Techtonic 2018

-

Thu . Nov 15

-

SAMSUNG SDS Tower
West Campus B1F
Magellan Hall /Pascal Hall

SAMSUNG SDS
Realize your vision

머신러닝이 쉬워진다!

오픈소스 AI분석 플랫폼
Brightics Studio

삼성SDS 민승재 마스터

Techtonic 2018

Agenda

- Data Analytics Trend
- Brightics Studio
- Demo
- Brightics AI

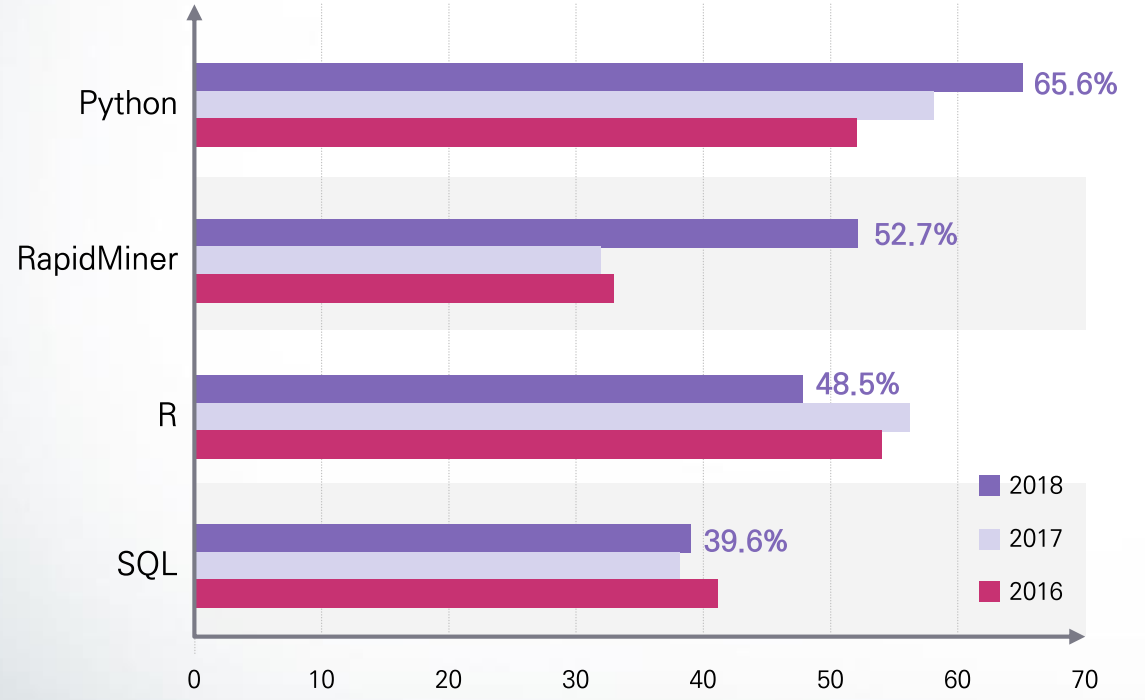
오픈소스 AI분석 플랫폼 Brightics Studio

Data Analytics Trend

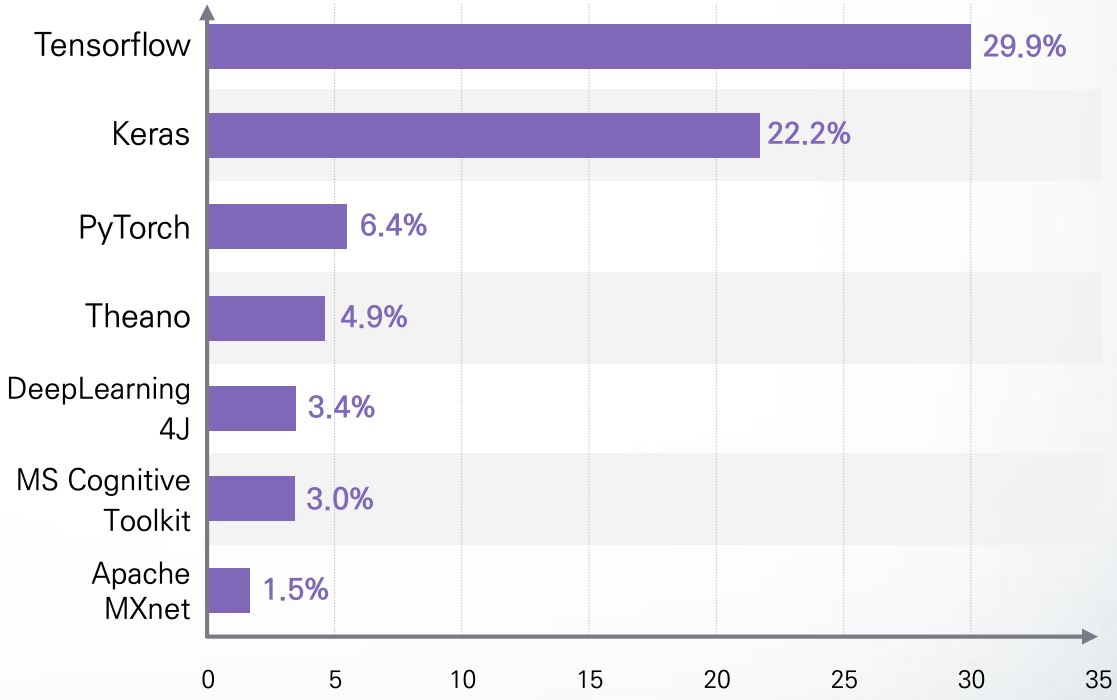
데이터분석 및 머신러닝 관련 S/W 선호도 조사

응답자 평균 7.0 개의 선호 S/W를 응답하였으며, 2017년의 6.75개 대비 소폭 상승됨

Machine Learning SW Poll, 2016–2018



Deep Learning Tools



※ Source : KDnuggets – Top Software for Machine Learning in 2018: Trends and Analysis

업종전문가를 위한 쉬운 Visual 분석환경

Professional Data Scientist 用 (분석 전문가)

R Code

```r
#Load Train and Test datasets

#Identify feature and response variable(s) and values must be numeric and numpy arrays

x_train <- input_var
y_train <- target_v
x_test <- input_var
x <- cbind(x_train,
# Train the model u
linear <- lm(y_trai
summary(linear)

#Predict Output

predicted= predict(
```

Python Code

```python
import matplotlib.pyplot as plt
import numpy as np
from sklearn import datasets, linear_model
from sklearn.metrics import mean_squared_error, r2_score
diabetes = datasets.load_diabetes()

diabetes_X = diabetes.data[:, np.newaxis, 2]

# Split the data into training/testing sets
diabetes_X_train = diabetes_X[:-20]
diabetes_X_test = diabetes_X[-20:]
regr = linear_model.LinearRegression()

# Train the model using the training sets
regr.fit(diabetes_X_train, diabetes_y_train)

# Make predictions using the testing set
diabetes_y_pred = regr.predict(diabetes_X_test)

# Plot outputs
plt.scatter(diabetes_X_test, diabetes_y_test,  color='black')
plt.plot(diabetes_X_test, diabetes_y_pred, color='blue', linewidth=3)
```

Citizen Data Scientist 用 (업종 전문가)

오픈소스 AI분석 플랫폼의 필요성

※ SDS AI분석플랫폼인 Brightics AI 사용자들의 Feedback기반

업종 전문가

"Coding을 할 줄 몰라도
내 업무에 AI를 적용해 보고 싶다."

개발자

"목적별로 적합한 머신러닝,
딥러닝 툴을 연계할 수 있고
분석모델을 운영에 쉽게
Deploy할 수 있는 기능이 필요"

Student / MBA

"라이선스 제한있는 상용제품보다
학기후에도 활용가능한 오픈소스 플랫폼으로
AI실습강의가 이루어졌으면 좋겠다"

오픈소스 AI분석 플랫폼 Brightics Studio

Brightics Studio

오픈소스 Brightics Studio의 주요 특징

1 Workflow Modeler
시각화된 분석 모델링 환경

2 User defined Function
사용자 정의함수로 개인화된
분석환경 제공 및 생태계 확산

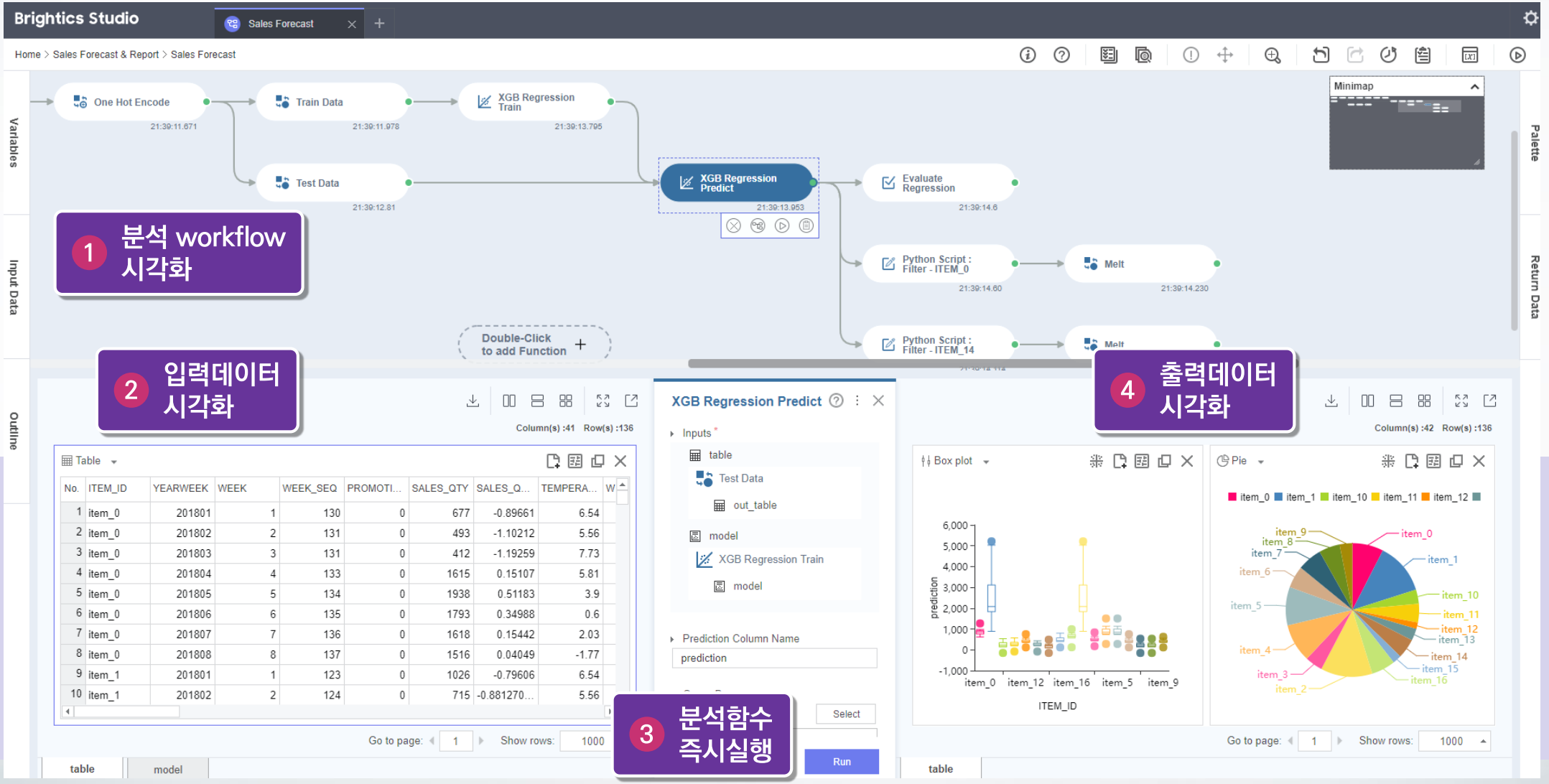
3 Ready to USE
분석모델의 입력/공유/배포기능
기본제공

4 Deep Learning
시각화 된 Deep Learning
모델링/분석 환경

5 3rd Party I/F
Public Cloud의 ML/DL tool 연계
분석모델링 환경

6 Open Source
Apache License,
Version 2.0 으로 배포

Feature ❶ Workflow Modeler – 시각화된 분석 모델링 환경



Feature 1 List of Machine Learning Functions

Descriptive Analytics

Manipulation

- Replace Missing Number
- Replace Missing String
- Simple Filter
- Sort

Statistics

- Bartlett's Test
- Correlation
- One Way ANOVA
- Pairplot
- Profile Table
- Tukey's Range Test

Transform

- Concatenate
- Melt
- Merge
- Pivot Table
- Random Sample
- Delete Missing Data
- Train Test Split
- Select Column

Extraction

- Add Shift
- Add Column If
- Label Encode (with Model)
- One Hot Encode (with Model)
- PCA (with Model)
- Scale (with Model)
- Add Column

Predictive Analytics

Regression

- Decision Tree Regression Train
- Decision Tree Regression Predict
- GLM Train
- GLM Predict
- Linear Regression Train
- Linear Regression Predict
- XGB Regression Train
- XGB Regression Predict

Classification

- Decision Tree Classification Train
- Decision Tree Classification Predict
- Logistic Regression Train
- Logistic Regression Predict
- Support Vector Classification Train
- Support Vector Classification Predict
- XGB Classification Train
- XGB Classification Predict

Clustering

- Kmeans Train & Predict
- Kmeans Train & Predict (Silhouette)
- Kmeans Predict

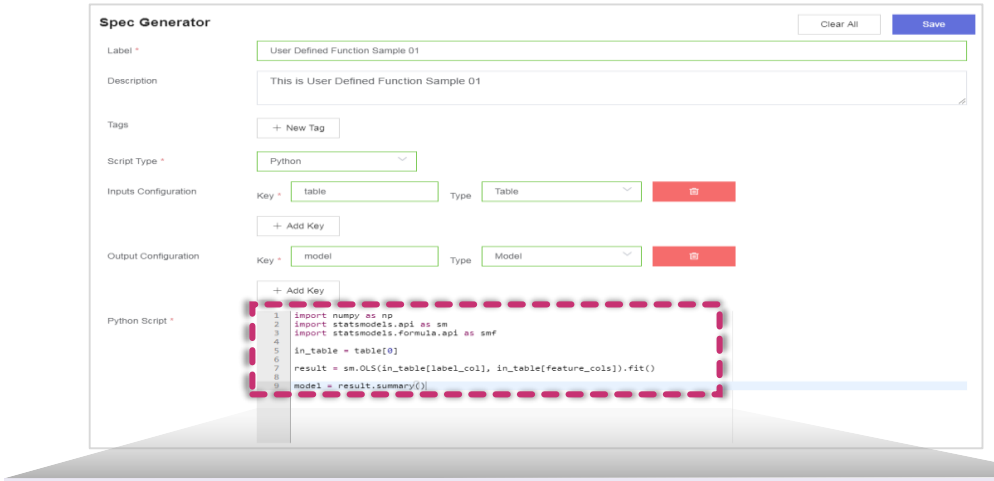
Evaluation

- Evaluate Classification
- Evaluate Regression
- Plot ROC Curve and PR Curve

Feature ❷ User Defined Function (UDF)

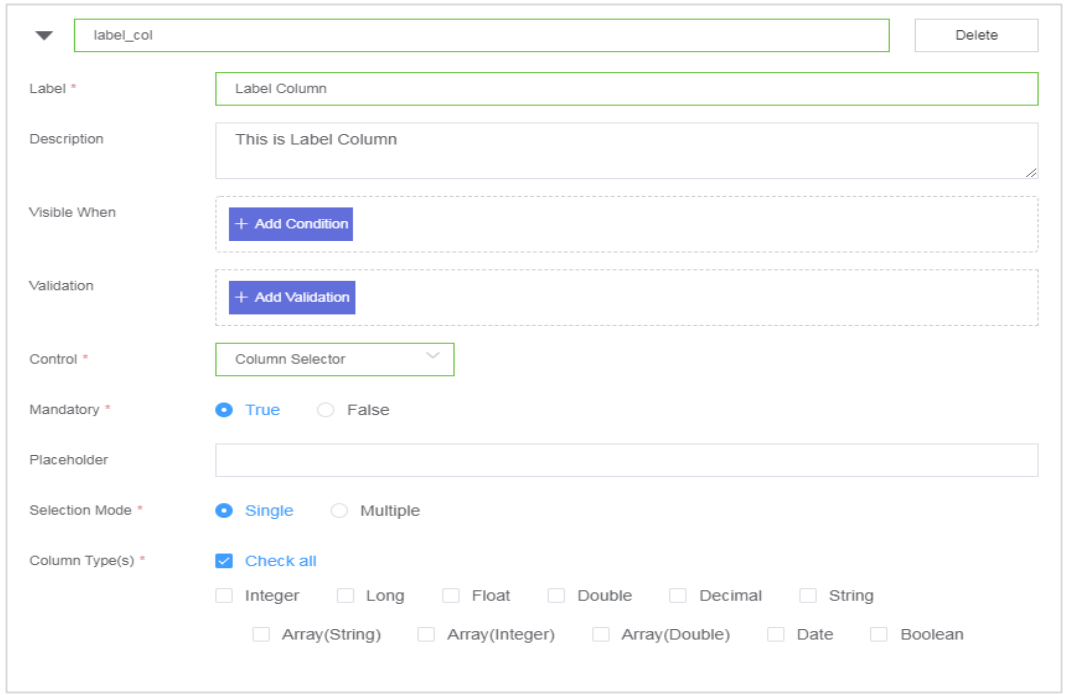
분석전문가의 전문 알고리즘을 쉽게 Workflow로 통합하여 **시각화된 분석환경으로 제공**

Write a Python Script




```
import numpy as np
import statsmodels.api as sm
import statsmodels.formula.api as smf
in_table = table[0]
result = sm.OLS(in_table[label_col], in_table[feature_cols]).fit()
model= result.summary()
```

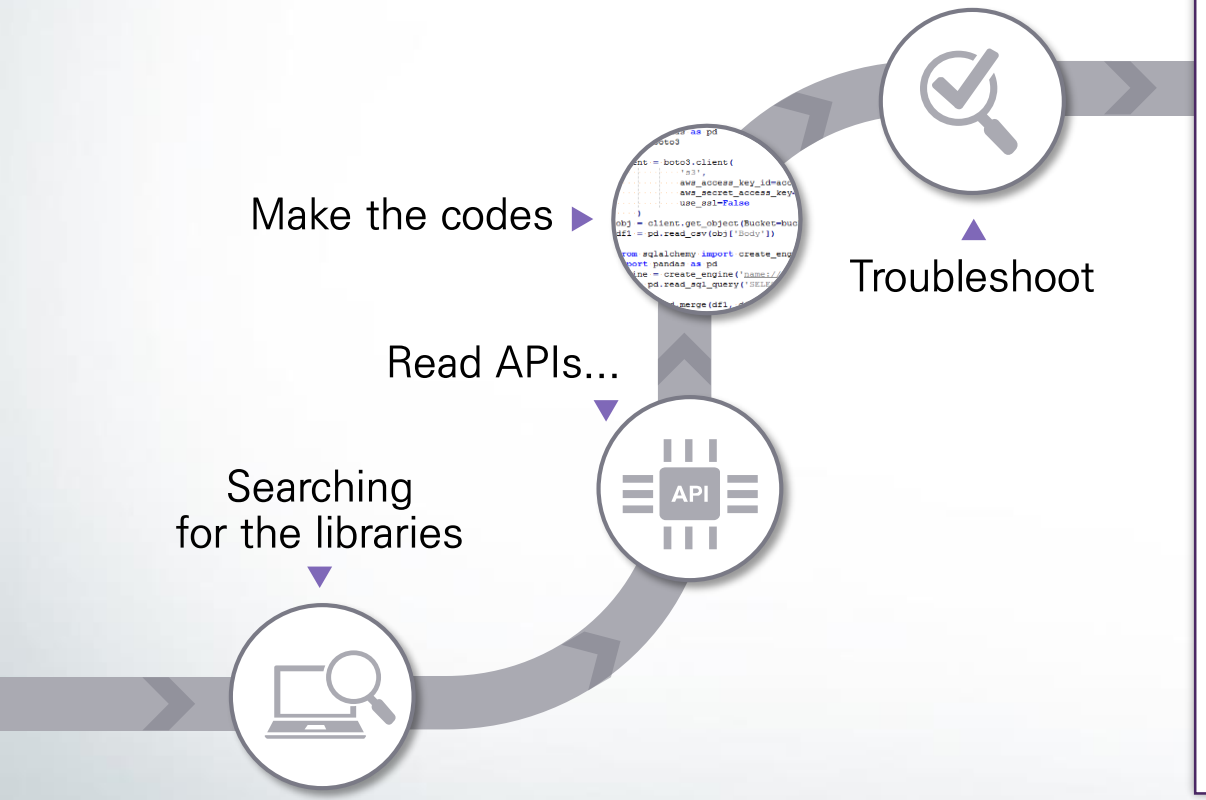
Set Parameters for your UDF



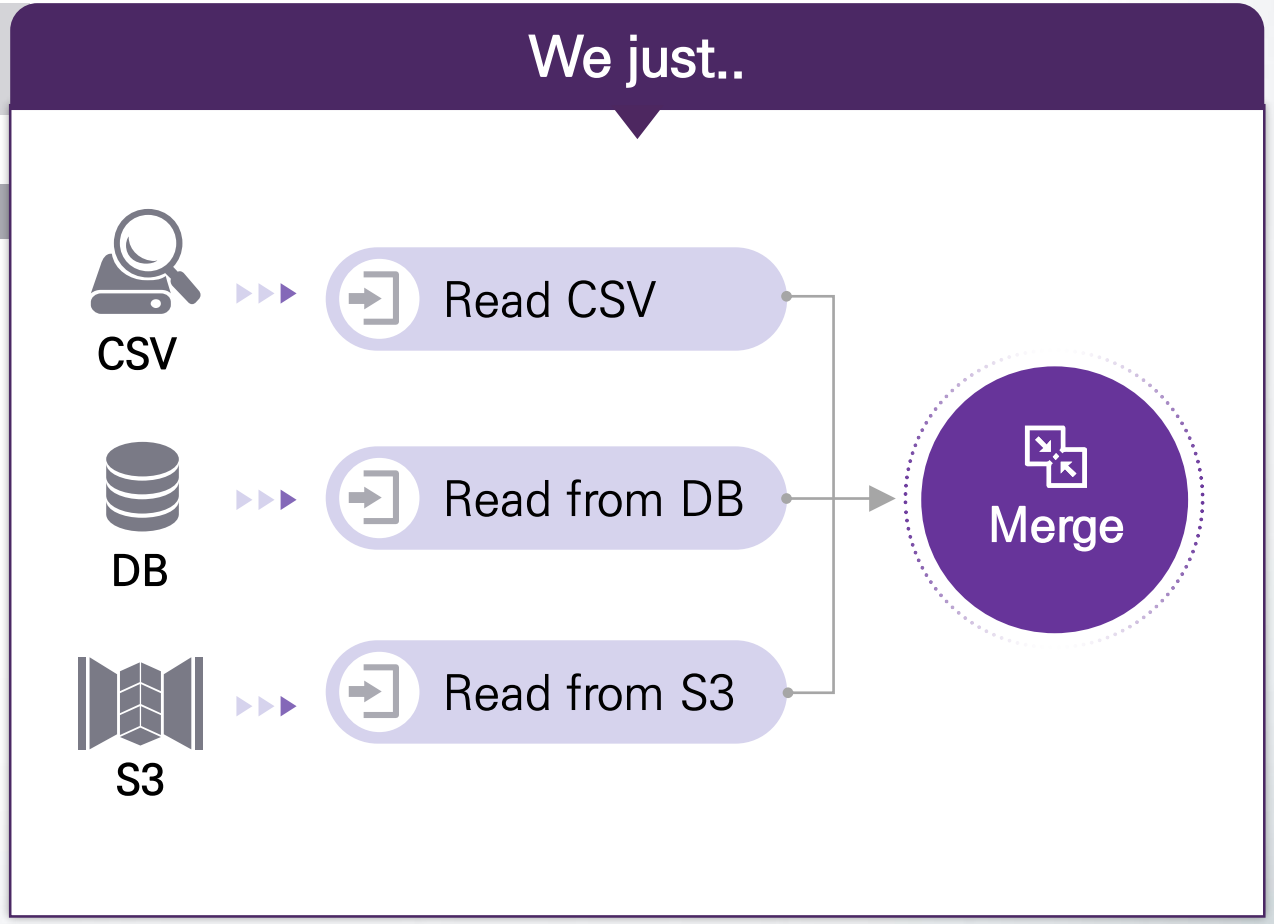
Feature ❸ Connection – easy to connect

데이터 준비를 위해 필요한 **주요 데이터 connection을 단일 함수(기능)**로 제공

You need..

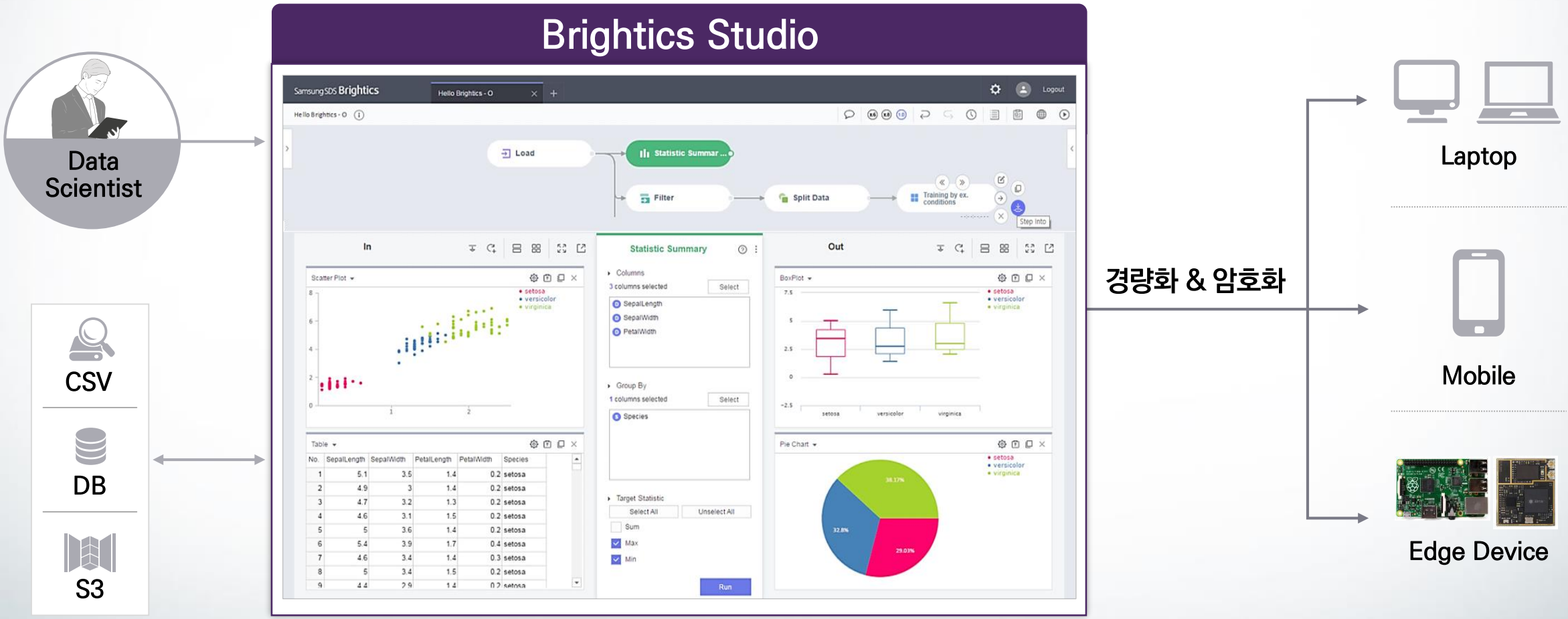


We just..



Feature ❸ Deploy – Lightweight deployable model

Edge Analytics를 지원하는 **경량화 분석모델 배포 제공**

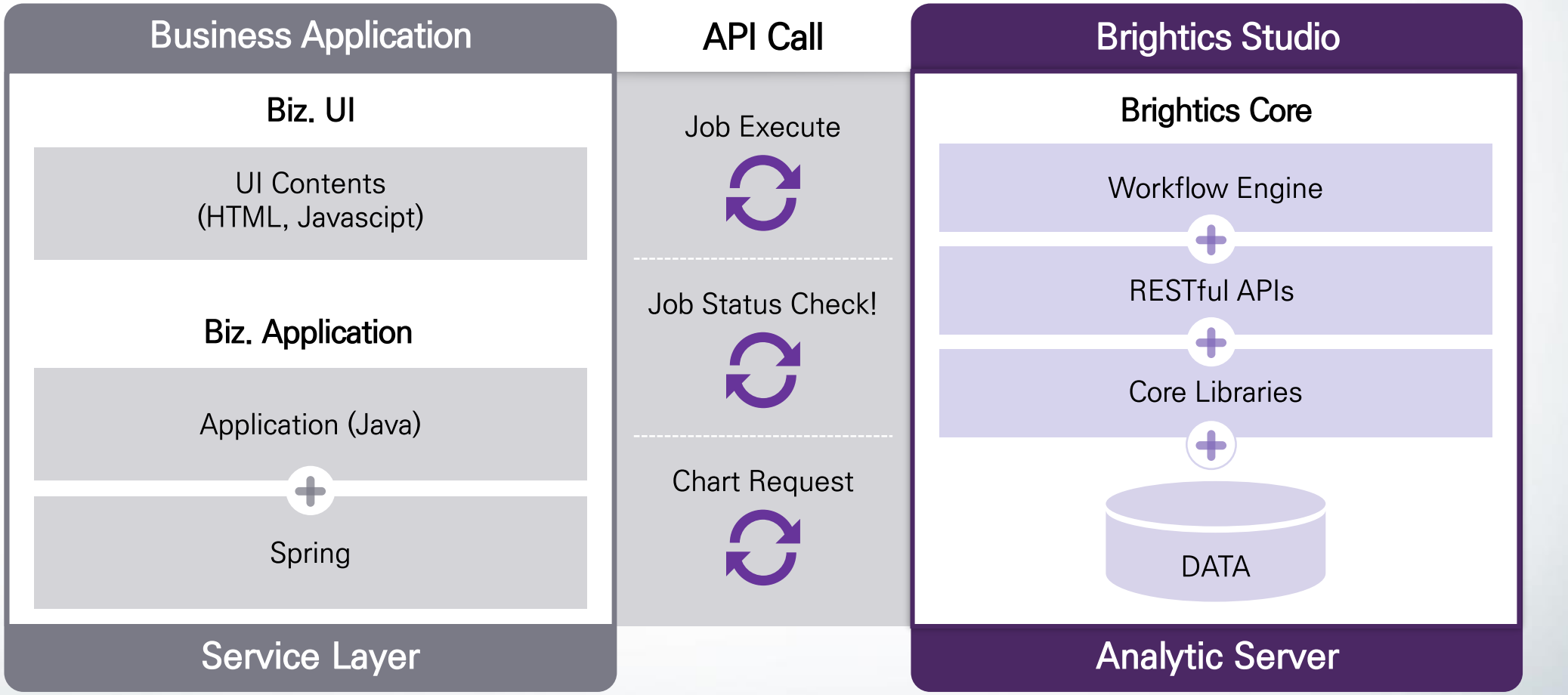


Feature ❸ Share – RESTful API

분석결과 (차트/데이터)의 활용을 위한 **RESTful API** 제공

Predict Results

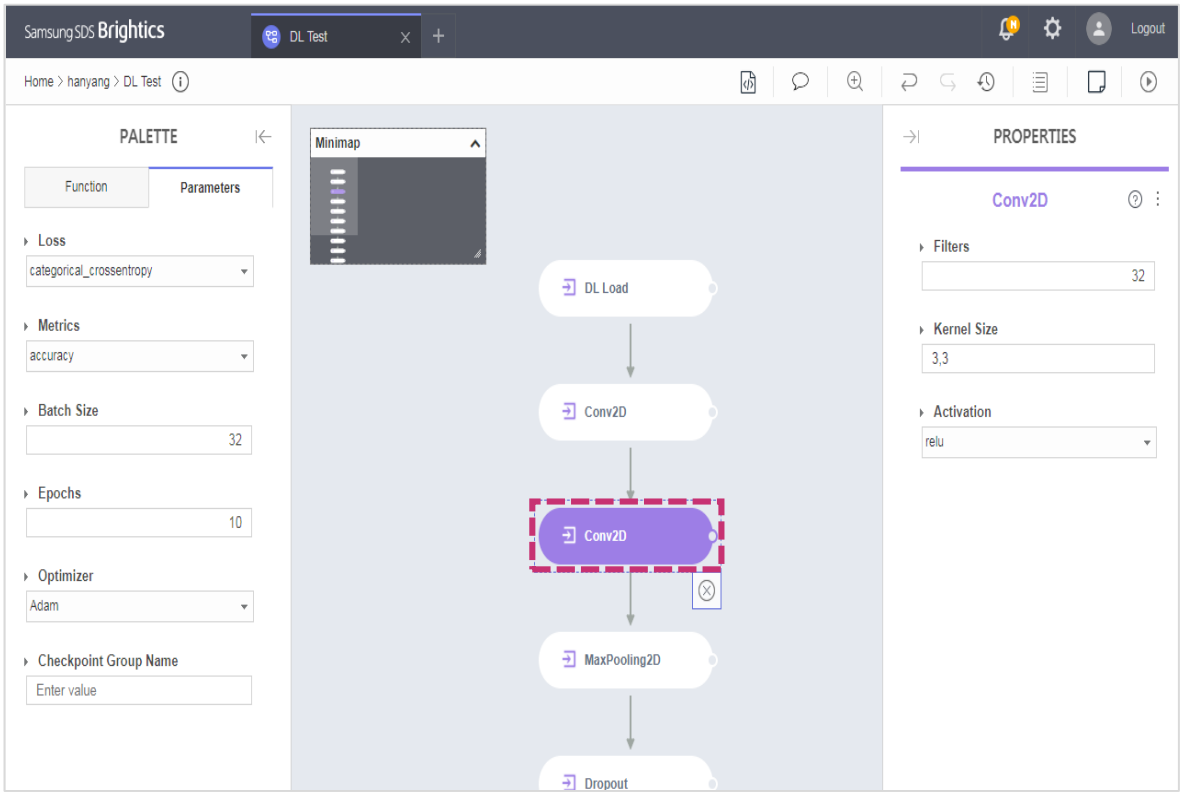
Business Report



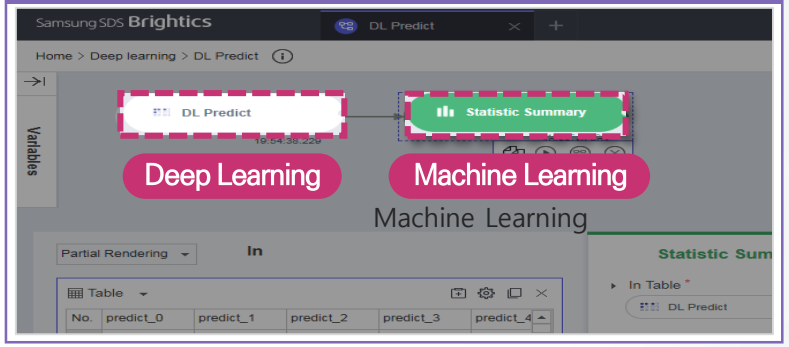
Feature ❹ Deep Learning 연계

시각화 된 Deep Learning 모델링/분석 환경 제공

Deep Learning Work-flow시각화



머신러닝 + 딥러닝 통합분석



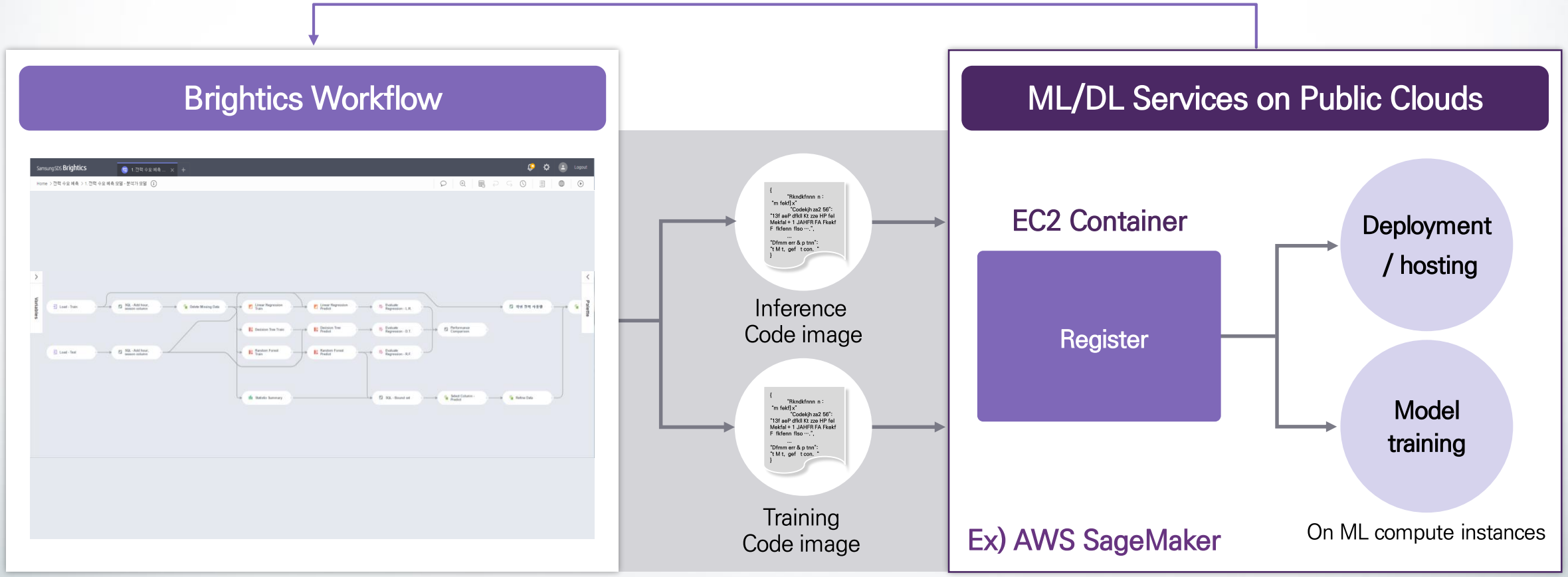
3rd Party Tool 지원

Python code Export

```
model.add(Conv2D(filters=32, kernel
model.add(MaxPooling2D(pool_size=(2
model.add(Dropout(rate=0.5))
model.add(Flatten())
model.add(Dense(units=128,activation
model.add(Dropout(rate=0.5))
```

Feature ❺ 3rd Party Tool 연계 – Public Cloud DL/ML Tool

각 Public Cloud(AWS, MS Azure, GC 등) 에서 제공하는 DL/ML Tool 을
Brightics 의 **시각화 된 workflow 와 연동**하여 분석모델링 환경 제공



Feature ❻ Open Source

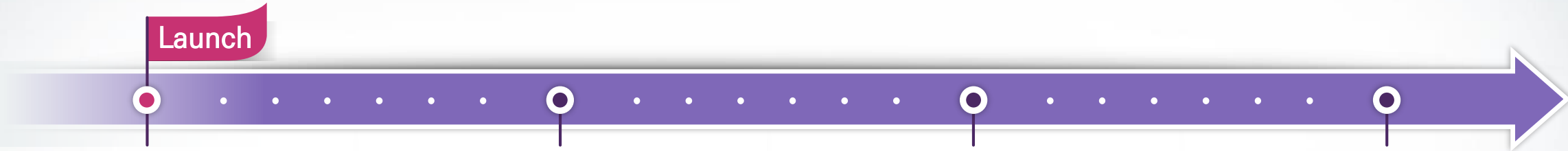
Brightics Studio Download

www.brightics.ai

Github : github.com/brightics/studio

※ Licensed under the Apache License, Version 2.0

Roadmap



Visual Modeler for ML

- 시각화된 분석 모델링 환경
- 분석함수 40 종, Chart 29 종
- User Defined Function (UDF)

Connection – External Data

- 시각화된 함수(기능) 제공
- CSV, RDB, S3 등

Share – RESTful API

- Analytic Server 기능 제공
- 분석모델 실행 및 조회 등

Visual Modeler for DL

- 시각화된 DL 모델링 환경
- 모델 최적화 및 추천

Model Deploy – Light & Secure

- Edge Analytics 기능 제공
- Linux 기반 Edge Device 지원

Public Cloud Service 연계

- Amazon Web Service,
  MS Azure, Google Cloud 등

Python Query Engine

- DB copy가 필요 없이 Pandas Dataframe Query 기능 제공

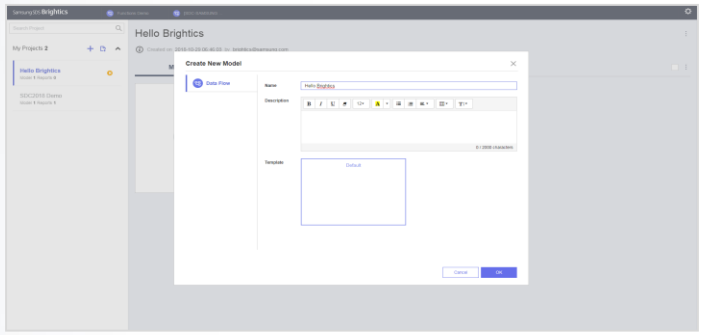
오픈소스 AI분석 플랫폼 Brightics Studio

Demo

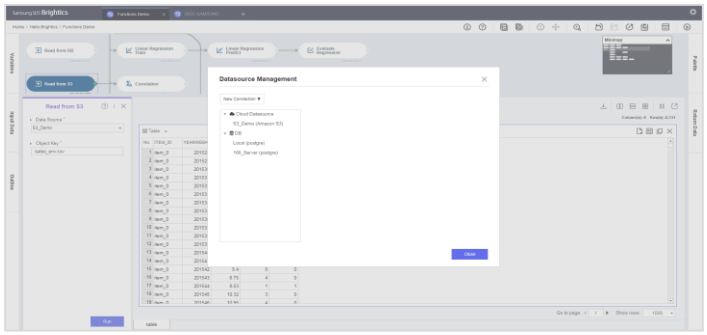
Brightics Studio Demo Scenario

플랫폼에서 분석모델 생성, 시각화, Report, Export

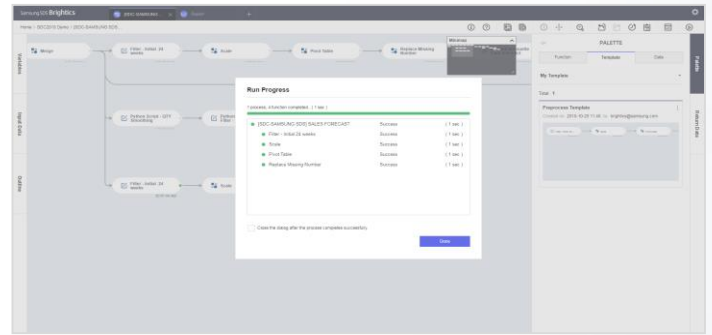
Create a model



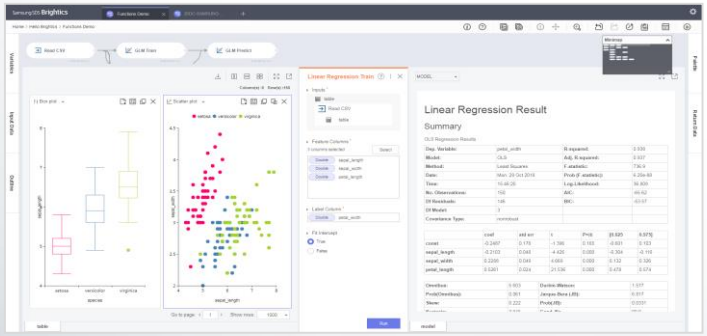
Select data sources



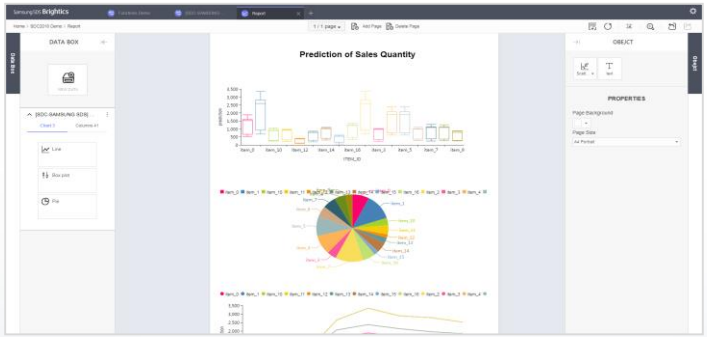
Build a model & run



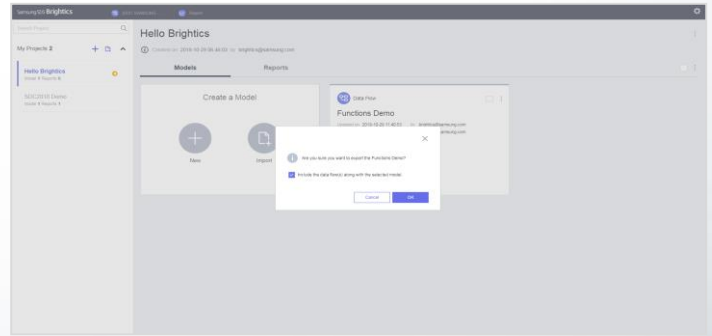
Visualization & summary



Create a report



Export a model

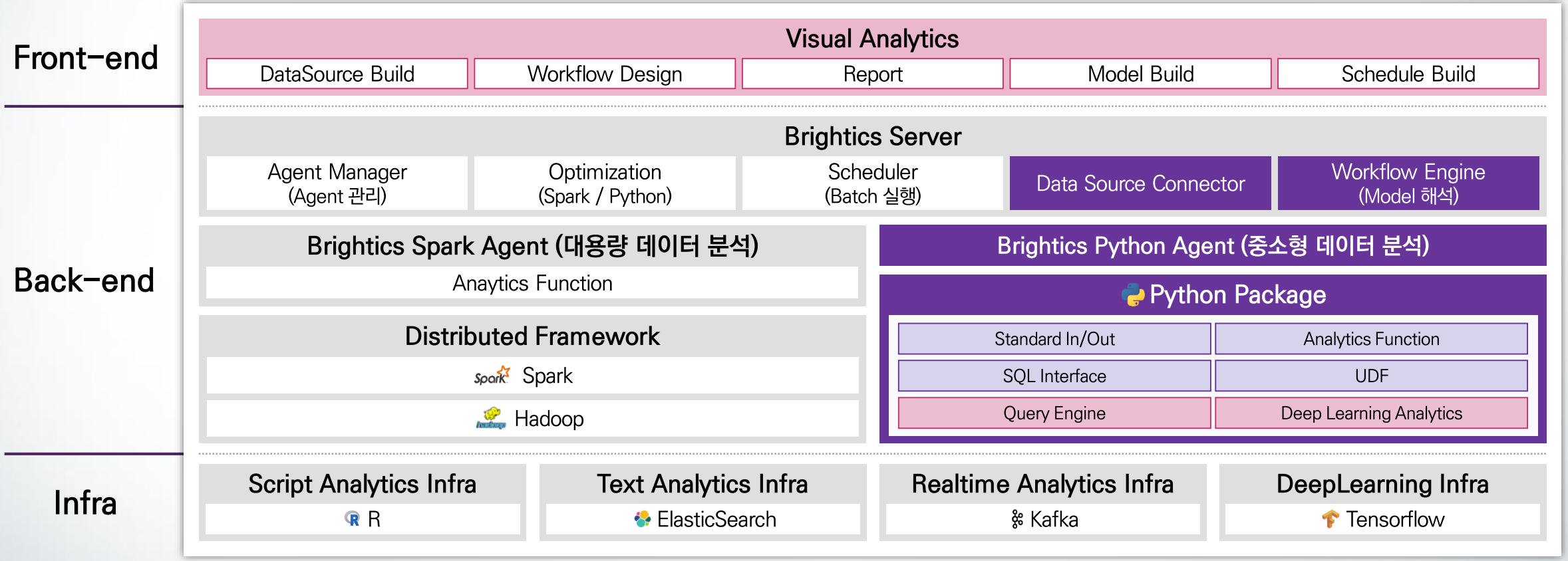


오픈소스 AI분석 플랫폼 Brightics Studio

Brightics AI

Brightics AI

다양한 기업데이터를 가공, 분석하여 **Business Layer에 AI 서비스를 제공하는 플랫폼**



※'18 Open Binary 제공

'18 Open Source '19 Open Source

업종별 주요 분석사례

제조, 마케팅, 물류, 보안, IoT, 헬스 분야 등 **90개 레퍼런스 확보**

제조	IoT	물류	마케팅	보안	헬스
설비 변경점 자동분석	**에너지** 수요예측	**글로벌 물류 위험** 예측	**판매 수요** 예측	**해킹 의심행위** 탐지	**유전체 샘플** 비교분석
품질불량 원인분석	**거주자 생활패턴** 분석	**배송지연** 사전예측	**방문객예측** 및 **직원 최적화**	Social Network 분석	**심질환 환자상태** 모니터링
설비 **예지보전**		Demand Sensing 예측	**시장반응(SNS)** 분석		

Q & A

Thank you