

Python & Azure Machine Learning

Python을 이용한 머신러닝 시작부터 예측모델 배포까지

<http://aka.ms/azure-ml-python>

한국마이크로소프트 | 김대우 | daewoo.kim@

2017-01-21

Python Machine Learning

Azure Machine Learning	01
Iris 분석 모델 workflow 구축	02
Data partition	03
Task flow	04
Accuracy check	05
Python code execution	06
Deploy to Web API and Excel	07

머신러닝 데모

- IRIS Data

(본사마와 무관한 통계업계의 “Hello World”)

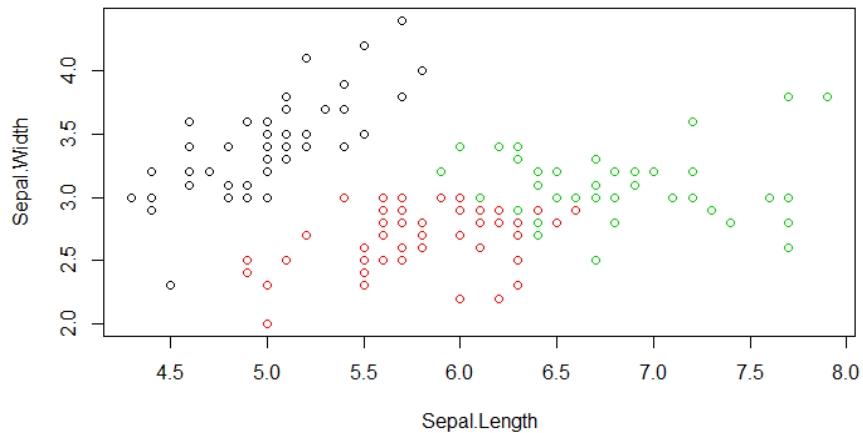
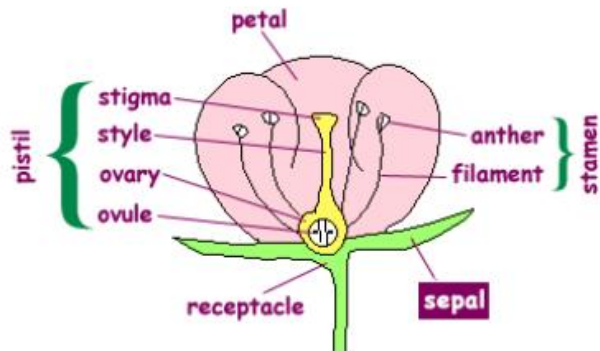
Iris setosa



Iris versicolor



Iris virginica



출처 : 디에스이트레이드 이성희

Iris 붓꽃 데이터 현황

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

- Column : 4개

- Row : 150개

- Sepal.Length : num

- Sepal.Width : num

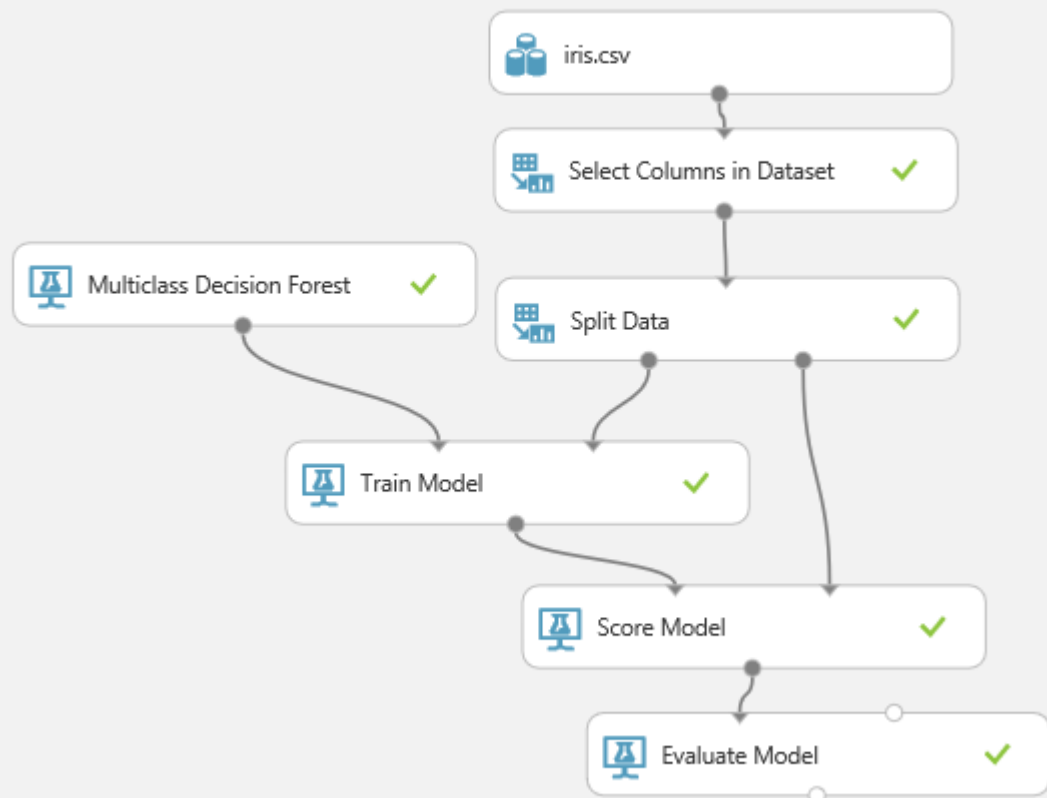
- Petal.Length : num

- Petal.Width : num

- Species : Factor

Iris모델 구축 데모 + 코드

iris 모델 - Azure Machine Learning



Demo

예측모델 생성 데모

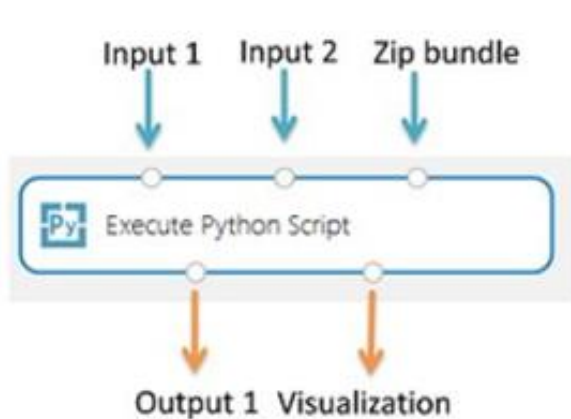
pydata & pandas

<https://github.com/wesm/pydata-book>

Materials and IPython notebooks for "Python for Data Analysis" by Wes McKinney, published by

O'Reilly Media

Python code execution



A diagram showing a Python code block with two inputs, "Input 1" and "Input 2", pointing to the function parameters `dataframe1` and `dataframe2` in the `azureml_main` function. The code defines the function, imports `pandas`, and includes comments for populating the return value and generating visualizations. Two orange arrows point from the `return` statement to the outputs "Output 1" and "Visualization".

```
def azureml_main(dataframe1, dataframe2):  
    import pandas  
    ## code to populate return value  
    result = pandas.DataFrame(...)  
    ## code to generate visualizations  
    return result,
```

```
0 1 2
1 5 6
2 0 11
```

dropna 메서드는 다루기 쉬운 여러 가지 추가 파라미터들을 지원한다.

```
# only drop rows where all columns are NaN
>>> df.dropna(how='all')

# drop rows that have not at least 4 non-NaN values
>>> df.dropna(thresh=4)

# only drop rows where NaN appear in specific columns (here: 'C')
>>> df.dropna(subset=['C'])
```

결측 데이터를 제거하는 것이 편리한 방법처럼 보일 수 있지만 이것 역시 단점이 있다. 예를 들면, 샘플을 너무 많이 제거해버리면 신뢰성 있는 분석이 불가능해질 수 있다. 혹은, 너무 많은 피쳐 열을 제거했다면 분류기가 분류들을 식별하는 데 필요한 특징 있는 정보를 잃게 될 수 있다. 다음 절에서는 결측값을 처리하는 방법으로 가장 많이 사용되는 것 중 하나인 보간법에 대해 살펴보려고 한다.

4.1.2 결측값의 보정

샘플을 제거하거나 전체 피쳐 열을 제거하면 가치 있는 데이터를 너무 많이 잃게 되어 사용할 수 없을 때가 자주 있다. 이런 경우, 우리는 여러 가지 보정법을 사용해 데이터 내의 다른 훈련 샘플들로부터 결측값을 추정할 수 있다. 가장 많이 사용되는 보정법 중 하나가 전체 피쳐열의 평균값으로 결측값을 간단히 대체하는 평균보정이다. 이것은 사이킷런의 `Imputer` 클래스를 사용해서 쉽게 구현할 수 있다. 다음의 코드를 참고하자.

```
>>> from sklearn.preprocessing import Imputer
>>> imr = Imputer(missing_values='NaN', strategy='mean', axis=0)
>>> imr = imr.fit(df)
>>> imputed_data = imr.transform(df.values)
>>> imputed_data
```



API로 호출

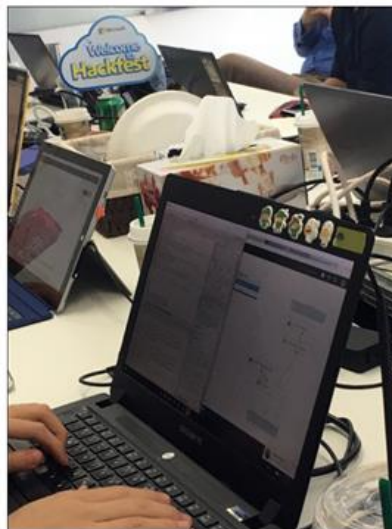
Python + node.js + C# + ...

Python 등에서 API 사용 DEMO

Restful Front-End를 이용한
Machine Learning API 호출
+ 대량 Batch 분석
DEMO




넥슨 Azure ML 머신러닝 케이스



Results

You've selected: Experiment ✖ [Clear all](#)

Sort by: Popular




EXPERIMENT

Sample 1: Download dataset from UCI: Adult 2 class dataset

This sample demonstrates how to download a dataset from a http location, add column names to the dataset and examine the dataset and compute some basic statistics.

50892 ↓ 116486 one year ago

Microsoft



EXPERIMENT


Tutorial: Building a classification model on Azure ML

This experiment serves as a tutorial on building a classification model using Azure ML. We will be using the...

Two-Class Decision Forest

20868 ↓ 8658 8 months ago

Raja Iqbal



EXPERIMENT


Telco Customer Churn

Customer churn can take different forms, such as switching to a competitor's service, reducing the number...

Two-Class Decision Forest

16617 ↓ 5037 4 months ago

weehyong tok



EXPERIMENT


Clustering: Find similar companies

This experiment clusters similar companies into same group given their Wikipedia articles and can be used...

K-Means Clustering

14590 ↓ 8979 9 months ago

Microsoft




MACHINE LEARNING API

Anomaly Detection

Detect different anomalous patterns in your time series data using machine learning algorithms. Level changes, trend changes, spikes are supported on seasonal and non-seasonal time series.

11263 9 days ago

Microsoft




EXPERIMENT

Online Fraud Detection: Step 1 of 5: Generate tagged data

This experiment demonstrates the steps in building an online transaction fraud detection solution.

16112 ↓ 4360 8 months ago

Microsoft



EXPERIMENT


Binary Classification: Twitter sentiment analysis

This experiment demonstrates the use of the Execute R Script, Feature Selection, Feature Hashing modules...

Two-Class Support Vector Machine

35381 ↓ 11997 one year ago

Microsoft




TUTORIAL

Sensor Data Analytics with ASA and Power BI

This is a tutorial for analyzing stream data in real time and creating a live dashboard using Power BI. We will be using a bluetooth sensor (TI sensor) for sending data to Eventhub, Azure Stream Analytics will analyze...

18958 one month ago

Microsoft



EXPERIMENT


Sample 5: Train, Test, Evaluate for Binary Classification: Adult...

This experiment demonstrates how we can build a binary classification model to predict income levels of ...

Two-Class Boosted Decision Tree

4986 ↓ 28859 one year ago

Microsoft



EXPERIMENT

Anomaly Detection: Credit Risk

Attempts to predict credit risk as an anomalies within the data.

PCA-Based Anomaly Detection, One-Class Support Vector Mac...

13071 ↓ 8013 one year ago

Microsoft

다양한 레퍼런스 머신러닝 아키텍처 모델 Cortana gallery

Q & A

Python & Azure Machine Learning

Python을 이용한 머신러닝 시작부터 예측모델 배포까지
<http://aka.ms/azure-ml-python>

한국마이크로소프트 | 김대우 | daewoo.kim@

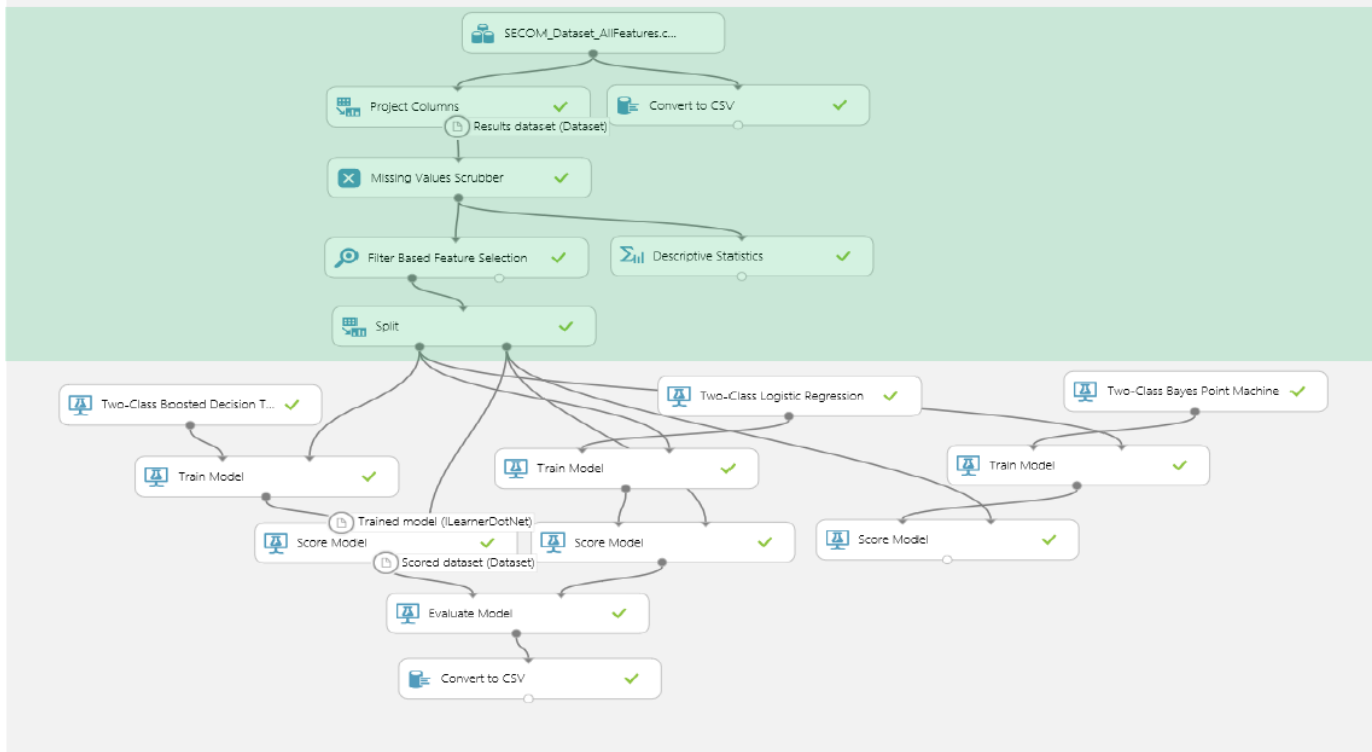
2017-01-21

Appendix



Azure ML 단계별 과정

Predictive Maintenance Model



데이터 수집

데이터 정리

알고리즘 선택

훈련 및 테스트

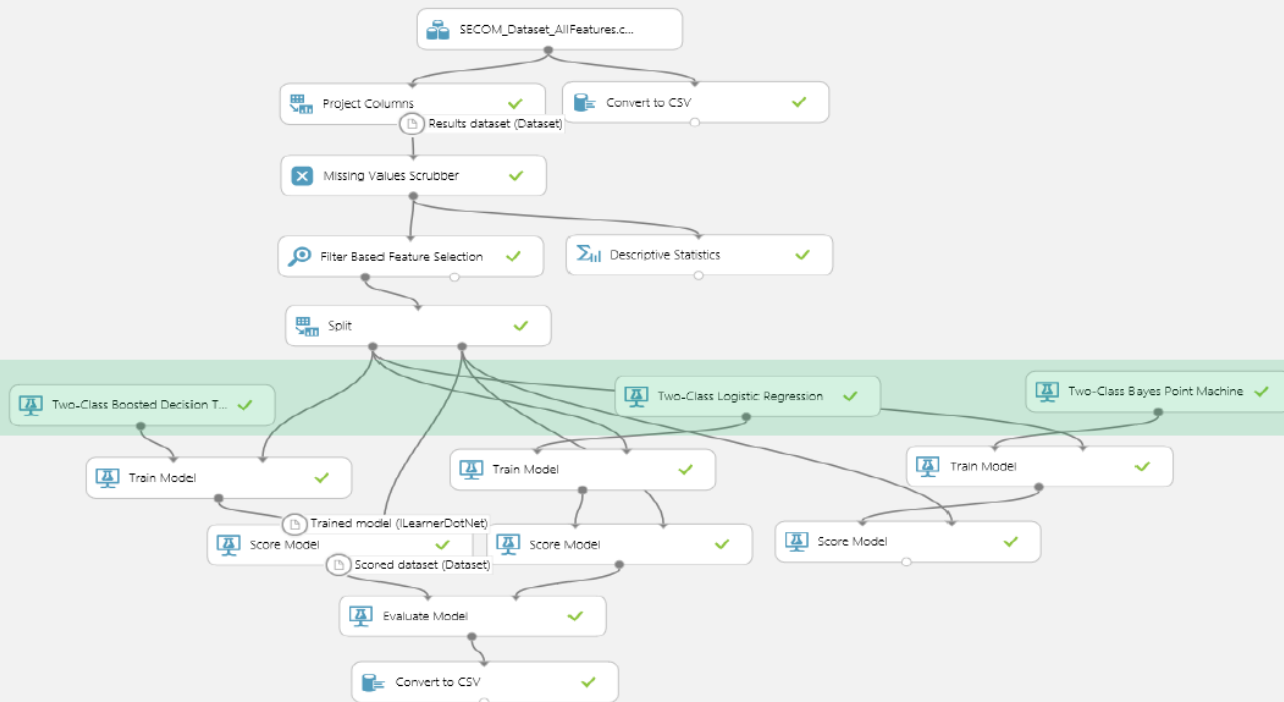
결과 추출

모델 배포



Azure ML 단계별 과정

Predictive Maintenance Model



데이터 수집

데이터 정리

알고리즘 선택

훈련 및 테스트

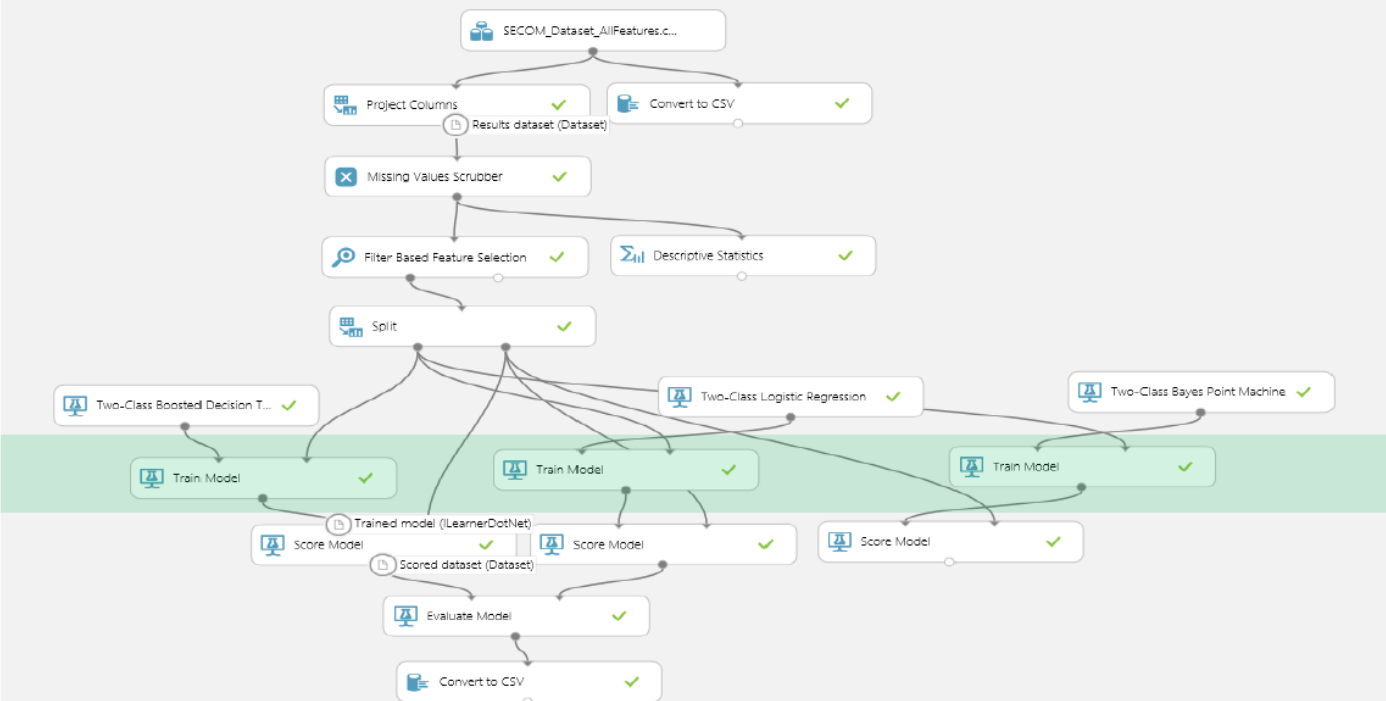
결과 추출

모델 배포



Azure ML 단계별 과정

Predictive Maintenance Model



데이터 수집

데이터 정리

알고리즘 선택

훈련 및 테스트

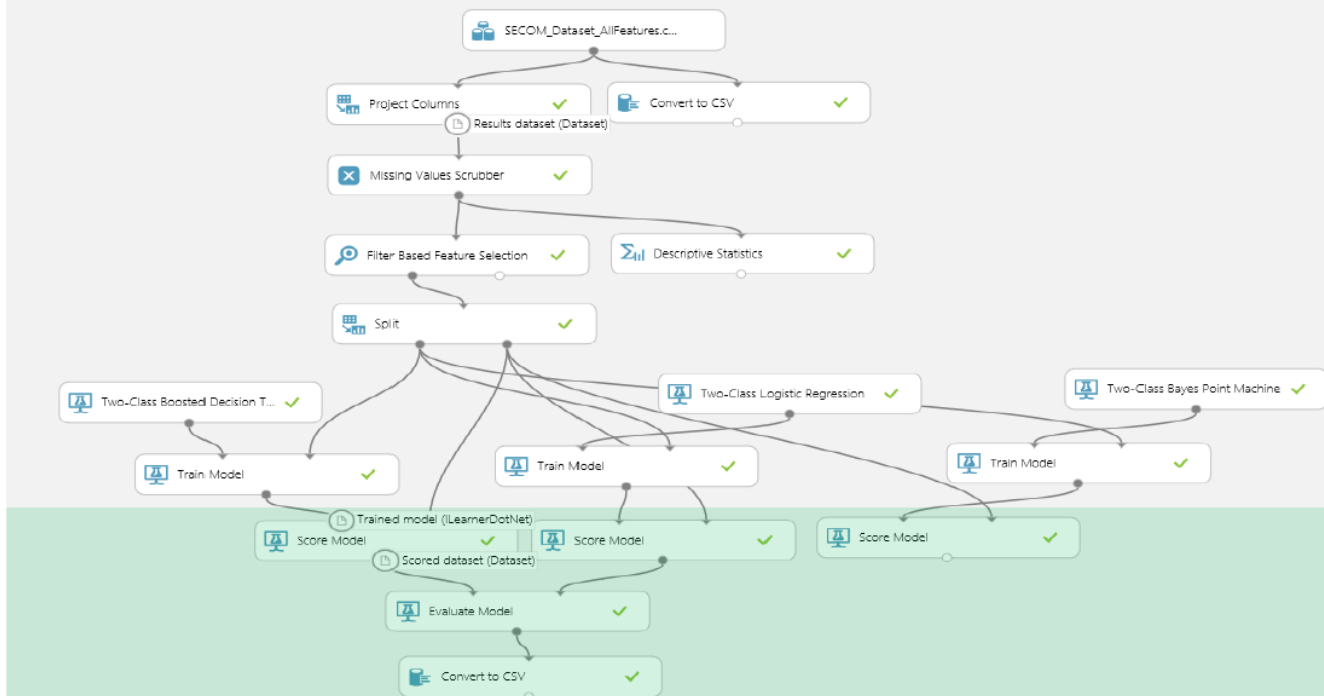
결과 추출

모델 배포



Azure ML 단계별 과정

Predictive Maintenance Model



데이터 수집

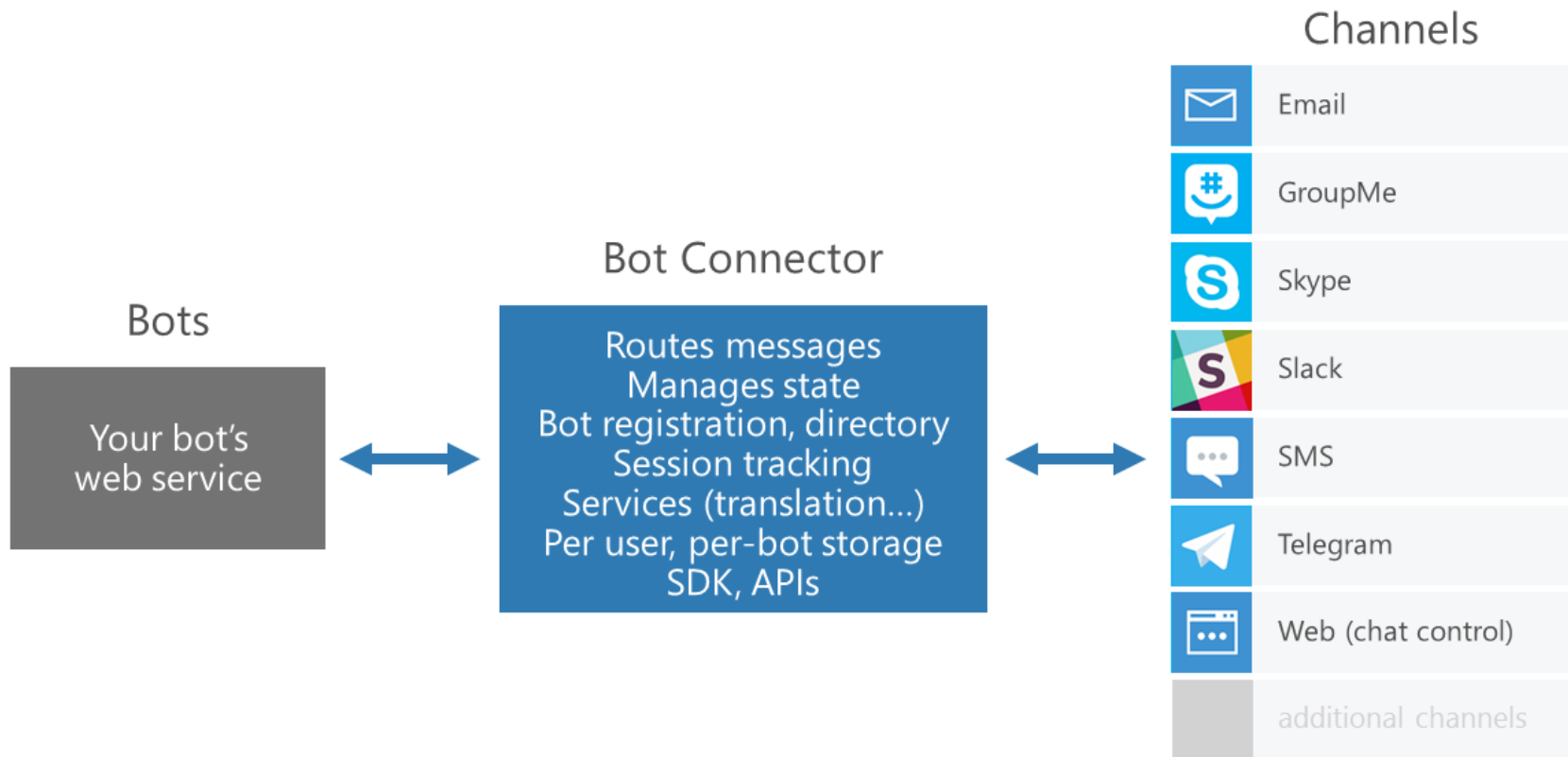
데이터 정리

알고리즘 선택

훈련 및 테스트

결과 추출

모델 배포



Bot Framework

Bot 커넥터

봇을 텍스트나 SMS, Mail, Skype, Slack, Facebook Messenger, Kik 등과 연동 가능

- 등록, 연결, 게시 및 관리 작업을 bot dashboard를 통해 수행 가능
- 메시지 라우팅
- 30개 이상의 언어 자동 번역 제공
- 사용자 및 상태(state) 관리
- 임베딩 가능한 웹 챗 제공
- 디버깅 도구 제공

Bot Builder SDKs

Node.js 또는 C#으로 봇을 개발 가능하며 필요한 API들을 모두 제공

- Github에서 오픈소스로 제공 <https://github.com/Microsoft/BotBuilder>
- 단순 빌트인 프롬프트부터 복잡한 명령 대화까지 제공
- 최고의 대화형 봇을 위한 다양한 라이브러리와 샘플 및 도구 제공
- 에뮬레이터 제공
- 다양한 Cognitive Service와 연계 가능

Cognitive Services

ML과 AI 서비스로 예측 및 지능을 모든 소프트웨어에서 구현 가능

- 자연어 처리(LUIS)
- 음성에 대해 실시간 트랜스크립트와 번역 서비스 제공
- 비전, 얼굴 인식 알고리즘
- 문장 인식(Sentence diagramming)
- Web language (hashtag) 파싱
- 다양한 서비스 API 제공

- 봇 등록
- 채널 설정
- 테스트
- 관리

Microsoft

Dae Woo Kim

Bot Framework

PREVIEW


My bots

Register a bot

Documentation

Bot Directory

Blog



dwkim-bot-d-party

Dae Woo Kim

In review

Details

Edit

Bot handle

dwkim-bot-d-party

Bot Framework Version

3.0

Messaging endpoint

https://dw-d-party-bot.azurewebsites.net/api/mess...





Microsoft App ID

()

Test connection to your bot






Test

Channels

	Test link	Issues	Enabled	Published		
	Skype		0	Yes (Preview)	Off	Edit
	Telegram	@dwkim_d_party_bot	20	Yes	Off	Edit
	Web Chat		0	Yes	Off	Edit

Get bot embed codes

Add another channel

	Direct Line	Add
	Email	Add
	Facebook Messenger	Add
	GroupMe	Add
	Kik	Add

Demo
<http://aka.ms/dpartybot>

EMERGENCY
TELEPHONE

Blue Line

Red Line



Browser window showing the **d_party_bot_framework** website. The page title is **d_party_bot_framework**. The content includes a chat interface and a JSON data input section.

Chat interface:

- Bot: Hi!! I'm dwkim-bot-d-party. Say "hi" if you'd like to chat
- Bot: {"SepalLength": "1", "SepalWidth": "1", "PetalLength": "1", "PetalWidth": "1"}
- You: 예측 결과: [{"Results": [{"output1": {"type": "table", "value": {"ColumnNames": ["SepalLength", "SepalWidth", "PetalLength", "PetalWidth"]}}]}]

JSON data input section:

```
{ "SepalLength": "1", "SepalWidth": "1", "PetalLength": "1", "PetalWidth": "1" }
{ "SepalLength": "2", "SepalWidth": "2", "PetalLength": "2", "PetalWidth": "2" }
```

스카이프 테스트 링크

[Add to Skype](#)

동작 안하면 서비스 지운거예요. 직접 아래 github repo로 만드시면 됩니다.

[Github repo 링크](#)

Visit [Bot Framework](#)

<http://dw-d-party-bot.azurewebsites.net/>
<http://aka.ms/dpartybot>

D.party Iris Machine Learning 봇
Github에서 코드 제공 및
웹 & 스카이프 실시간 테스트 가능

Transformational trends



Advanced Analytics

Business Intelligence와의 비교

