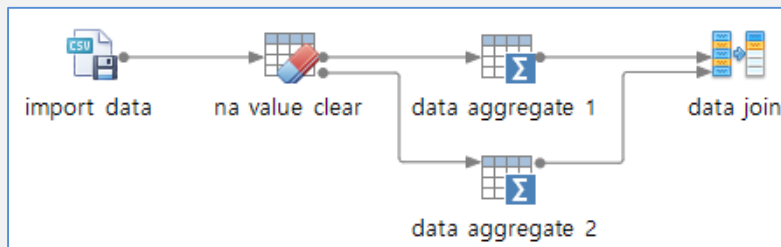
- Change Colum Name
- Conv. Data Type
- Conv. data.frame
- Create data.frame
- Create Dummy
- Data Anonymize
- Data Categorize
- Data Encoding (beta)
- Data Imputation
- Data Scaling
- Data Scaling by Scale Info.
- Derived Variable

String

- Merge String
- Replace String
- Split String
- Trim Whitespace

- Data Subset
- Data Filter
- Data Aggregate
- Data Sort
- Data Join
- Data Merge
- Data Merge (2 data.frame)
- Data Sampling

▶▶ 데이터 전처리는 더 이상 가장 많은 시간을 소요하는 작업이 아님



03. Data Aggregate

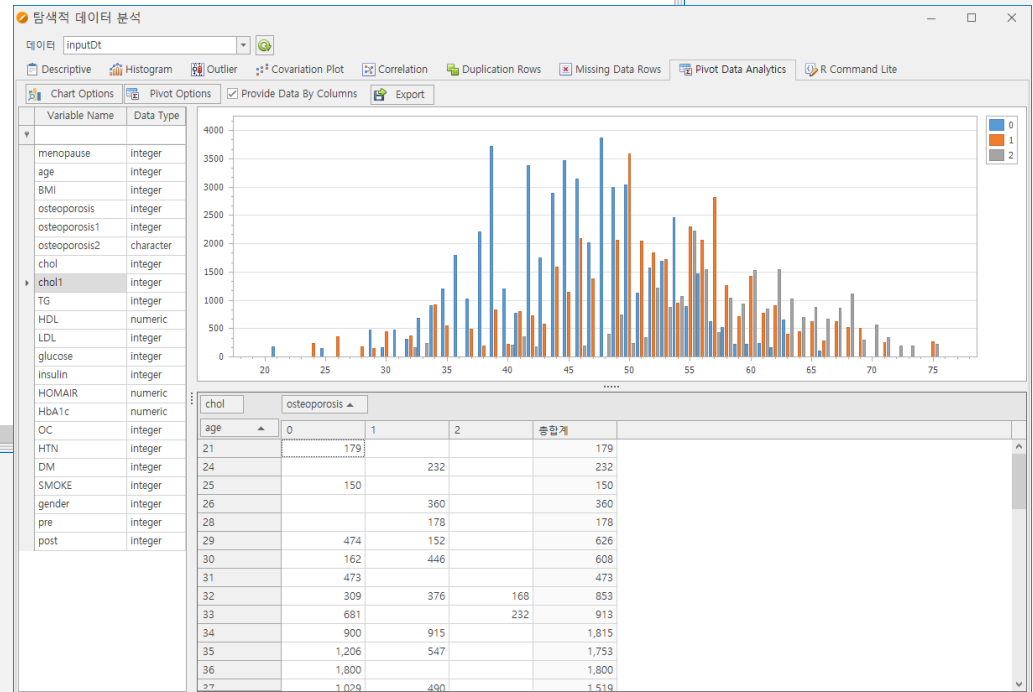
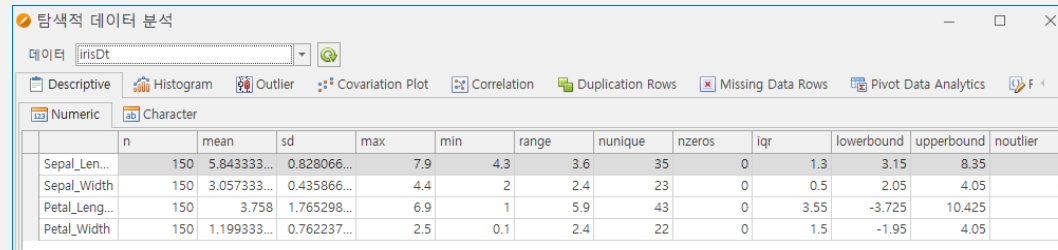
Workflow Task Result R Command Lite

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	df.agg Numeric Descriptive												
2			n	mean	sd	max	min	range	nunique	nzeros	iqr	lowerbound	upperbound
3		osteoporos	3	1	1	2	0	2	3	1	1.25	-1.375	3.375
4		donor	3	204	34.04165634	282	115	167	3	0	116.75	-10.125	423.625
5		bchol	3	195.9039133	5.878497615	202.9913043	189.2553191	13.7359852	3	0	0.091344147	180.2232015	211.3652265
6		fchol	3	186.1394397	7.713665776	194.1478261	178.7588652	15.38896084	3	0	0.8911812204	168.7675283	203.1974453
7		TG	3	442	38.97435054	484	407	77	3	0	40.25	360.625	519.875
8		chol.gap	3	9.764473511	0.842542034	10.4964539	8.843478261	1.65297564	3	0	1.245751493	7.529856076	12.09359838

Workflow – 탐색적 데이터 분석(EDA)

▶▶ EDA는 별도의 창에서 실행되어 분석과 동시에 탐색함

- ✓ Descriptive
- ✓ Histogram
- ✓ Boxplot/Outlier
- ✓ Covariation Plot
- ✓ Correlation
- ✓ Duplication Rows
- ✓ Missing Data Rows
- ✓ Pivot Data Analytics



Workflow 분석 모델

분석 모델

Statistics

Machine Learning

Optimization

✓ **다양한 모델** – 60여종의 분석 모델을 선별하여 제공하며, 최신 인기 모델을 주기적으로 반영

✓ **즉시 사용 가능 모델** – 모델 분석 시 선택 옵션을 자동 제공하며, 기본 분석결과 외에 추가적인 분석결과도 제공

※ 회귀분석인 경우 stepwise와 같은 변수선택 기능, LP인 경우 최적해뿐만 아니라 민감도 분석까지 기본적으로 제공

✓ **모델 별 검증 포함** – 모델과 입력 데이터 유형에 맞는 검증까지 한번에 제공

※ 모델별로 적합한 훈련과 테스트 할당, Confusion Matrix, ROC 등 각종 검증 지표를 자동 산출하여 제공

Workflow 분석 모델 - Statistics

▶▶▶ 전통적인 통계 모형 - 기본적인 검정에서 최신 회귀모형까지 제공

Statistics

- ▶ Correlation Analysis
- ▶ Hypothesis Tests
- ▶ Analysis of Variance
- ▶ Analysis of Covariance
- ▶ Regression Analysis
- ▶ Time Series Analysis
- ▶ Survival Analysis
- ▶ Factor Analysis

Hypothesis Tests

- ▶ One Sample T-Test
- ▶ Paired T-Test
- ▶ Two Sample T-Test
- ▶ One Sample Wilcoxon-Test
- ▶ Two Sample Wilcoxon-Test
- ▶ One Sample Chi-square Test
- ▶ Two Sample F Test
- ▶ One Sample Proportions Test
- ▶ N Sample Proportions Test
- ▶ Goodness of Fit test
- ▶ Test of Independence
- ▶ Test of Homogeneity
- ▶ Correlation Test
- ▶ Kolmogorov-Smirnov Test
- ▶ One Sample Z-Test
- ▶ Two Sample Z-Test
- ▶ Test for Normality

Regression Analysis

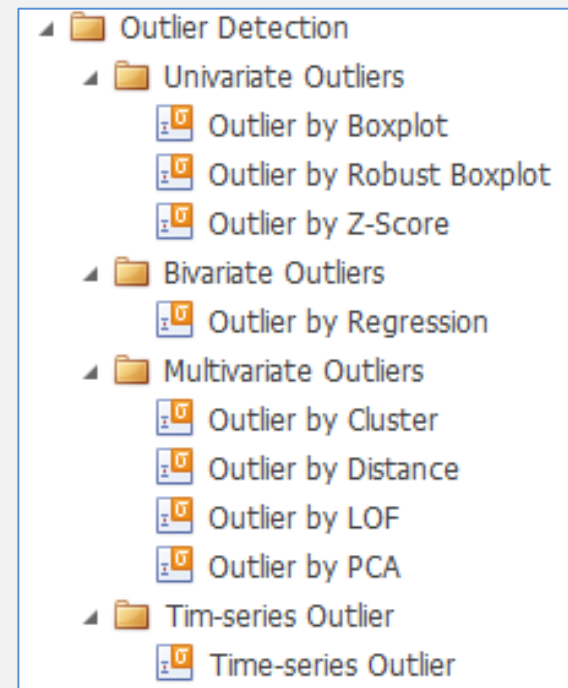
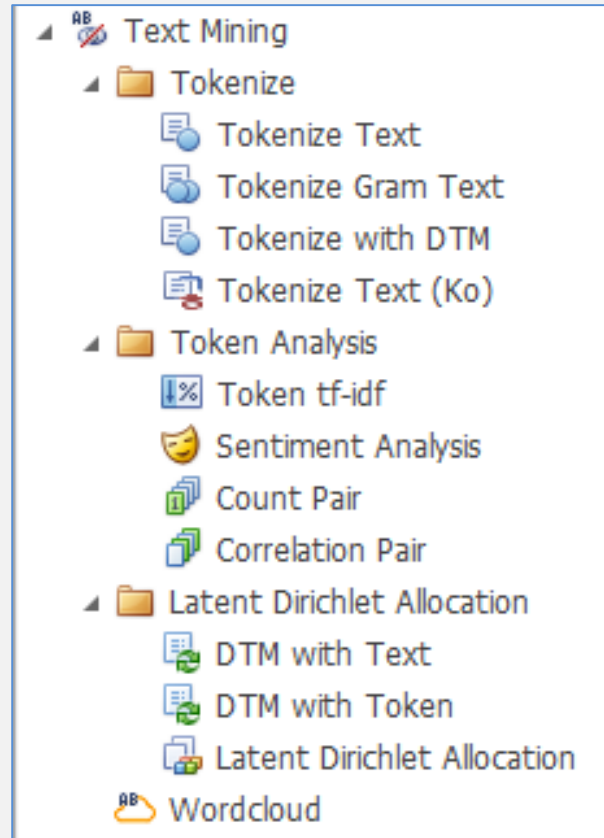
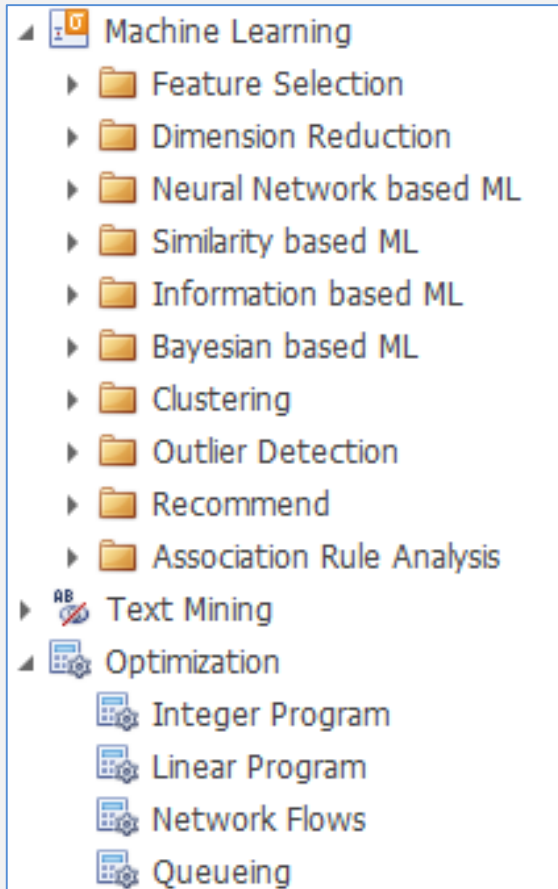
- ▶ Lasso, Elastic-Net, Ridge
 - ▶ Optimal lambda
 - ▶ Lasso, Elastic-Net, Ridge
- ▶ Multiple Linear Regression
- ▶ Partial Least Squares Reg.
- ▶ Principal Component Reg.
- ▶ Simple Linear Regression

Survival Analysis

- ▶ Kaplan Meier Curve
- ▶ Log-Rank test
- ▶ Cox Regression

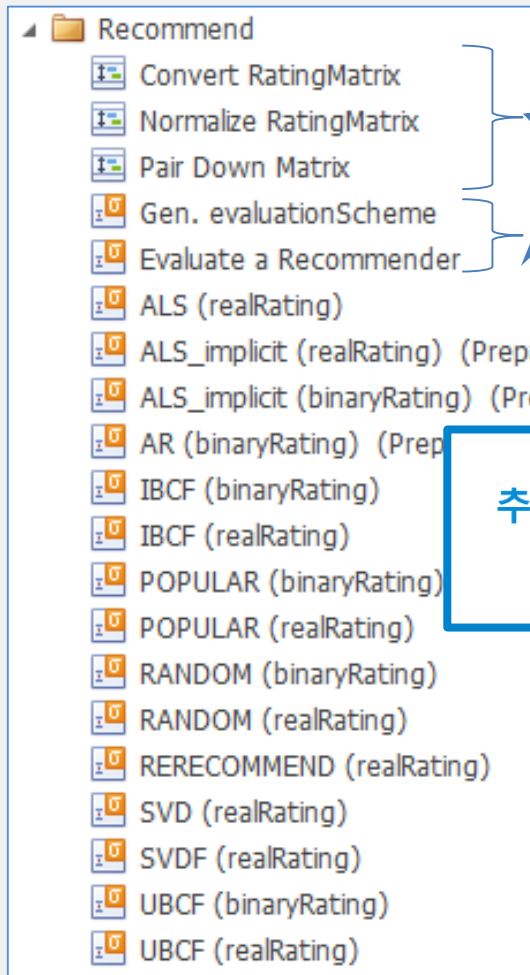
Workflow 분석 모델 – Machine Learning, Optimization

- ▶▶ 머신러닝과 최적화 모형을 제공하는 한편, 주제별 특화된 모형(텍스트 마이닝, 이상치 탐지 등)도 선별하여 제공



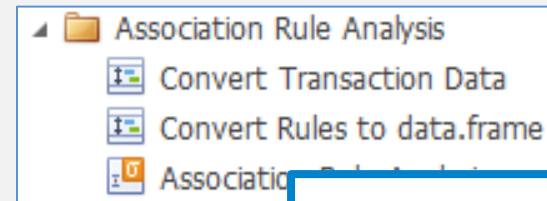
Workflow 분석 모델 - 모형 특화 데이터 처리

▶▶ 모형에 따라 특화된 데이터 전/후 처리가 필요한 경우에 해당 기능을 함께 제공

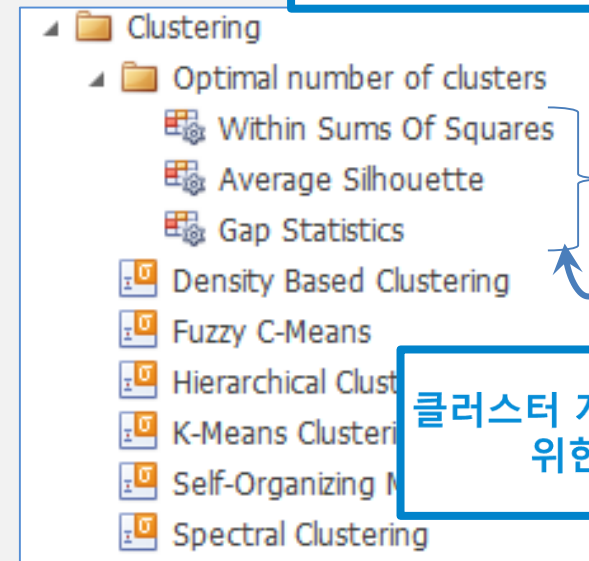


추천 모형에서 필요로
하는 RatingMatrix로
데이터 처리

추천 모형에 특화된
검증을 처리



Association 분석 시 입력
데이터 형식 처리와 결과 값의
DB구조로의 저장을 위한
데이터 후처리



클러스터 개수 최적화를
위한 처리

Workflow – 사용자 정의 Task (유료 버전)

▶▶▶ 사용자가 직접 Task를 만들어서 사용

- ✔ 나만의 워크플로우를 구성하는 것처럼 나만의 Task를 정의하여 사용
- ✔ Task 이용 시 설정할 파라미터와 워크시트에 나타낼 결과 데이터도 사용자가 정의

Custom Task

Custom Task List

Task Name
R1C

Task Information

Task Title.

Description

Date of Publication

Task Script Set Task Parameter Draw data.frame in Result Sheet

1

Task 기능을 스크립트로 정의

설정할 파라미터를 정의

Task 실행 시 저장할 결과를 정의

Workflow 실행 - 로컬 모드 vs. 서버 모드 (서버 버전)

▶▶ 사용자가 직접 Task를 만들어서 사용

✓ 로컬 모드 사용 예 - 간단한 테스트 워크플로우는 내 PC에서 실행

✓ 서버 모드 사용 예 - 검증된 공용 워크플로우나 대용량 컴퓨팅 자원이 필요한 경우는 서버에서 실행



Schedule (서버 버전)

- ▶▶▶ 주기적인 모델 학습이나 주기적인 예측 모델 실행이 필요한 경우, 해당 Workflow를 서버에 등록하여 주기적으로 자동실행합니다

