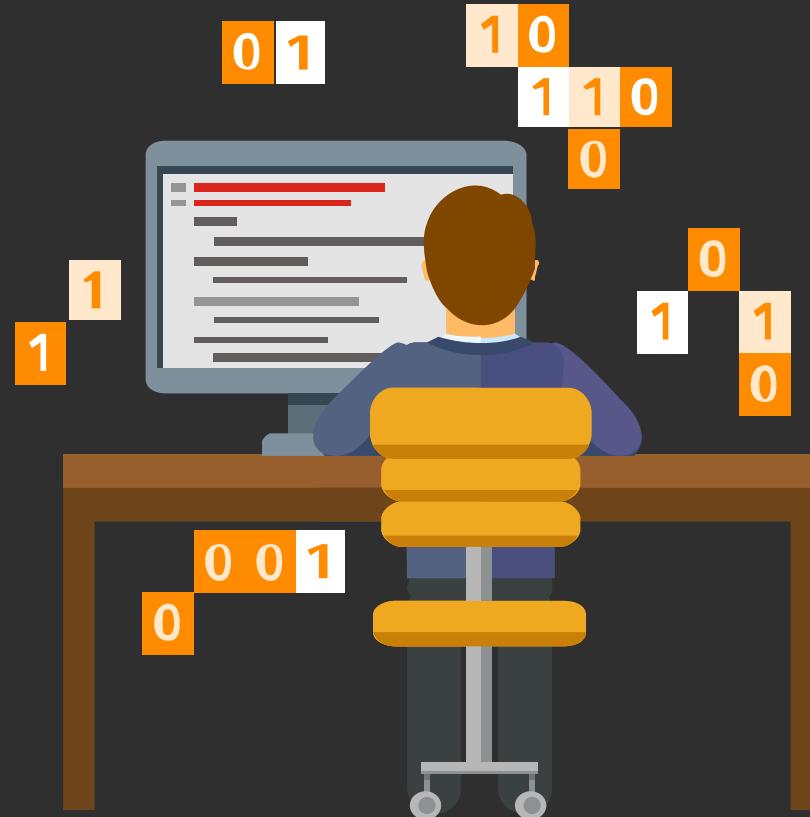


Azure ML Studio

-코드없이 만드는 친절한 머신러닝-

Speaker : Mijeong Jeon

Assistant : Jeonghwan Kim



Contents

- 1부 Azure AI 둘러보기
- 2부 Data 준비, 전처리
- 3부 모델 훈련, 평가, 배포

1

Azure AI 둘러보기

- Azure
- Azure AI
- Azure ML Studio
- Azure Pass 등록





Security & Management



Platform Services

The Platform Services section contains six main categories:

- Compute**: Cloud Services, Service Fabric, Batch, Remote App.
- Web and Mobile**: Web Apps, API Apps, API Management, Mobile Apps, Logic Apps, Notification Hubs.
- Developer Services**: Visual Studio, Azure SDK, Team Project, Application Insights.
- Integration**: Storage Queues, BizTalk Services, Hybrid Connections, Service Bus.
- Analytics & IoT**: HDInsight, Machine Learning (highlighted with a red box), Data Factory, Event Hubs, Stream Analytics, Mobile Engagement.
- Data**: SQL Database, SQL Data Warehouse, Redis Cache, Search, DocumentDB, Tables.
- Media & CDN**: Media Services, Content Delivery Network (CDN).

Hybrid Operations



Infrastructure Services

Compute



Storage



Networking



Azure Machine Learning 서비스



이 서비스에서 데이터 과학자가 CLI/Python 도구 및 라이브러리와 다양한 Azure 데이터 및 계산 서비스를 함께 사용하여 AI 솔루션을 개발하고 관리하도록 지원하는 방법을 알아봅니다.
자세히 알아보기...

Azure Machine Learning Studio



이 공동 작업을 위한 끌어서 놓기 도구를 사용하여 데이터에 대한 예측 분석 솔루션을 빌드, 테스트 및 배포하는 방법을 알아봅니다. 자세히 알아보기...

Cognitive Services



앱, 웹 사이트 및 봇에 지능형 알고리즘을 제공하는 API입니다.

Microsoft Machine Learning 서버



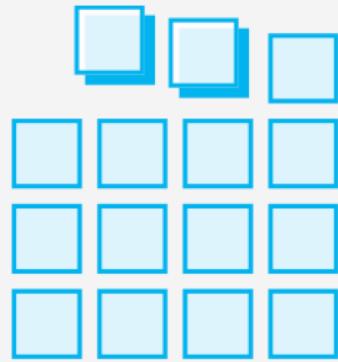
유연한 엔터프라이즈 기계 학습 플랫폼으로, Windows, Linux 및 HDInsight Spark와 Hadoop 클러스터에서 R 및 Python 배포를 지원합니다.

데이터 과학 Virtual Machine



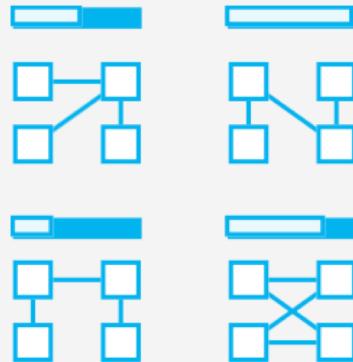
DSVM(Data Science Virtual Machine) 및 DLVM(Deep Learning Virtual Machine)을 사용하여 소프트웨어를 설치하거나 구성하지 않고 실험합니다. 인기 있는 데이터 분석, 기계 학습 및 AI 도구가 사전 설치되어 있습니다.

Azure AI



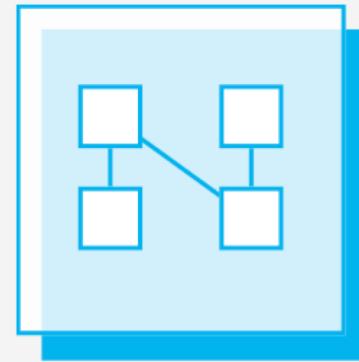
데이터 준비

다양한 원본에 연결하여 데이터 수집



빌드 및 학습

모델 설정 및 데이터를 사용하여 학습



배포

모델 배포 및 성능 추적

Azure Machine Learning

- Cloud-based: 서버비용, 관리비용, 초기 구축비용 절약
- Data Integration: 다양한 데이터 활용가능
- Common Toolset: 협업에 용이한 일반적인 모델과 알고리즘
- Simple Deployment: 배포 및 관리 용이

Azure Machine Learning 서비스



이 서비스에서 데이터 과학자가 CLI/Python 도구 및 라이브러리와 다양한 Azure 데이터 및 계산 서비스를 함께 사용하여 AI 솔루션을 개발하고 관리하도록 지원하는 방법을 알아봅니다.
자세히 알아보기...

Azure Machine Learning Studio



이 공동 작업을 위한 끌어서 놓기 도구를 사용하여 데이터에 대한 예측 분석 솔루션을 빌드, 테스트 및 배포하는 방법을 알아봅니다. 자세히 알아보기...

Cognitive Services



앱, 웹 사이트 및 봇에 지능형 알고리즘을 제공하는 API입니다.

Microsoft Machine Learning 서버



유연한 엔터프라이즈 기계 학습 플랫폼으로, Windows, Linux 및 HDInsight Spark와 Hadoop 클러스터에서 R 및 Python 배포를 지원합니다.

데이터 과학 Virtual Machine



DSVM(Data Science Virtual Machine) 및 DLVM(Deep Learning Virtual Machine)을 사용하여 소프트웨어를 설치하거나 구성하지 않고 실험합니다. 인기 있는 데이터 분석, 기계 학습 및 AI 도구가 사전 설치되어 있습니다.



Search experiment items



▶ Saved Datasets



▶ Trained Models



▶ Transforms
▶ Data Format Conversions



▶ Data Input and Output



▶ Data Transformation



▶ Feature Selection



▶ Machine Learning



▶ OpenCV Library Modules
▶ Python Language Modules



▶ R Language Modules



▶ Statistical Functions

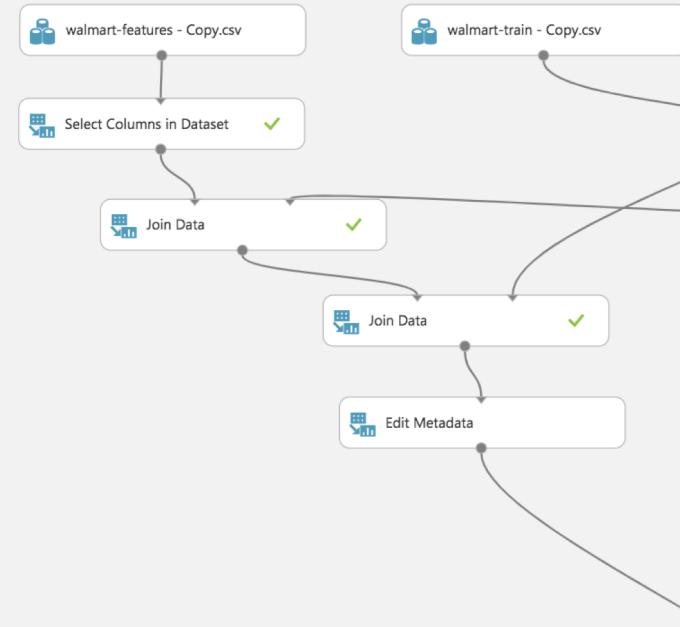


▶ Text Analytics

Training experiment

Predictive experiment

Walmart Sales Forecasting Using Regres



Azure ML Studio

- Visual interface, Drag and drop ML
- 터미널 x, 커맨드 라인 x, 코드 x
- 고급 데이터 마이닝 기술 사용: R, Python, SQL
- 다양한 개발 기술 사용 : RESTful API, C#, R, Python

Best tool for ML beginner!/Fast ML

Azure ML Studio

데이터 전처리에서 모델 구성까지

총 111개 모듈 제공

<https://docs.microsoft.com/azure/machine-learning/studio-module-reference/a-z-module-list>

문제 정의

데이터셋 준비

모델 설정

모델 훈련

111 모듈

모델 평가

모델 활용

Machine Learning in ML Studio

Anomaly Detection

- One-class Support Vector Machine
- Principal Component Analysis-based Anomaly Detection
- Time Series Anomaly Detection*

Classification

Two-class Classification

- Averaged Perceptron
- Bayes Point Machine
- Boosted Decision Tree
- Decision Forest
- Decision Jungle
- Logistic Regression
- Neural Network
- Support Vector Machine

Multi-class Classification

- Decision Forest
- Decision Jungle
- Logistic Regression
- Neural Network
- One-vs-all

Clustering

K-means Clustering

Recommendation

Matchbox Recommender

Regression

- Bayesian Linear Regression
- Boosted Decision Tree
- Decision Forest
- Fast Forest Quantile Regression
- Linear Regression
- Neural Network Regression
- Ordinal Regression
- Poisson Regression

Statistical Functions

- Descriptive Statistics
- Hypothesis Testing T-Test
- Linear Correlation
- Probability Function Evaluation

Text Analytics

- Feature Hashing
- Named Entity Recognition
- Vowpal Wabbit

Computer Vision

- OpenCV Library

Data/Model Visualization

- Scatterplots
- Bar Charts
- Box plots
- Histogram
- R and Python Plotting Libraries
- REPL with Jupyter Notebook
- ROC, Precision/Recall, Lift
- Confusion Matrix
- Decision Tree*

Training

- Cross Validation
- Retraining
- Parameter Sweep

<https://studio.azureml.net>

Guest Access Workspace: Free trial access without logging in.

Free Workspace: Free persisted access, no Azure subscription needed.

Standard Workspace: Full access with SLA under an Azure subscription.

Cross browser drag & drop ML workflow designer.
Zero installation needed.

Import Data

Preprocess

Split Data

Train Model

Score Model

Training Experiment

One-click Operationalization

Predictive Experiment

Make Prediction with Elastic APIs

- Request-Response Service (RRS)
- Batch Execution Service (BES)
- Retraining API

Data Source

- Azure Blob Storage
- Azure SQL DB
- Azure SQL DW*
- Azure Table
- Desktop Direct Upload
- Hadoop Hive Query
- Manual Data Entry
- OData Feed
- On-prem SQL Server*
- Web URL (HTTP)

Data Format

- ARFF
- CSV
- SVMLight
- TSV
- Excel
- ZIP

Data Preparation

- Clean Missing Data
- Clip Outliers
- Edit Metadata
- Feature Selection
- Filter
- Learning with Counts
- Normalize Data
- Partition and Sample
- Principal Component Analysis
- Quantize Data
- SQLite Transformation
- Synthetic Minority Oversampling Technique

Enterprise Grade Cloud Service

- SLA: 99.95% Guaranteed Up-time
- Azure AD Authentication
- Compute at Large Scale
- Multi-geo Availability
- Regulatory Compliance*

Community

- Gallery (<http://gallery.azureml.net>)
- Samples & Templates
- Workspace Sharing and Collaboration
- Live Chat & MSDN Forum Support

* Feature Coming Soon



Azure Machine Learning Studio Capabilities Overview

친절한 Repository

http://bit.ly/181215_Azure_HOL

Azure Pass 등록

http://bit.ly/181215_azure_pass

Azure ML Studio 로그인

<https://studio.azureml.net>

2

2부: 데이터 준비, 전처리

- 데이터 형식
- 데이터 업로드
- 데이터 합치기
- 데이터 전처리



지원 데이터 형식

- 헤더가 있거나(.csv) 없는(.nh.csv) 쉼표로 구분된 값(CSV)
- 헤더가 있거나(.tsv) 없는(.nh.tsv) 탭으로 구분된 값(TSV)
- 일반 텍스트(.txt)
- Excel 파일
- Azure 테이블
- Hive 테이블
- SQL 데이터베이스 테이블
- SVMLight 데이터(.svmlight)
- 특성 관계 파일 형식(ARFF) 데이터(.arff)
- Zip 파일(.zip)
- R 개체 또는 작업 영역 파일(.RData)

데이터 유형

- 문자열
- 정수
- double
- BOOLEAN
- Datetime
- timespan

Machine Learning in ML Studio

Anomaly Detection

- One-class Support Vector Machine
- Principal Component Analysis-based Anomaly Detection
- Time Series Anomaly Detection*

Classification

Two-class Classification

- Averaged Perceptron
- Bayes Point Machine
- Boosted Decision Tree
- Decision Forest
- Decision Jungle
- Logistic Regression
- Neural Network
- Support Vector Machine

Multi-class Classification

- Decision Forest
- Decision Jungle
- Logistic Regression
- Neural Network
- One-vs-all

Clustering

K-means Clustering

Recommendation

Matchbox Recommender

Regression

- Bayesian Linear Regression
- Boosted Decision Tree
- Decision Forest
- Fast Forest Quantile Regression
- Linear Regression
- Neural Network Regression
- Ordinal Regression
- Poisson Regression

Statistical Functions

- Descriptive Statistics
- Hypothesis Testing T-Test
- Linear Correlation
- Probability Function Evaluation

Text Analytics

- Feature Hashing
- Named Entity Recognition
- Vowpal Wabbit

Computer Vision

- OpenCV Library

Data/Model Visualization

- Scatterplots
- Bar Charts
- Box plots
- Histogram
- R and Python Plotting Libraries
- REPL with Jupyter Notebook
- ROC, Precision/Recall, Lift
- Confusion Matrix
- Decision Tree*

Training

- Cross Validation
- Retraining
- Parameter Sweep

<https://studio.azureml.net>

Guest Access Workspace: Free trial access without logging in.

Free Workspace: Free persisted access, no Azure subscription needed.

Standard Workspace: Full access with SLA under an Azure subscription.

Cross browser drag & drop ML workflow designer.
Zero installation needed.

Import Data

Preprocess

Unlimited Extensibility

- R Script Module
- Python Script Module
- Custom Module
- Jupyter Notebook

Built-in ML Algorithms

Split Data

Train Model

Score Model

Training Experiment

One-click Operationalization

Predictive Experiment

Make Prediction with Elastic APIs

- Request-Response Service (RRS)
- Batch Execution Service (BES)
- Retraining API

Data Source

- Azure Blob Storage
- Azure SQL DB
- Azure SQL DW*
- Azure Table
- Desktop Direct Upload
- Hadoop Hive Query
- Manual Data Entry
- OData Feed
- On-prem SQL Server*
- Web URL (HTTP)

Data Format

- ARFF
- CSV
- SVMLight
- TSV
- Excel
- ZIP

Data Preparation

- Clean Missing Data
- Clip Outliers
- Edit Metadata
- Feature Selection
- Filter
- Learning with Counts
- Normalize Data
- Partition and Sample
- Principal Component Analysis
- Quantize Data
- SQLite Transformation
- Synthetic Minority Oversampling Technique

Enterprise Grade Cloud Service

- SLA: 99.95% Guaranteed Up-time
- Azure AD Authentication
- Compute at Large Scale
- Multi-geo Availability
- Regulatory Compliance*

Community

- Gallery (<http://gallery.azureml.net>)
- Samples & Templates
- Workspace Sharing and Collaboration
- Live Chat & MSDN Forum Support

* Feature Coming Soon



Azure Machine Learning Studio Capabilities Overview

© 2015 Microsoft Corporation. All rights reserved.

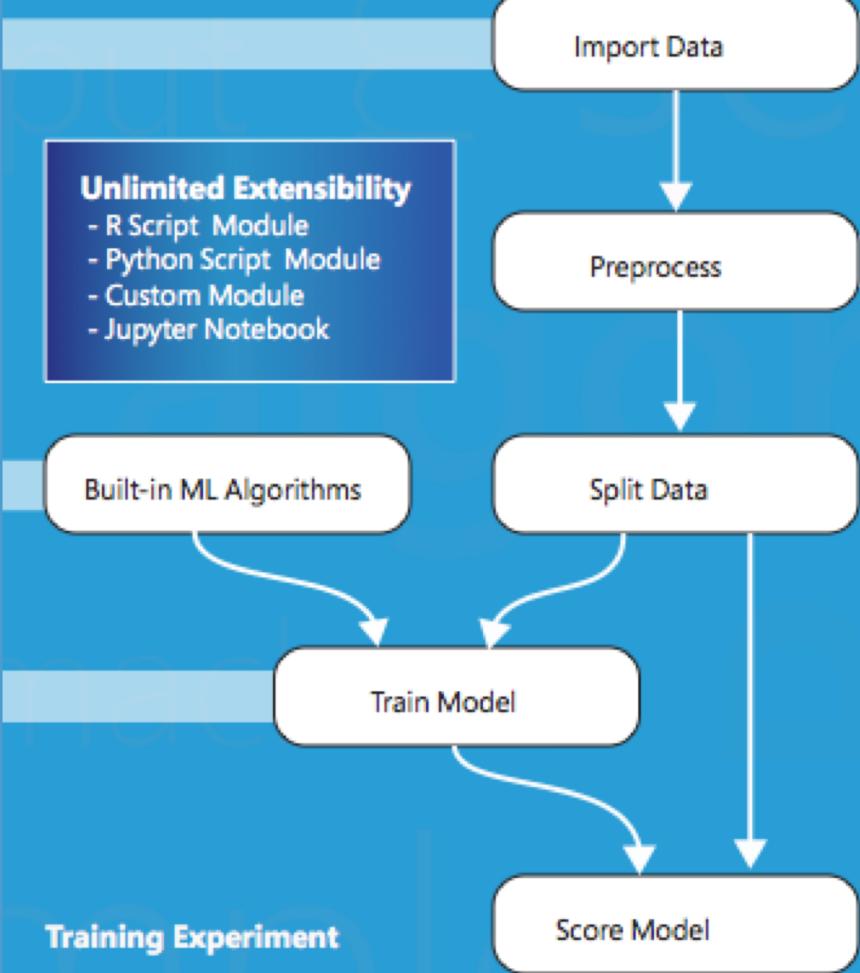
Created by the Azure Machine Learning Team

Email: AzurePoster@microsoft.com

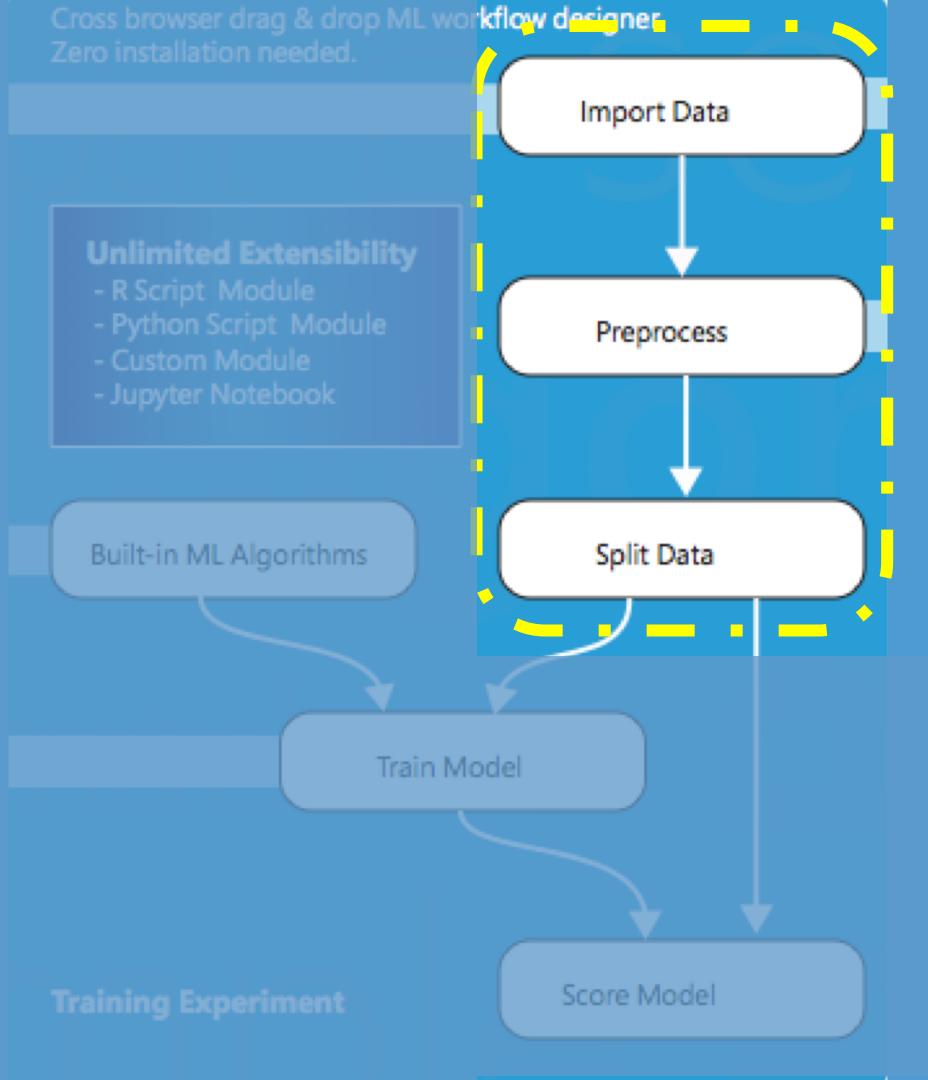
Download this poster: <http://aka.ms/MLStudioOverview>



Cross browser drag & drop ML workflow designer.
Zero installation needed.



Cross browser drag & drop ML workflow designer
Zero installation needed.



데이터 다운받기

http://bit.ly/181215_hol_dataset

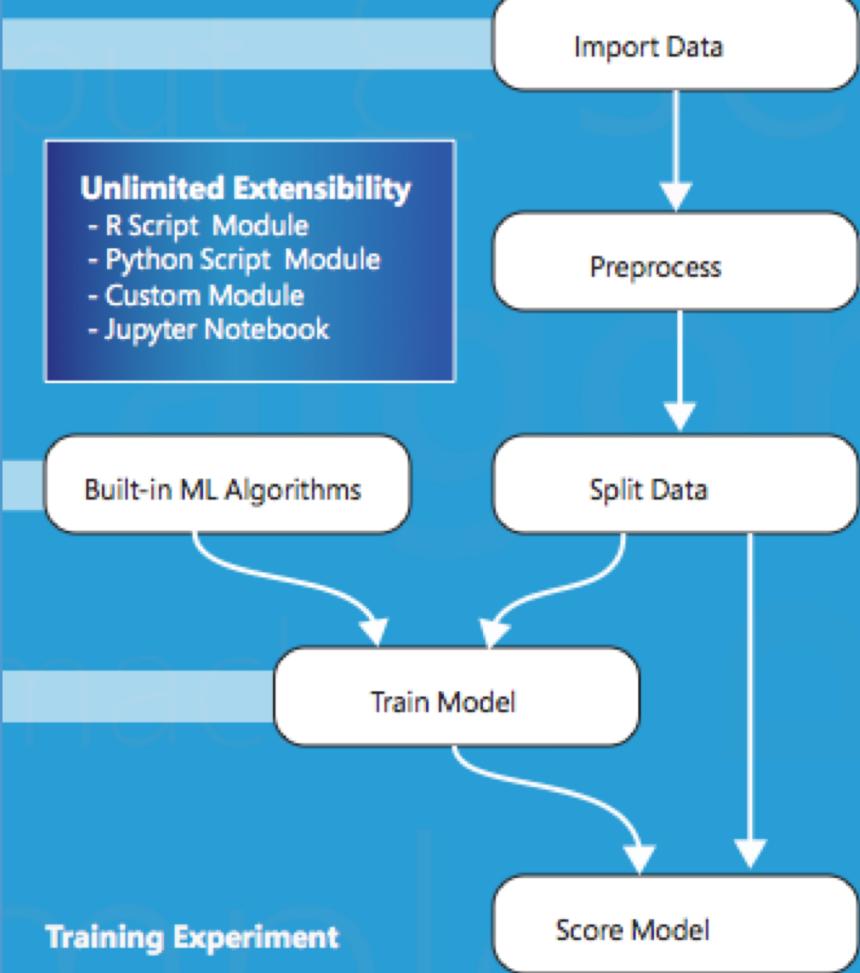
3

3부: 모델 학습, 예측, 배포

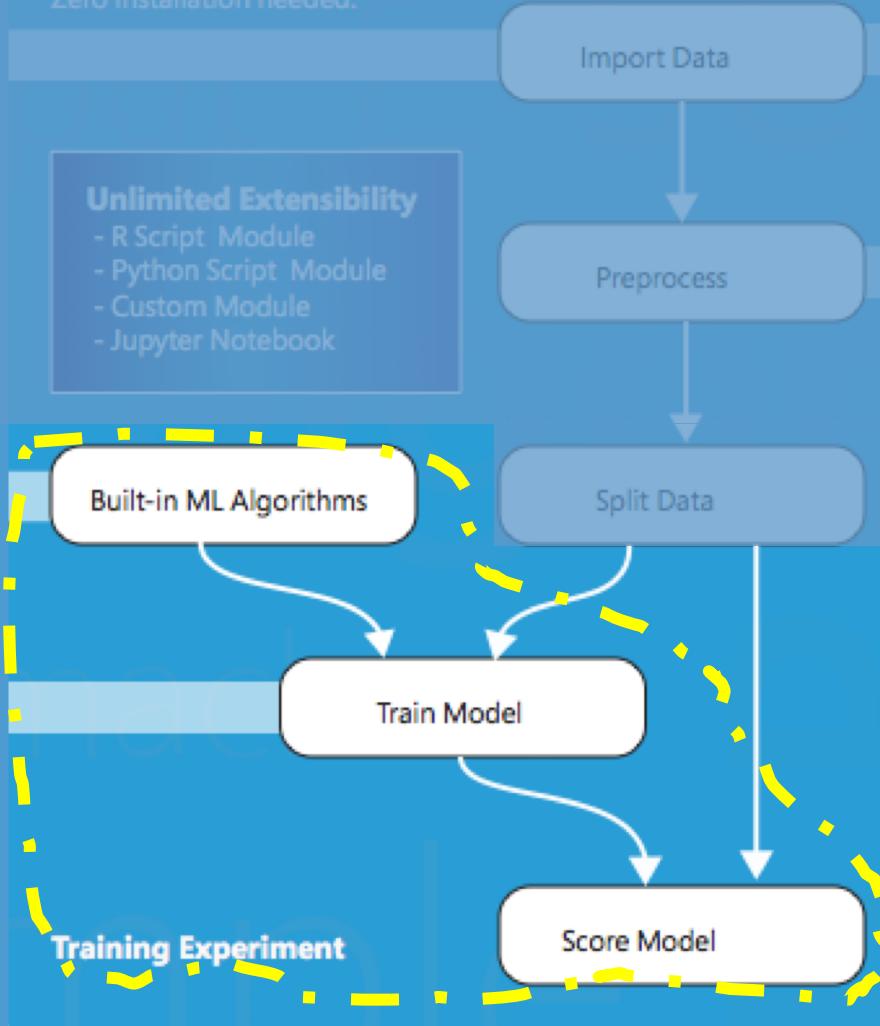
- 다양한 모델 살펴보기
- 올바른 모델 선택
- 모델 훈련, 평가, 배포
- 모델 활용 방안



Cross browser drag & drop ML workflow designer.
Zero installation needed.



Cross browser drag & drop ML workflow designer.
Zero installation needed.



“What machine learning algorithm should I use?”

“What machine learning algorithm should I use?”

“It depends.”

On size & quality of data, train time, computer, application

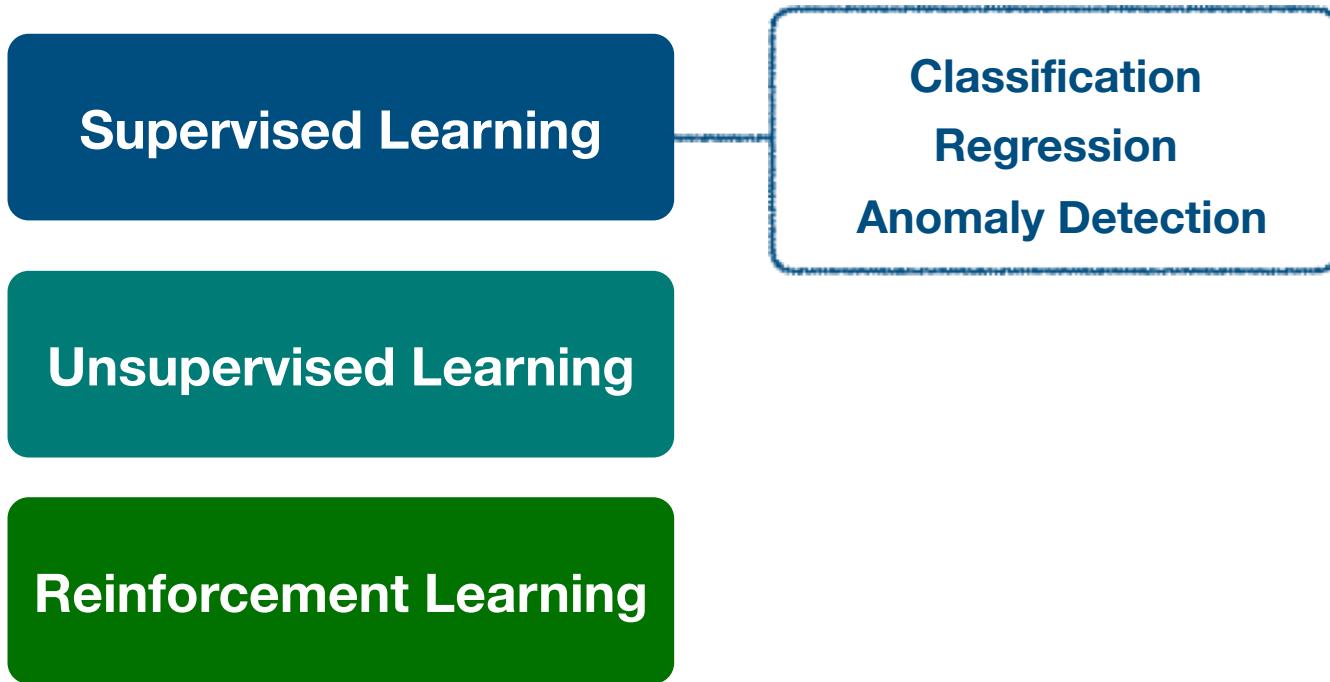
Machine Learning

Supervised Learning

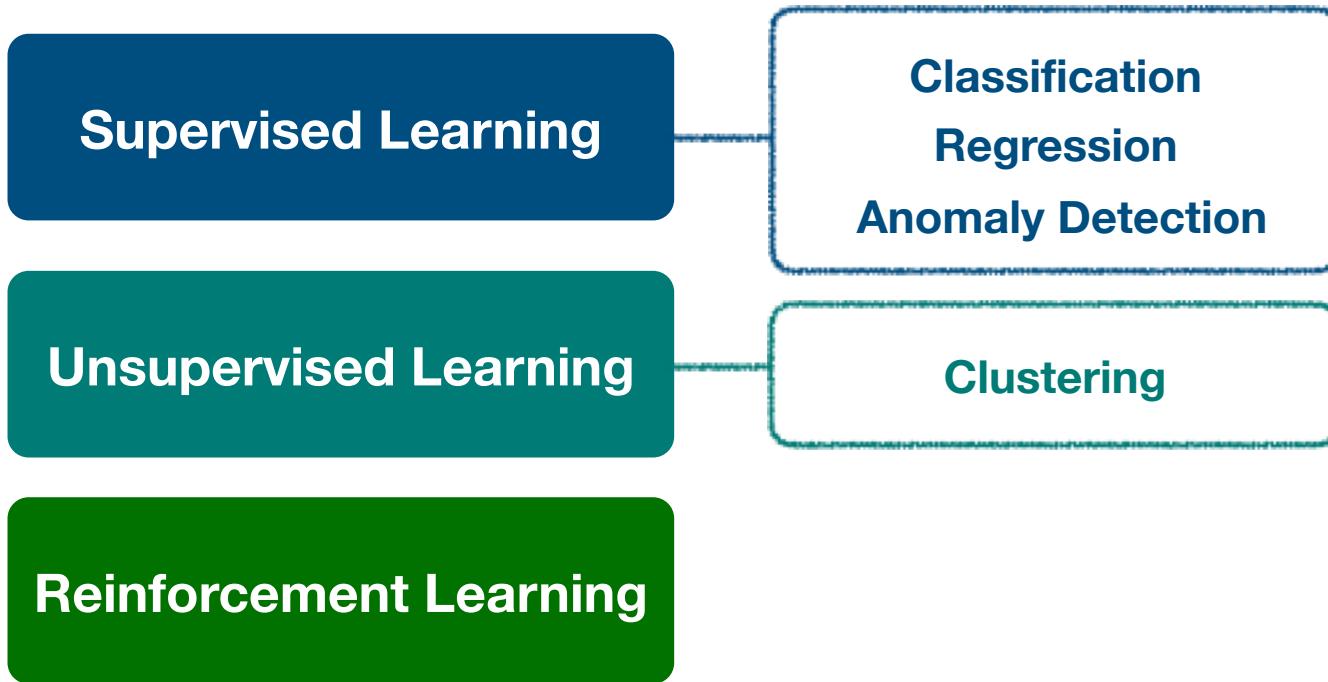
Unsupervised Learning

Reinforcement Learning

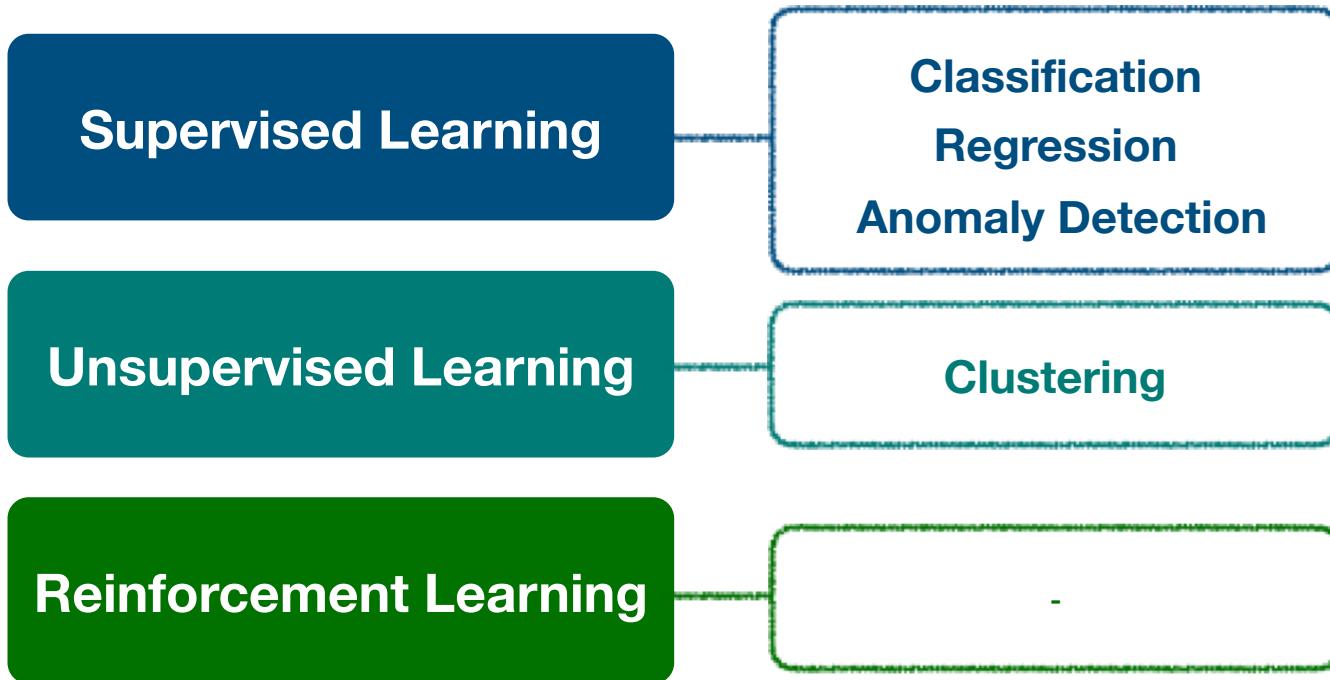
Machine Learning



Machine Learning



Machine Learning



Accuracy

Parameter

Training Time

Feature

Linearity

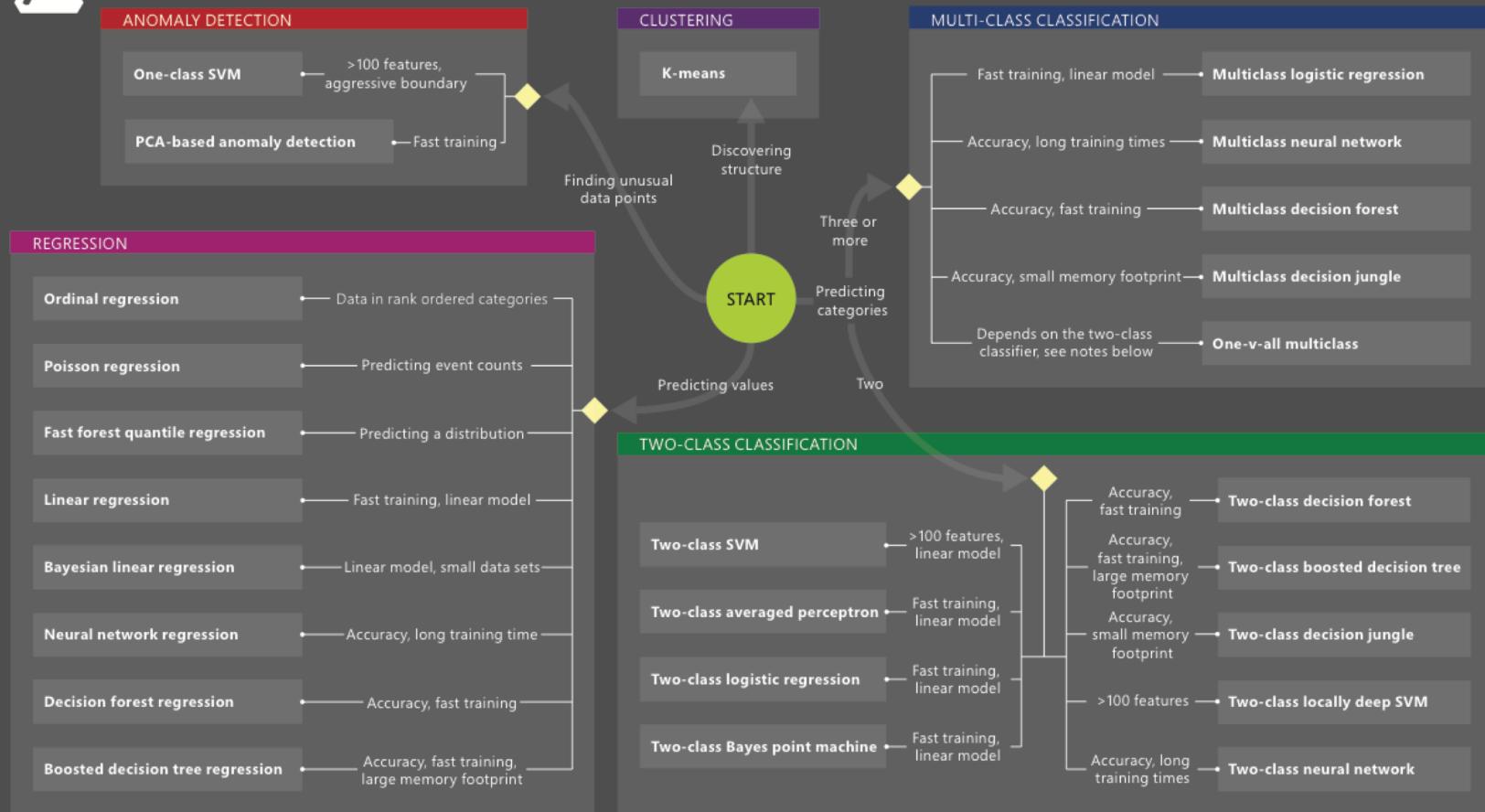
Etc.

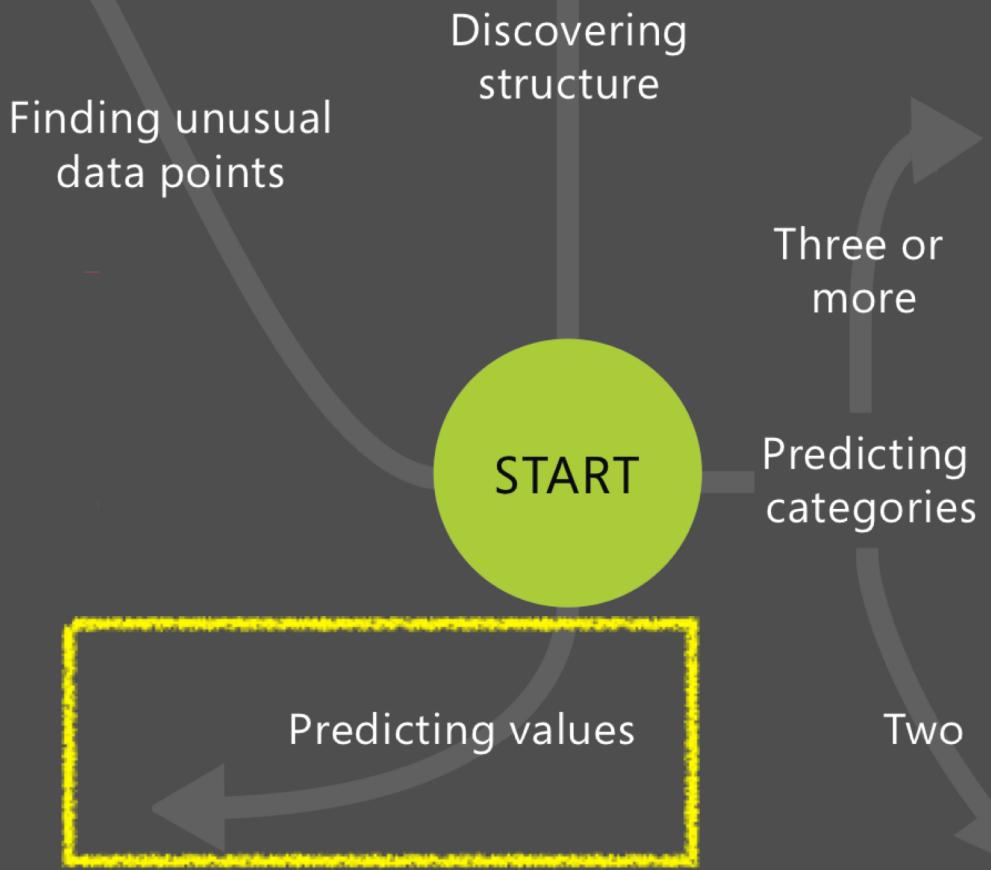
<https://docs.microsoft.com/ko-kr/azure/machine-learning/studio/algorith-choice>



Microsoft Azure Machine Learning: Algorithm Cheat Sheet

This cheat sheet helps you choose the best Azure Machine Learning Studio algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the question you're trying to answer.





Ordinal regression

→ Data in rank ordered categories

Poisson regression

→ Predicting event counts

Fast forest quantile regression

→ Predicting a distribution

Linear regression

→ Fast training, linear model

Bayesian linear regression

→ Linear model, small data sets

Neural network regression

→ Accuracy, long training time

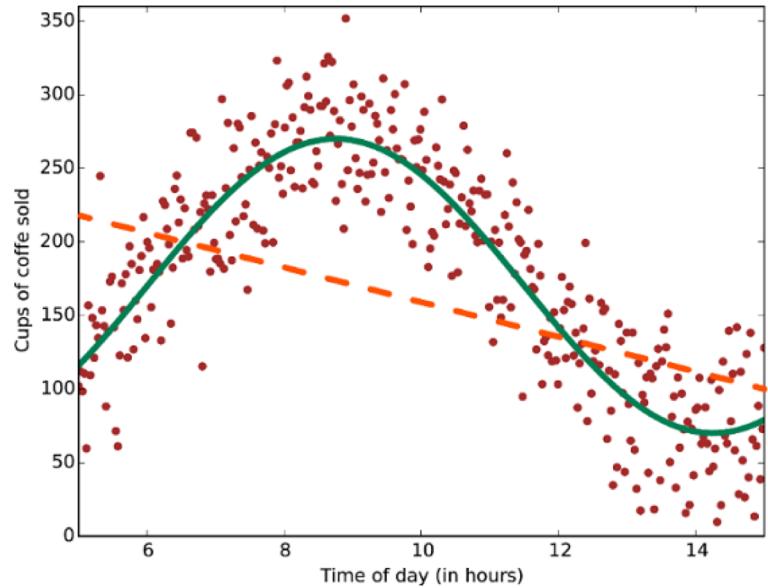
Decision forest regression

→ Accuracy, fast training

Boosted decision tree regression

→ Accuracy, fast training,
large memory footprint

REGRESSION



- 키에 따른 신발 사이즈
- 시간에 따른 커피 소비량
- 햇빛 노출 시간과 주근깨 개수
- 달 위상에 따른 주요 도시의 범죄 수
- 기온과 인터넷 쇼핑 장바구니 물품 수

REGRESSION

Ordinal Regression	데이터 내 상대적 순서나 랭킹 예측 ex) 강연 참석자의 선호도, URL 즐겨찾기 순서	0
Poisson Regression	어떤 이벤트가 발생할 횟수 예측 이산분포를 따르며 음의 정수값 X ex) 비행기 탑승에 따른 병원 방문 횟수	5
Fast Forest Quantile Regression	값의 분산/분포 예측 ex) 성적 예측률 통한 학생들의 발달 단계 평가	9
Linear Regression	가장 일반적인 선형 회귀 알고리즘	4

REGRESSION

Bayesian Linear Regression	Bayesian 접근법을 선형회귀에 적용한 방식 확률 분포에 관련된 사전 정보가 있을 때 활용	2
Neural Network Regression	신경망 회로(DNN), 비선형 문제에 활용 가능 Customizable algorithm	9
Decision Forest Regression	의사 결정 트리, 비선형 문제에 활용 가능 효율적인 메모리 사용 및 계산 (overfitting 주의)	6
Boosted Decision Tree Regression	이전 트리에 종속되어 있어 메모리 사용이 큼 정확도가 높음, 앙상블 모델에 활용 가능	5

Ordinal regression

→ Data in rank ordered categories

Poisson regression

→ Predicting event counts

Fast forest quantile regression

→ Predicting a distribution

Linear regression

→ Fast training, linear model

Bayesian linear regression

→ Linear model, small data sets

Neural network regression

→ Accuracy, long training time

Decision forest regression

→ Accuracy, fast training

Boosted decision tree regression

→ Accuracy, fast training,
large memory footprint

“All models are wrong but some are useful”

- George Box



◀ Evaluate

Cross Validate Model 

Evaluate Model 

Evaluate Recommender 

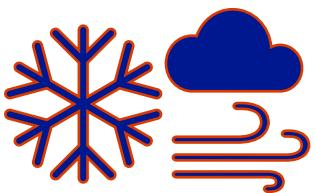
◀ Metrics

Mean Absolute Error	3197.596091
Root Mean Squared Error	5469.159134
Relative Absolute Error	0.211708
Relative Squared Error	0.058495
Coefficient of Determination	0.941505

Evaluate Regression

Mean Absolute Error(MAE)	$MAE = \frac{\sum_{i=1}^n p_i - a_i }{n}$
Root Mean Squared Error(RMSE)	$RMSE = \sqrt{\frac{\sum_{i=1}^n (p_i - a_i)^2}{n}}$
Coefficient of Determination	R^2

Relative Absolute Error(RAE)	$RAE = \frac{\sum_{i=1}^n p_i - a_i }{\sum_{i=1}^n \bar{a} - a_i }$
Relative Squared Error(RSE)	$RSE = \frac{\sum_{i=1}^n (p_i - a_i)^2}{\sum_{i=1}^n (\bar{a} - a_i)^2}$



Weather Forecast API

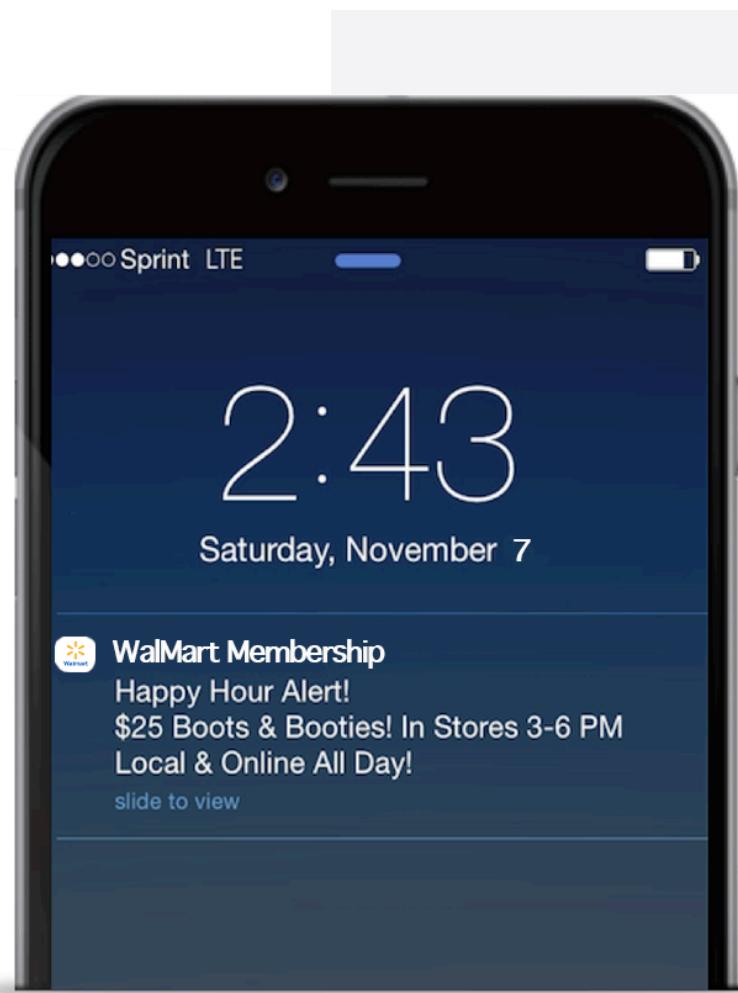


Fuel Price Forecast API



	A	B	C	D	E	F	G	H	I
1	Store	Date	Temperature	Fuel_Price	Dept	Weekly_Sales	IsHoliday	Type	Size
2	1	2010-02	42.31	2.572	1	24924.5	FALSE	A	151315
3	1	2010-02	43.31	2.813	2	50605.27	FALSE	A	151315
4	1	2010-02	42.31	2.572	3	13740.12	FALSE	A	151315
5	1	2010-02	42.31	2.572	4	39954.04	FALSE	A	151315
6	1	2010-02	42.31	2.572	5	32229.38	FALSE	A	151315
7	1	2010-02	43.31	2.813	1	24924.5	FALSE	A	151315
8	1	2010-02	42.31	2.572	2	50605.27	FALSE	A	151315
9	1	2010-02	43.31	2.813	3	13740.12	FALSE	A	151315
10	1	2010-02	42.31	2.572	4	39954.04	FALSE	A	151315
11	1	2010-02	43.31	2.572	5	32229.38	FALSE	A	151315
12	1	2010-02	42.31	2.813	1	24924.5	FALSE	A	151315
13	1	2010-02	43.31	2.572	2	50605.27	FALSE	A	151315
14	1	2010-02	42.31	2.813	3	13740.12	FALSE	A	151315
15	1	2010-02	43.31	2.572	4	39954.04	FALSE	A	151315
16	1	2010-02	43.31	2.813	5	32229.38	FALSE	A	151315

Storage(Azure Blob)



- <https://www.kaggle.com/datasets>
- <http://archive.ics.uci.edu/ml/datasets.html>
- <https://gallery.azure.ai>
- <https://studio.azureml.net>
- <https://docs.microsoft.com/ko-kr/azure/machine-learning/studio/algorithm-choice>
- <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/machine-learning-initialize-model-regression>

GLOBAL AI BOOTCAMP

NEW INTELLIGENT WORLDS

감사합니다.