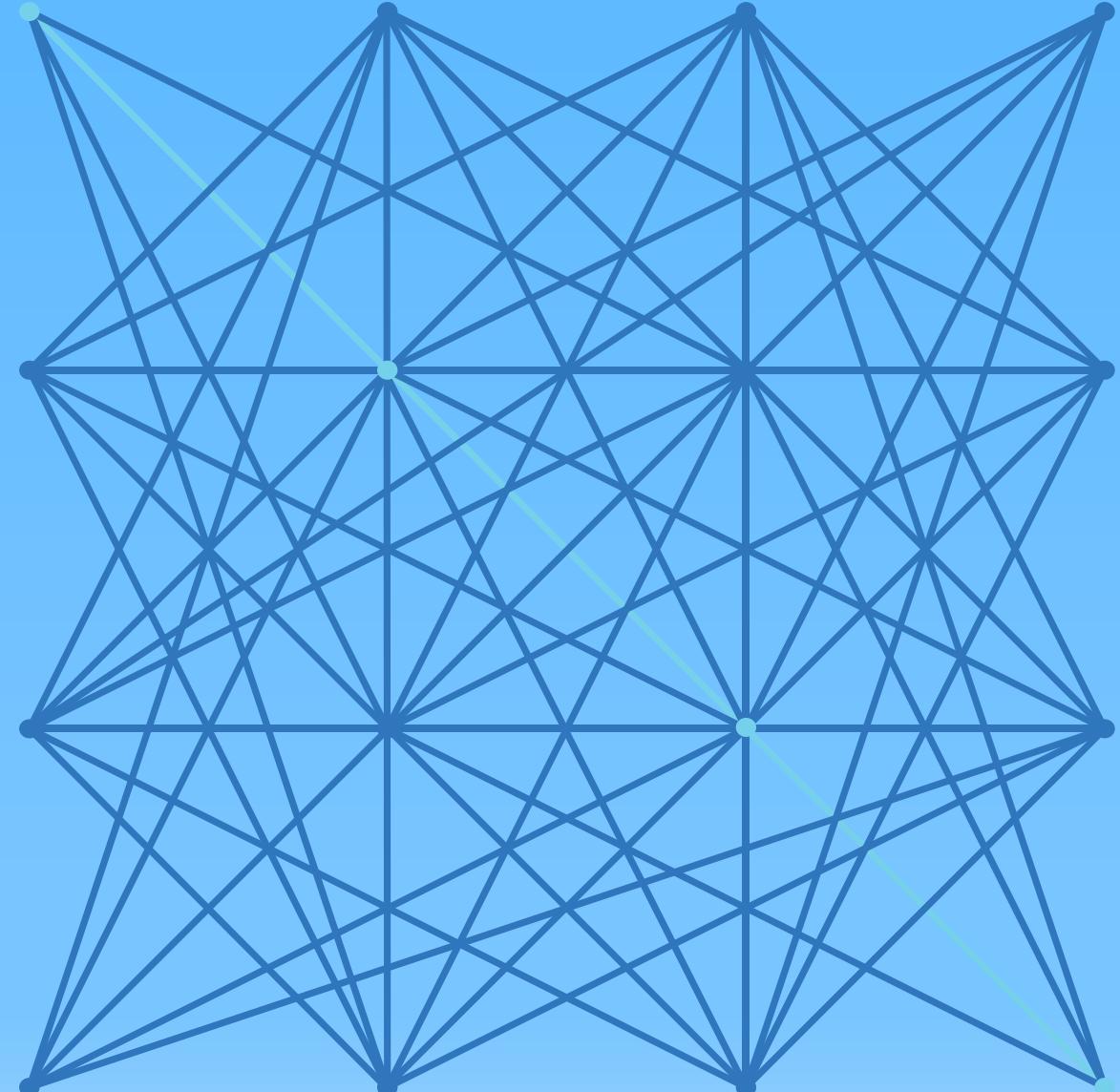


Automated ML 동향 알아보기

2019.7.4 (목) 15:00 사이언스홀
디플러스 김영하





6년동안
직업군인
하다가...



Embrace Data Chaos

Go from data to business outcomes
faster than ever before with Splunk

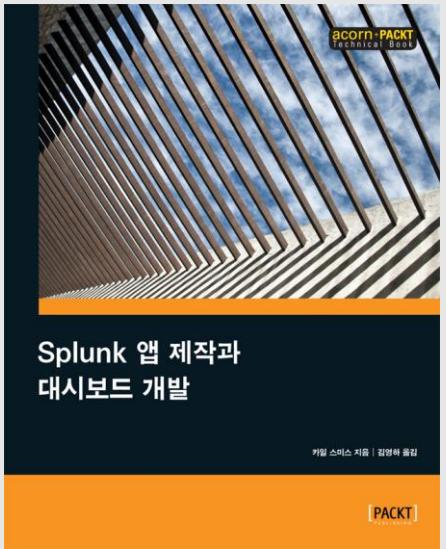
[Start Your Journey](#)

.conf19 Call For Papers is now open!

[Share Your Story](#)

Turn Data Into Action

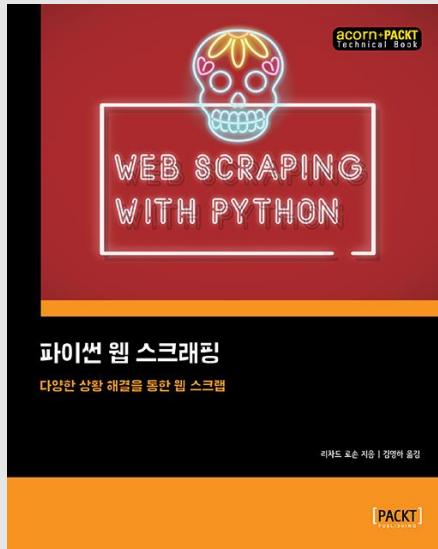
어쩌다 운좋게 빅데이터를 만지게 되어



Splunk 앱 제작과
대시보드 개발

키릴 스미스 지음 | 김영하 옮김

[PACKT]



파이썬 웹 스크래핑
다양한 상황 해결을 통한 웹 스크랩



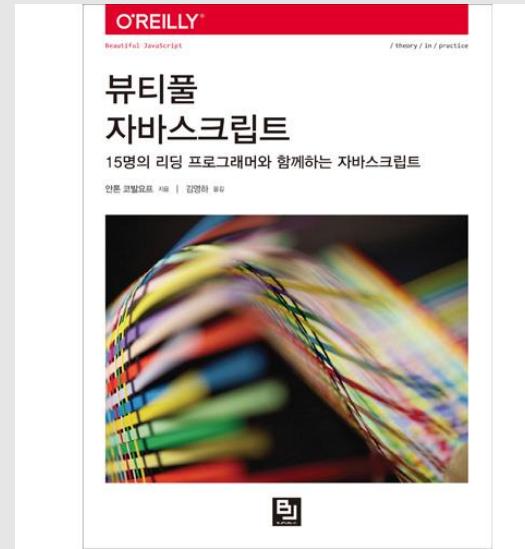
세상의 속도를
따라갈고 싶다면
Do it!
데이터 분석을 위한
판다스 입문

테슬라 주식 · 우버 택시 · 빌보드 차트 · 애플리케이션 데이터 등
현실 데이터 기반의 실제 예제로 판다스 실무 기본기 원장
Daniel Y. Chen 지음 | 김영하 옮김



국어판 2판 판다스

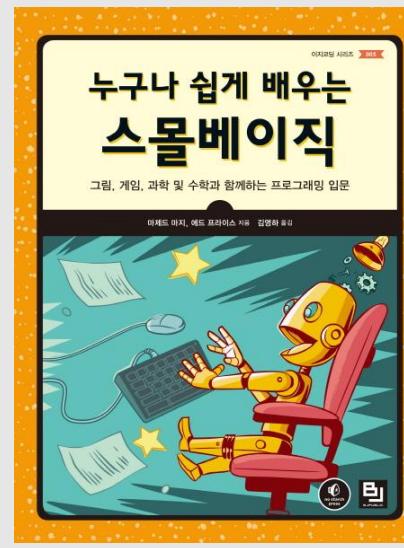
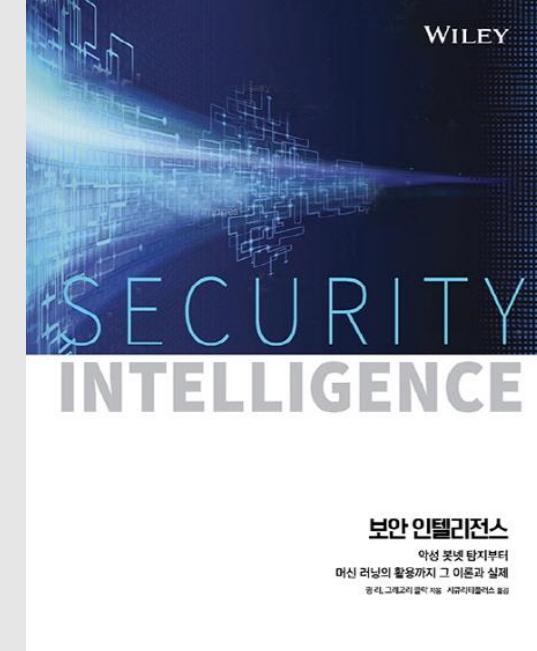
최신 버전 번역판
Pandas for Everyone



뷰티풀
자바스크립트

15명의 리딩 프로그래머와 함께하는 자바스크립트

안톤 코발트로프 저음 | 김영하 옮김



누구나 쉽게 배우는
스몰베이직

마체드 마시, 에드 프赖스 저음 | 김영하 옮김



Korea ML.NET User Group

공개 그룹

정보

토론

1

채팅

공지

멤버

이벤트

사진

그룹 인사이트

그룹 관리

이 그룹 검색



게시물 작성

사진/동영상 추가

라이브 방송



더 보기



글쓰기...



사진/동영상



함께 시청하기



친구 태그하기



멤버 초대

초대 펴가기

+ 이름이나 이메일 주소를 입력하세요...



멤버

멤버 124명



이번 주에 새로운 멤버가 17명 있습니다. 새 멤버를 환영하는 게시물을 작성해보세요.

추천 멤버

송기기

친구

사진

파일

이 그룹 검색



바로가기

따빠마드레 Tapa M... 20+

Korea ML.NET User Gr...

TensorFlow KR 20+

OpenLab.AI

책쓰는 프로그래머 ... 1



가입함 ▾ 알림 공유하기 ... 더 보기

우분투 공식 한국 지역팀입니다.

우분투는 사용자 중심의 오픈소스 기반 OS입니다.
그렇다고 개발자가 못 쓰진 않아요.

매월 정기 세미나도 합니다.
참여대상자는 사람! 행사기념품은 지식!

자세한건 ubuntu-kr.org 공지를 참조하세요.

Forum_ ubuntu-kr.org
wiki_ wiki.ubuntu.or.kr
facebook_ facebook.com/groups/ubuntu.ko
irc_ irc.ubuntu.com/#ubuntu-ko
YouTube_ youtube.com/user/UbuntuKorea

가입함 ▾ 알림 공유하기 ... 더 보기

게시물 작성 사진/동영상 추가 라이브 방송 ... 더 보기

글쓰기...

사진/동영상

함께 시청하기

친구 태그하기 ...

더 보기

관리 페이지

Ubuntu Korea Community
564명이 좋아합니다

문의하기

멤버 초대 + 이름이나 이메일 주소를 입력하세요...

멤버

멤버 9,435명

MEMBERS

데이터 분석을 돋는

모두의 희망
재택근무!



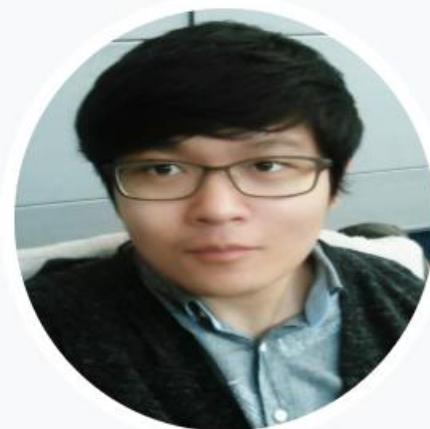
정원혁

대표



김영하

연구원

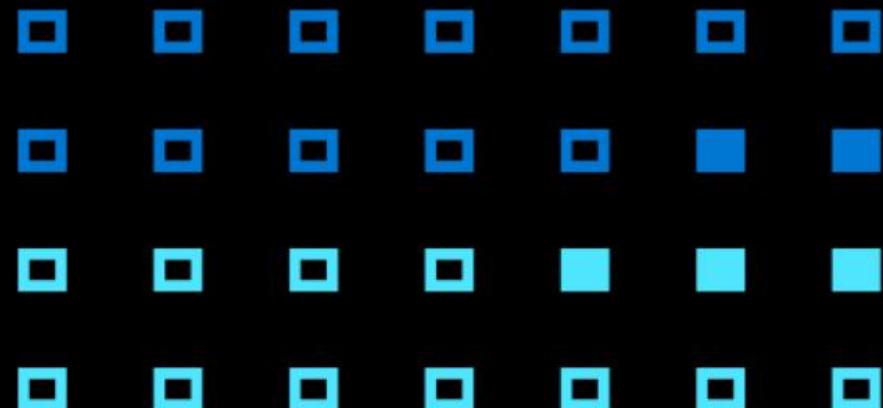


김정훈

데이터 엔지니어



Microsoft Azure Discovery Day



Azure Discovery Day

Microsoft Azure의 클라우드 서비스는 분석, 컴퓨팅, 데이터베이스, 모바일, 네트워킹, 스토리지, 웹 서비스를 제공하여, 합리적인 비용으로 더 빠르면서도 더 많은 것을 성취할 수 있게 합니다.

매달 진행되는 Azure Discovery Day는 Microsoft Azure를 통해 귀사의 인프라의 효율적인 운영 및 개발을 지원할 수 있는 방법을 소개해 드리고자 마련되었습니다.

본 세미나는 2일 속성 과정으로

- 첫째 날인 **Azure Discovery Day for Azure 101**에서는 클라우드 전반에 걸쳐 다양한 서비스에 대한 기본적인 소개와 함께 Azure의 주요 기능과 활용 방안을 데모와 함께 전달 드릴 예정입니다.

- 둘째 날인 **Azure Discovery Day for AI & ML**에서는 Microsoft Azure에서 구현 가능한 다양한 AI 서비스를 살펴볼 예정입니다. Azure Machine Learning Studio의 기본 활용부터 다양한 App에 연동할 수 있는 개발 기술까지 Microsoft 가 준비한 다양한 Demo와 함께 AI 기술을 학습할 수 있는 시간을 가져 보시기 바랍니다.

Welcome to Bokeh

Bokeh is an interactive visualization library that targets modern web browsers for presentation. Its goal is to provide elegant, concise construction of versatile graphics, and to extend this capability with high-performance interactivity over very large or streaming datasets. Bokeh can help anyone who would like to quickly and easily create interactive plots, dashboards, and data applications.

To get started using Bokeh to make your visualizations, start with the [User Guide](#).

For examples of how you might use Bokeh with your own data, check out the [Gallery](#).

For detailed information about specific Bokeh components, consult the [Reference Guide](#).

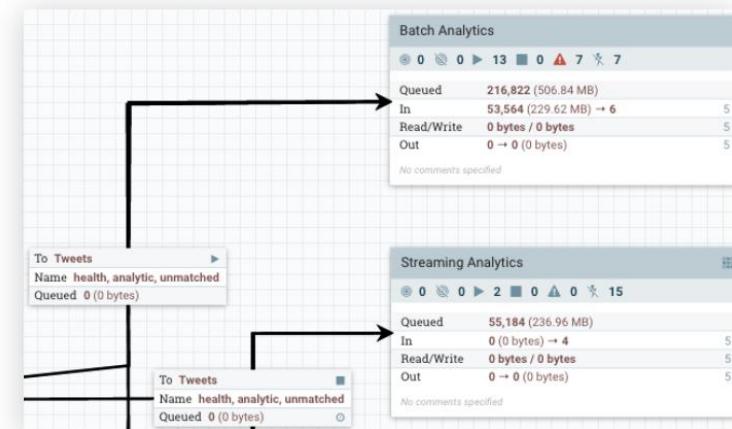
If you are interested in contributing to Bokeh, or extending the library, see the [Developer Guide](#).



3개월째
무시무시한 제조데이터
를 만지고 있습니다!



An easy to use, powerful, and reliable system to process and distribute data.



Features

Apache NiFi supports powerful and scalable directed graphs of data routing, transformation, and system

mediation logic. Some of the high-level capabilities and objectives of Apache NiFi include:

- Web-based user interface

박사 학위는 없지만
기술을 좋아하고 전파하고 싶은
Too Much Information Searcher



딥러닝 컨퍼런스에
맞지않게 쉬운 내용들이니
편하게 들어주세요

Automated

Automated

자동화

뭐가 자동화?

Automated Factory Automated System

넓은 의미의 Automated

좁은 의미의 Automated

Data Science

"컴퓨터 도구를 **효율적으로** 이용하고,
적절한 **통계학 방법**을 사용하여
실제적인 문제에 답을 내리는 활동"

“Data Scientist (noun): Person who is better at statistics
than any software engineer and better at software
engineering than any statistician.”

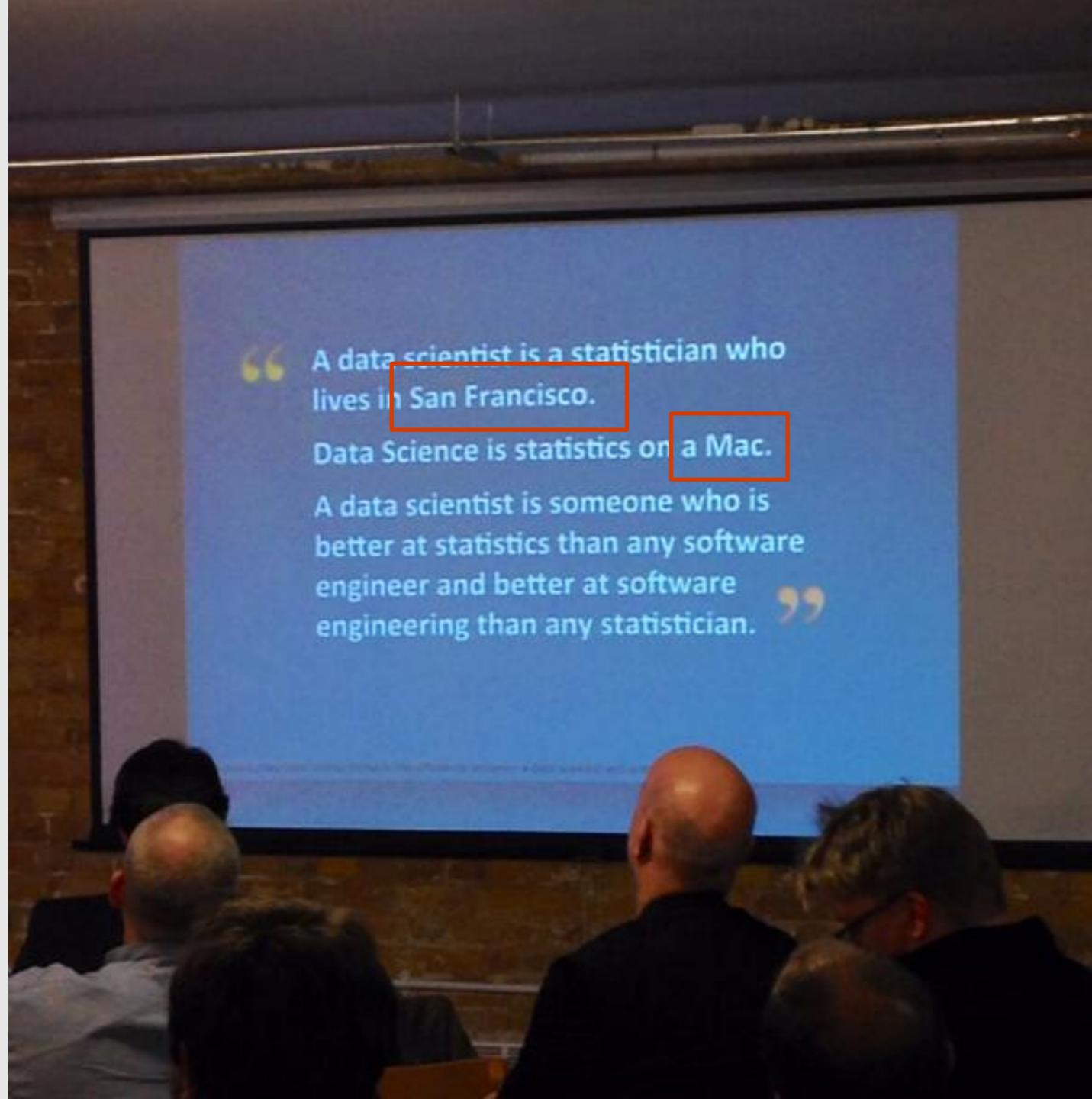
긍정적인 마음

-- Josh Wills, Director of Data Science at Cloudera
(formerly Google)

부정적인 마음`

“Data Scientist (noun): Person who is worse at statistics
than any statistician and worse at software engineering
than any software engineer.”

-- Will Cukierski, Data Scientist at Kaggle



자! 이 정의에 의하면

한국에는 진정한
데이터과학자가
없습니다!

용기를 가지세요!

데이터 과학자의 일상 1



출처 : <https://datasciencedegree.wisconsin.edu/data-science/what-do-data-scientists-do/>

데이터 과학자의 일상 2



데이터에 숨겨진
의미를 찾기 위해 모델
및 알고리즘 개발



패턴과 흐름을
확인하기 위해
데이터 분석



해결책과 기회를
발견하기 위해
데이터 해석



시각화 등의 방향으로
현업담당자와 논의

데이터 과학자들은

데이터를 분석하기에도 **시간이 부족합니다!**

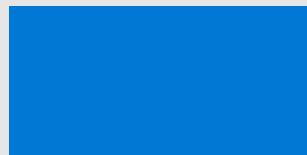
분석 환경 준비 및 보안까지 신경쓰기에는...

인프라 및 운영 담당자, 보안관리자들도
조직의 서버 및 소프트웨어 관리 시간이 부족합니다.
그리고, 통합적으로 관리하지 않으면
일관적이지 않고 예외가 있을 수 있습니다.

뭐든 시작은
데이터!

데이터는 얼마나 많이 늘어날까요?

Global Datasphere : 생성, 캡처, 복사한 모든 데이터의 총량



2018년 33 ZettaBytes



2025년 175 ZettaBytes

출처 : <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-korea-whitepaper.pdf>



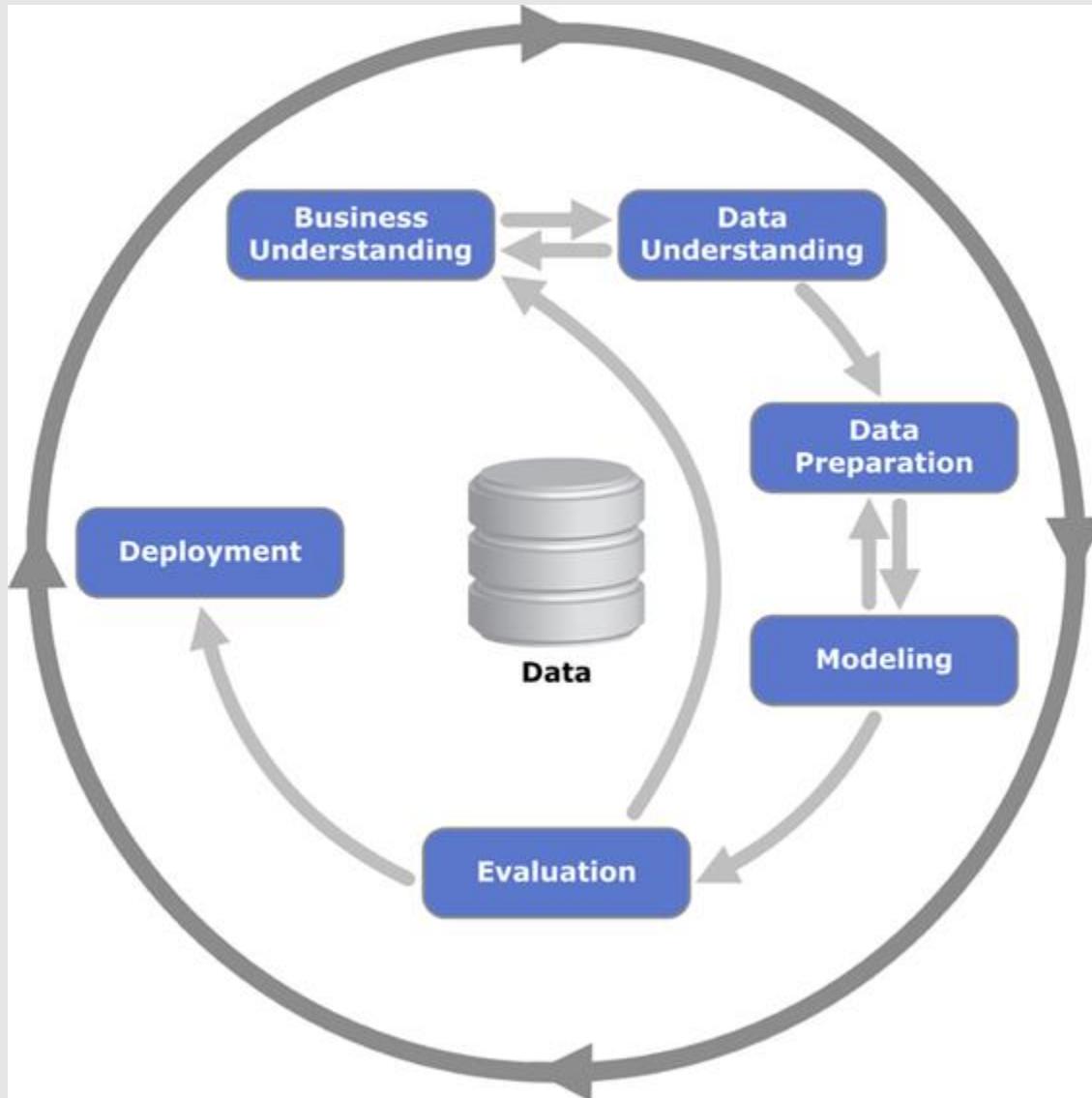
5 PetaBytes
(5,242,880
GigaBytes)

Yotta	Y	10^{24}	= 1 000 000 000 000 000 000 000 000
Zetta	Z	10^{21}	= 1 000 000 000 000 000 000 000 000
Exa	E	10^{18}	= 1 000 000 000 000 000 000 000 000
Peta	P	10^{15}	= 1 000 000 000 000 000 000
Tera	T	10^{12}	= 1 000 000 000 000
Giga	G	10^9	= 1 000 000 000
Mega	M	10^6	= 1 000 000
Kilo	K	10^3	= 1 000
Hecto	H	10^2	= 100
Deca	Da	10^1	= 10

참으로
데이터가
많군요!

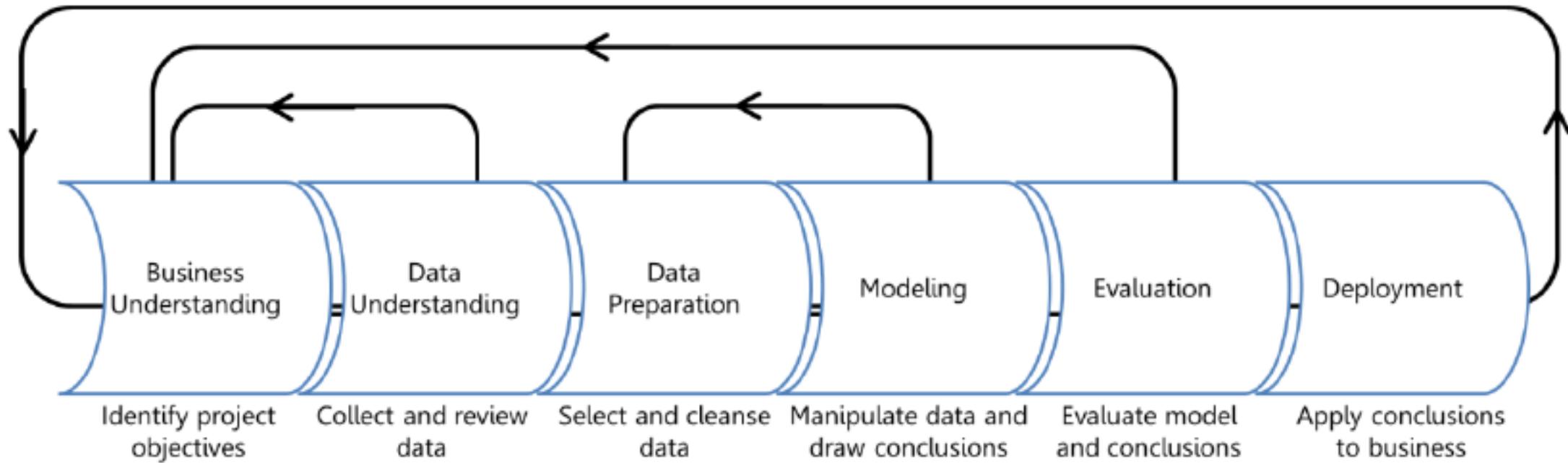
좁은 의미부터

CRISP-DM (Cross-industry standard process for data mining)



출처 : <https://data.sngular.com/en/art/40/crisp-dm-the-methodology-to-put-some-order-into-data-science-projects>

The lifecycle of Data for AI



좀더
체계적으로!

DataOps - <https://www.dataopsmanifesto.org/>



DataOps Manifesto

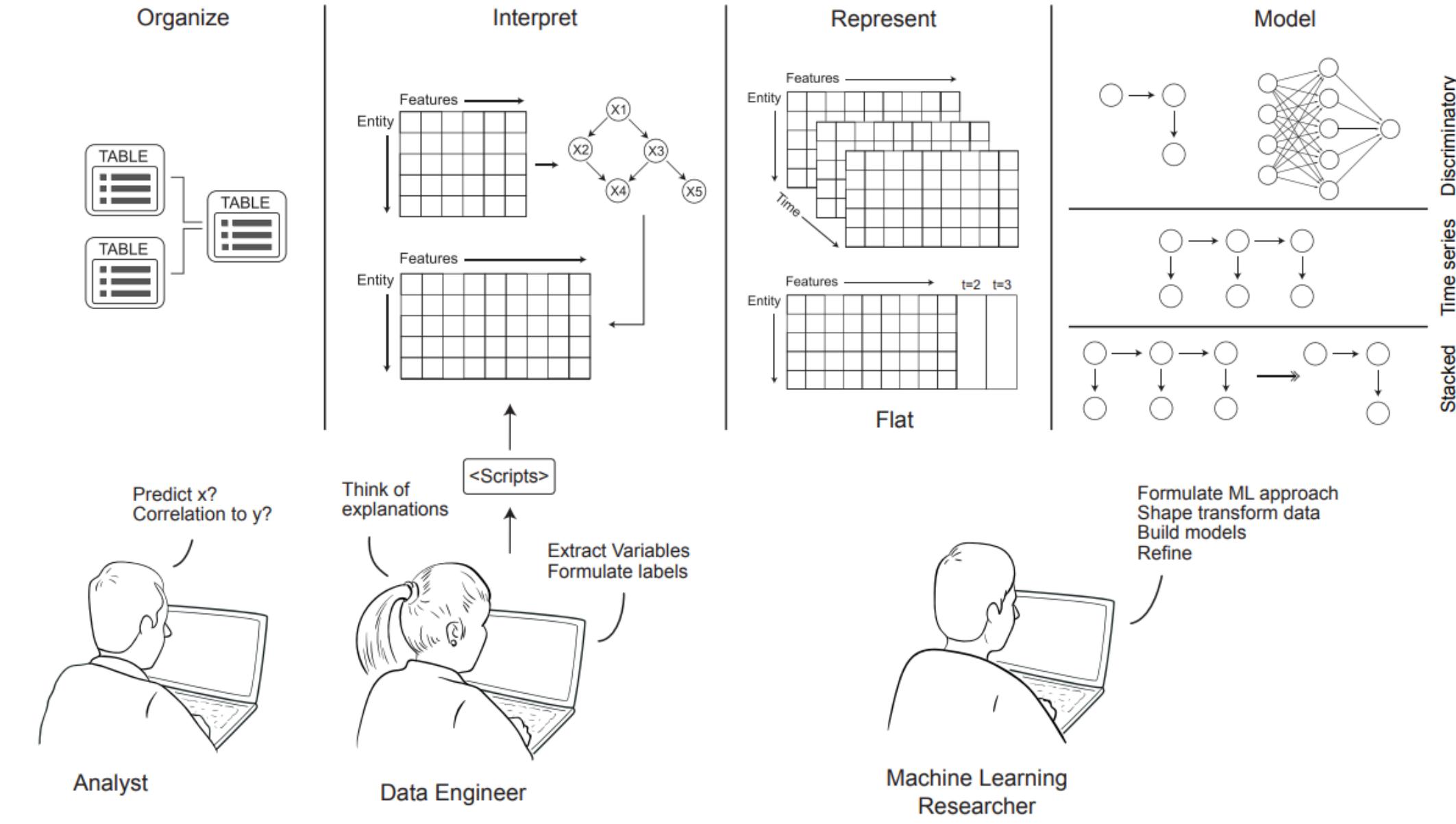
조직, 도구 및 산업에 걸쳐 데이터로 작업하는 직접적 경험을 통해 우리는 우리가 DataOps라고 부르는 분석을 개발 및 전달하기 위한 보다 좋은 방법을 알아냈습니다.

데이터 과학, 데이터 엔지니어링, 데이터 관리, 빅 데이터, 비즈니스 인텔리전스라고 언급되든 아니든 우리의 작업을 통해 우리는 분석의 가치에 도달하게 되었습니다:

- 프로세스 및 도구보다 개인 및 상호 작용
- 포괄적 문서보다 작동하는 분석
- 계약 협상보다 고객 협업
- 광범위한 업프런트 설계보다 실험, 반복 및 피드백
- 단절된 책임보다 교차 기능적인 운영의 소유

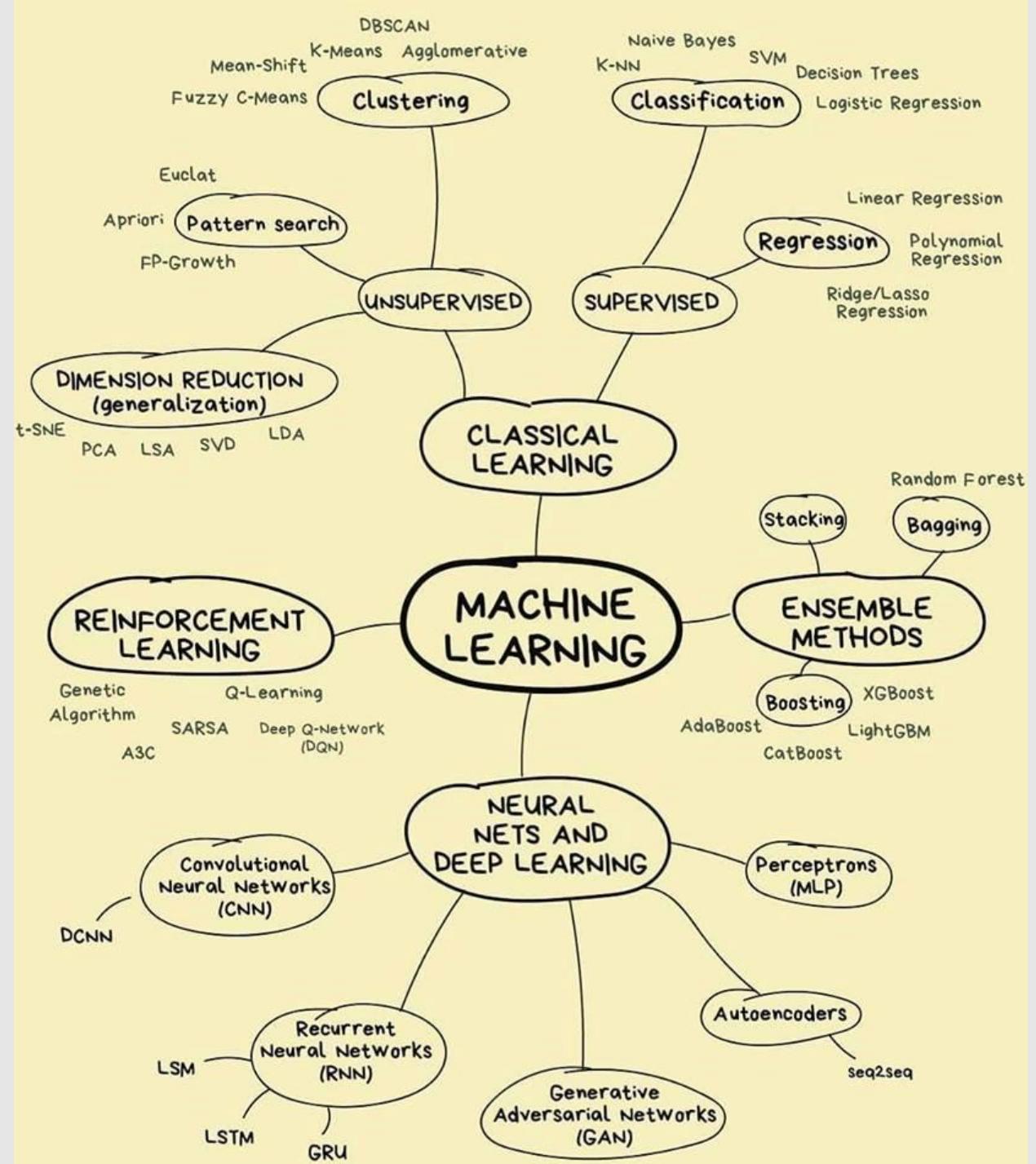
좀더 이쁘고
전문적으로

Deep Feature Synthesis: Towards Automating Data Science Endeavors



이쁘게 잘
정리해주셨는데...

왠지 겁나고
내 일이 아닌 것 같습니다



너도 해보고
나도 해보는

데이터 분석의
민주화!

Data Analysis
Democracy!

Automated ML



WIKIPEDIA
The Free Encyclopedia

Not logged in Talk Contributions Create account Log in

Article Talk

Read Edit View history

Search Wikipedia



Automated machine learning

From Wikipedia, the free encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

This article has multiple issues. Please help improve it or discuss these issues on the talk page. (Learn how and when to remove these template messages) [hide]



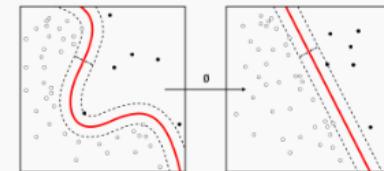
- This article may be incomprehensible or very hard to understand. (November 2018)
- This article may be too technical for most readers to understand. Please help improve it to make it understandable to non-experts, without removing the technical details. (November 2018)
- This article is in list format, but may read better as prose. (March 2018)

Automated machine learning (AutoML) is the process of automating the end-to-end process of applying machine learning to real-world problems. In a typical machine learning application, practitioners must apply the appropriate data pre-processing, feature engineering, feature extraction, and feature selection methods that make the dataset amenable for machine learning. Following those preprocessing steps, practitioners must then perform algorithm selection and hyperparameter optimization to maximize the predictive performance of their final machine learning model. As many of these steps are often beyond the abilities of non-experts, AutoML was proposed as an artificial intelligence-based solution to the ever-growing challenge of applying machine learning.^{[1][2]} Automating the end-to-end process of applying machine learning offers the advantages of producing simpler solutions, faster creation of those solutions, and models that often outperform models that were designed by hand.

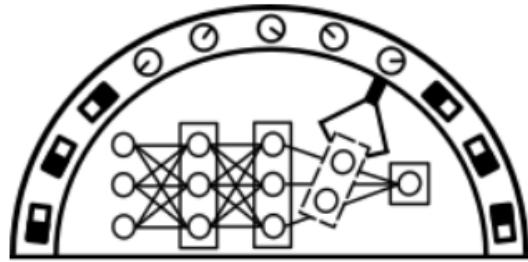
Contents [hide]

- 1 Targets of automation
- 2 Examples
 - 2.1 Hyperparameter optimization and model selection
 - 2.2 Full pipeline optimization
 - 2.3 Deep neural network architecture search

Machine learning and data mining



- Problems** [show]
Supervised learning (classification • regression) [show]
Clustering [show]
Dimensionality reduction [show]
Structured prediction [show]
Anomaly detection [show]



AutoML.org Freiburg

Search

Follow Us



AutoML Freiburg

Home

Blog

AutoML ▾

AAD ▾

Analysis ▾

Book

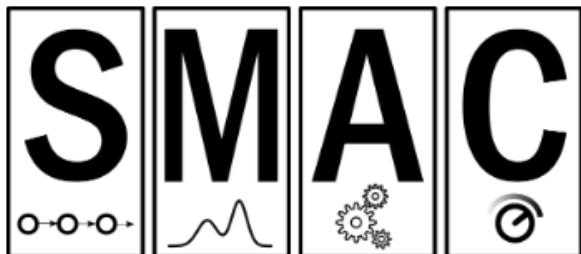
Events

Team & Partners ▾

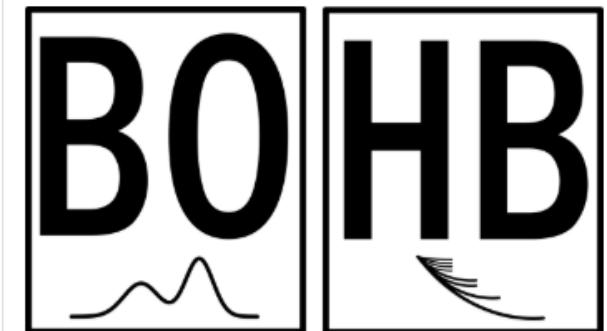
AutoML ...

<https://www.automl.org/>

provides methods and processes to make Machine Learning available for non-Machine Learning experts, to improve efficiency of Machine Learning and to accelerate research on Machine Learning. Machine learning (ML) has achieved considerable successes in recent years and an ever-growing number of disciplines rely on it. However, this success crucially relies on human machine learning experts to perform manual tasks. As the complexity of these tasks is often beyond non-ML-experts, the rapid growth of machine learning applications has created a demand for off-the-shelf machine learning methods that can be used easily and without expert knowledge. We call the resulting research area that targets progressive automation of machine learning AutoML.



SMAC





Books ▾



24/7 Customer service. Learn more

Deliver to
Korea, Republic of

Departments ▾

Today's Deals

Your Amazon.com

Buy Again

Gift Cards

Help

Registry

Sell

EN

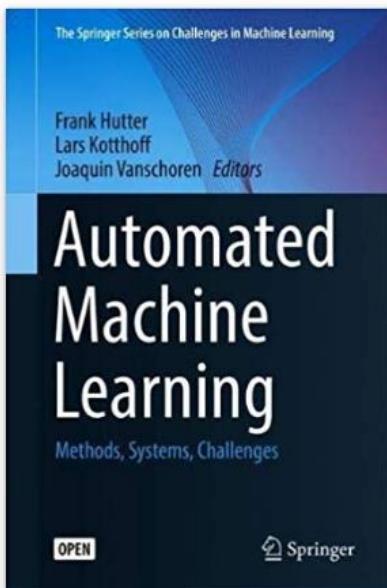
Hello, Sign in
Account & Lists ▾

Orders



Books Advanced Search New Releases Amazon Charts Best Sellers & More The New York Times® Best Sellers Children's Books Textbooks Textbook Rentals Sell Us Your Books Best Books of the Month Kindle eBooks

◀ Back to results



See this image

Automated Machine Learning: Methods, Systems, Challenges (The Springer Series on Challenges in Machine Learning)

Hardcover – July 10, 2019

by Frank Hutter (Editor), Lars Kotthoff (Editor), Joaquin Vanschoren (Editor)

[See all formats and editions](#)**Hardcover**

\$53.61

1 New from \$53.61

prime student

College student? Get FREE shipping and exclusive deals

[LEARN MORE](#)

This open access book presents the first comprehensive overview of general methods in Automated Machine Learning (AutoML), collects descriptions of existing systems based on these methods, and discusses the first series of international challenges of AutoML systems. The recent success of commercial ML applications and the rapid growth of the field has created a high demand for off-the-shelf ML methods that can be used easily and without expert knowledge. However, many of the recent

[◀ Read more](#)

[Share](#) **Pre-order**

Qty: 1 ▾

\$53.61

List Price: \$59.99

Save: \$6.38 (11%)

Pre-order Price

Guarantee.

This title will be released on July 10, 2019.

Ships from and sold by Amazon.com.



Pre-order: Add to Cart



Pre-order now

This item ships to Korea; Republic of (South Korea). [Learn more](#)

Deliver to Korea, Republic of

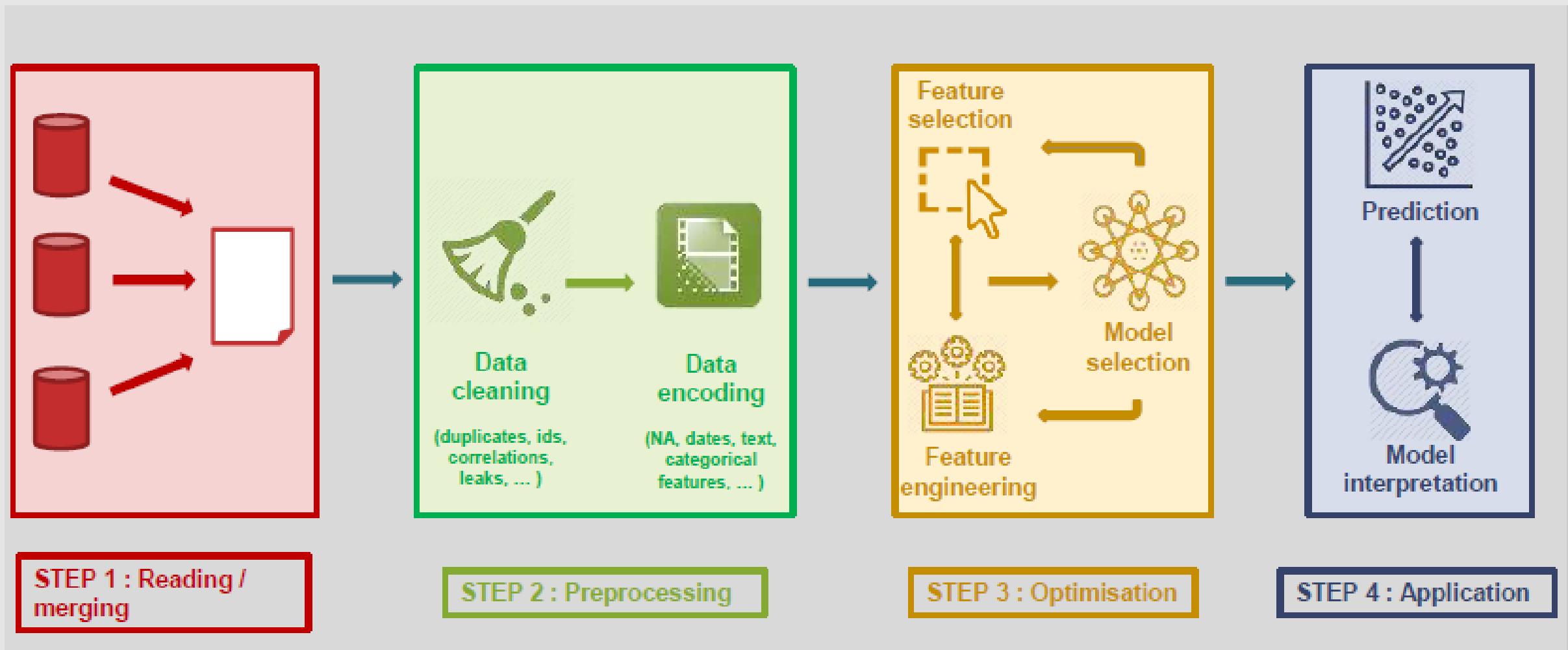
[Add to List](#)

Special offers and product promotions

- Pre-order Price Guarantee! Order now and if the Amazon.com price decreases between your order time and the end of the day of the release date, you'll receive the lowest price. [Here's how](#) ▾

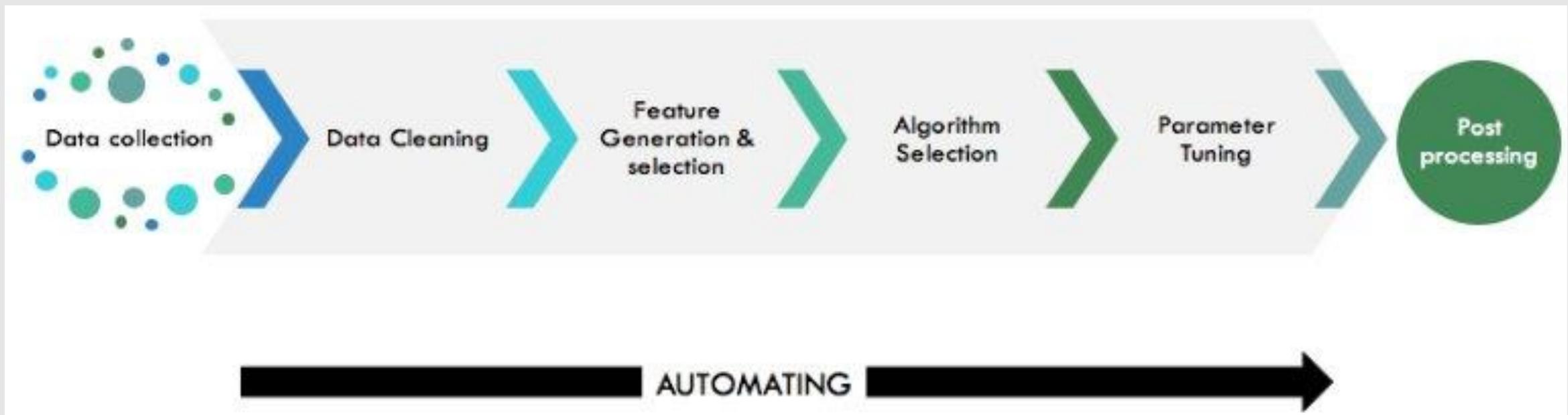
무료 다운로드 가능 : <https://link.springer.com/book/10.1007%2F978-3-030-05318-5>

일반적인 ML 파이프라인



출처 : AUTOMATED MACHINE LEARNING – Axel de Romblay
<https://www.slideshare.net/AxeldeRomblay>

이렇게 하고자하고 실현하고 있습니다!

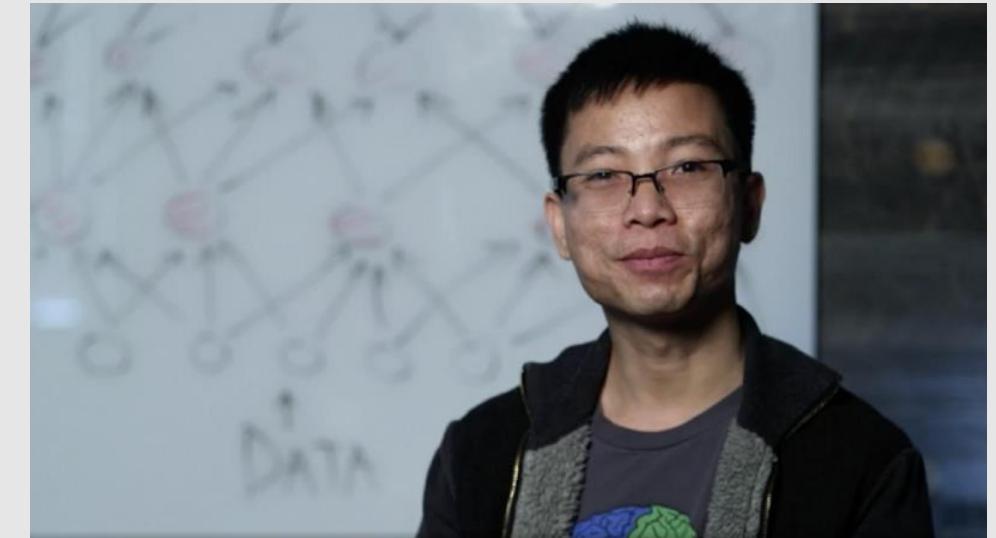
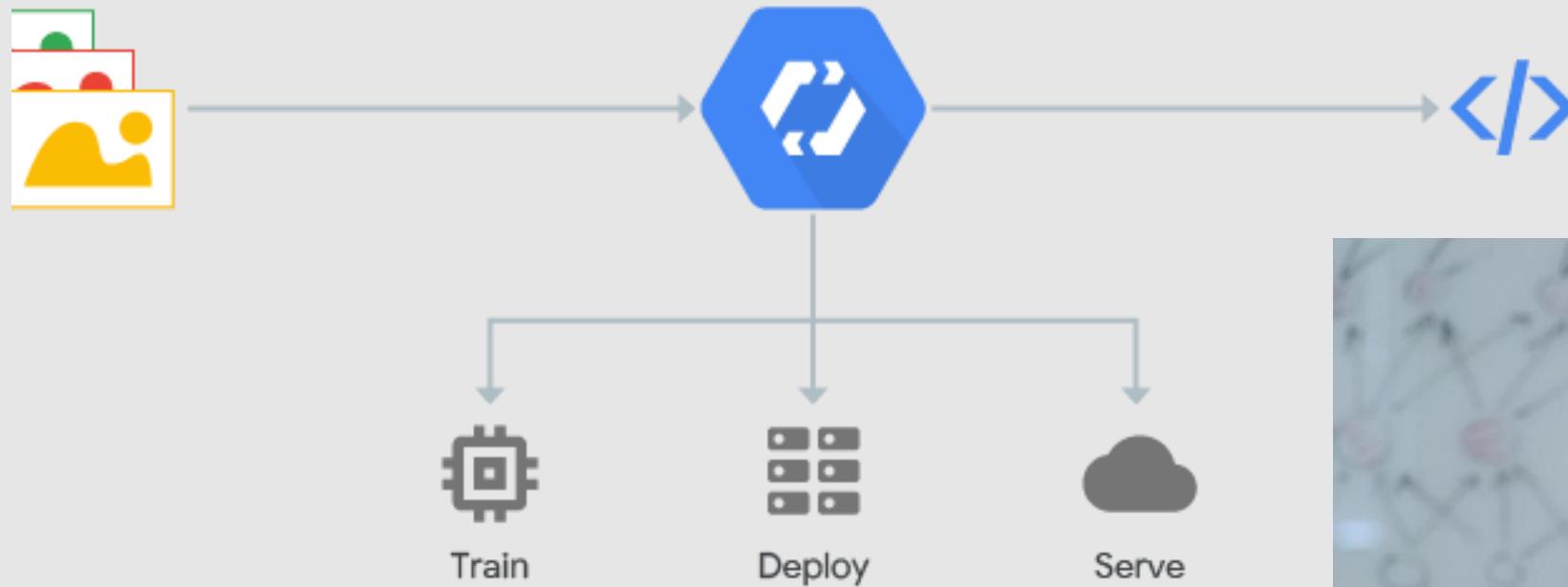


구글에서는 AutoML로

dataset

AutoML

Generate prediction
with a REST API



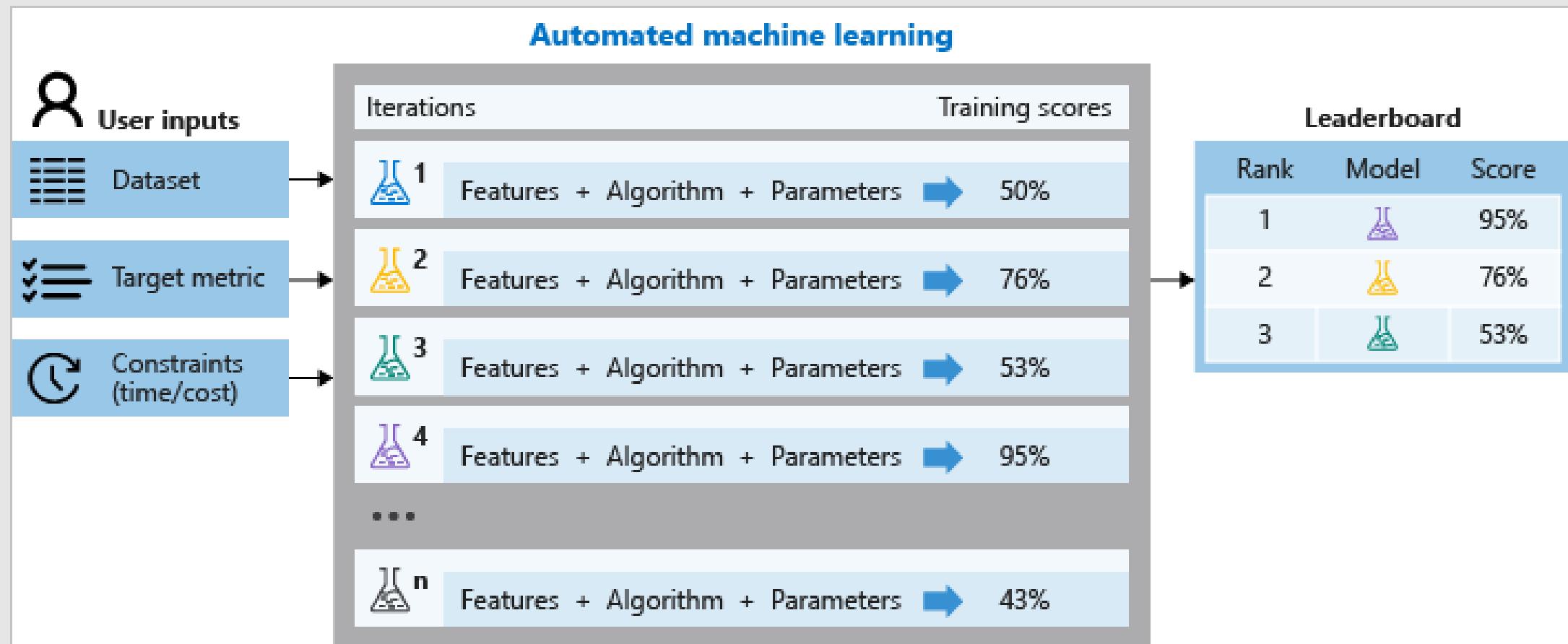
Quoc Le

<https://arxiv.org/abs/1112.6209>

MS Azure에서는

Azure Machine Learning service로!

오늘 F 트랙에서 13~15시에 전미정님께서 실습해 주셨습니다!





An open source platform for the machine learning lifecycle

Latest News

Bay Area MLflow Meetup @ Microsoft in Sunnyvale, CA (13 Jun 2019)

MLflow 1.0 Released! (04 Jun 2019)

MLflow 1.0 Release Candidate Available (22 May 2019)

MLflow 0.9.1 Released! (21 Apr 2019)

[News Archive](#)

<https://mlflow.org/>



WORKS WITH ANY ML LIBRARY, LANGUAGE &



RUNS THE SAME WAY IN ANY CLOUD



DESIGNED TO SCALE FROM 1 USER TO LARGE OPS



SCALES TO BIG DATA WITH APACHE SPARK™

Ready-to-use AI Powered Applications

Invest a little and gain a lot with our data science, business-focused use cases, combining Signal Processing and AI to deliver previously impossible levels of Accuracy, Stability, and Performance.

[Start a trial](#)

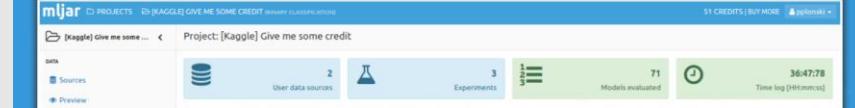
mljar Features Pricing Docs Academy Blog

Machine Learning for Humans!

Automated machine learning platform
With mljar anyone can train great machine learning models!

Get free credits and start today!

[Test it out](#) [Learn more](#)



Features Resources Blog Login Request a demo

The most comprehensive deep learning platform to manage experiments, data and resources more frequently, at scale and with greater confidence.

Enter your work email [Request a demo](#)

Watch how it works (1 min)

Train computer vision models faster

aidoc nanit SAMSUNG iotivity WPS

AutoML 스타트업

- Binah : <https://www.binah.ai/>
- DataRobot : <https://www.datarobot.com/>
- BigML : <https://bigml.com/>
- MissingLink : <https://missinglink.ai/>
- mljar : <https://mljar.com/>

bigml PRODUCT ▾ GETTING STARTED PRICING ▾ SUPPORT

Machine Learning made beautifully simple for everyone

Take your business to the next level with the leading Machine Learning platform

Sign up with [Amazon](#) [GitHub](#) [Google](#)
or
Email
Korea, Republic of
Username Password
SIGN UP

By signing up, you agree to our [terms of service](#) and [privacy policy](#). We'll occasionally send you BigML related emails.

Australian or New Zealand users sign up here
Already registered on BigML? [Sign in](#)

PLATFORM SOLUTIONS RESOURCES ▾ PARTNERS ▾ ABOUT ▾ WE'RE HIRING 41+ CONTACT US

Enabling the AI-Driven Enterprise with Automated Machine Learning

LEARN HOW DATAROBOT EMPOWERS YOU SEE OUR PLATFORM IN ACTION

DataRobot works for... Data Analysts Advance your career with automated machine learning Data Scientists Amplify your data science skills Executives Become an AI-driven enterprise Software Easily add AI to your applications

We use cookies and other tracking technologies to improve your experience on our site. To learn more, please read our update policy. By choosing "OK", you consent to our use of cookies and tracking technologies. [Find out more](#).

Automated ML 애플리케이션 1

- ABM: Automatic Business Modeler : <http://e-abm.com/>
- AutoDiscovery from ButlerScientifics : <http://www.butlerscientifics.com/>
- Automatic Statistician project : <http://www.automaticstatistician.com/index/>
- auto-sklearn : <https://github.com/automl/auto-sklearn>
- Auto-WEKA : <http://www.cs.ubc.ca/labs/beta/Projects/autoweka/>
- Bicedeep AI : <https://bicedeepai.com/>
- Darwin : <https://www.sparkcognition.com/product/darwin>
- DataRobot : <https://www.datarobot.com/>
- DataRPM Cognitive Data Science platform : <http://www.datarpm.com/>
- DMWay : <http://dmway.com/>

Automated ML 애플리케이션 2

- dotData : <https://dotdata.com/>
- Emcien : <http://emcien.com/>
- ForecastThis DSX : <http://forecastthis.com/platform>
- FeatureLab : <http://www.featurelab.co/how-it-works>
- Kogentix Automated Machine Learning Platform : <http://kogentix.com/>
- Loom Systems : <http://www.loomsystems.com/>
- machineJS : <https://github.com/ClimbsRocks/machineJS>
- MLJAR : <https://mljar.com/>
- Quill from Narrative Science : <https://www.narrativescience.com/quill>
- SAP Predictive Analytics : <http://help.sap.com/pa>

Automated ML 애플리케이션 3

- Savvy from Yseop : <https://savvy.yseop.com/>
- Skytree Machine Learning Software : <http://www.skytree.net/>
- Tree-based Pipeline Optimization Tool (TPOT) : <https://github.com/rhiever/tpot>
- Wordsmith from Automated Insights : <https://automatedinsights.com/products/>
- Xpanse Analytics : <http://www.xpanseanalytics.com/>
- H2O : <https://www.h2o.ai/products/h2o/>

Maxs-MacBook-Pro:automl-test maxwoolf\$

```
from tpot import TPOTClassifier
from sklearn.datasets import load_digits
from sklearn.model_selection import train_test_split

digits = load_digits()
X_train, X_test, y_train, y_test = train_test_split(digits.data, digits.target,
                                                    train_size=0.75, test_size=0.25)
```

```
tpot = TPOTClassifier(generations=5, population_size=50, verbosity=2, n_jobs=-1)
tpot.fit(X_train, y_train)
```

```
Optimization Progress:  0% |  0/300 [00:00<?, ?pipeline/s]
```

```
print(tpot.score(X_test, y_test))
```

Women

Men

Toys

Kids

Home

Vintage

Beauty

Electronics

Sports

Handmade

Other

The Selling App.

Sell or buy. Almost anything.

[Sell now](#)

Sell it.

List in minutes. Take a few photos. Add a description. Set your price.

Ship it.

No meetups. Printable shipping label emailed to seller.

Get paid.

Listings are free. Flat 10% selling fee charged when sale completes.

Sold in the last 10 minutes.

<https://www.mercari.com/>



[AutoML1 \(2015-2016\)](#)[AutoML2 \(2017-2018\)](#)[AutoML3 \(current\)](#)

Navigation

[Home](#)

Home

[How to cite us?](#)[Winning Software](#)[Data](#)[Platform](#)

EVENTS

[Competition@NeurIPS2...](#)[Workshop@PRICAI2018](#)[Workshop@ICML2018](#)[Workshop@PAKDD2018](#)[Workshop@ICML2017](#)[Competitions@WCCI2016](#)[Workshop@ICML2016](#)[GPUtrack@ICML2016](#)[Workshop@NIPS2015](#)[Bootcamp@Stanford2015](#)[MLIschool@Petersburg2...](#)[Hackathon@ICML2015](#)[Workshop@ICML2015](#)[Hackathon@ESPCI2015](#)[Workshop@ICML2014](#)[BeatAutoSKLearn \(v1\)](#)[AutoSKLearn worksheet](#)

LINKS

[CiML](#)[Fact sheet](#)

Just started:

[AutoML 6 \(Automatic Computer Vision = AutoCV\),](#)
part of the AutoDL series, 28,000 USD in prizes.

Still running:

[AutoML5 : KDDCup2019 - AutoML for Temporal
Relational Data](#)
30,000USD+ in prizes

Recently ended:

[AutoML4 \(rematch of AutoML3\)](#)

News:

Now running: KDDCup - AutoML on
temporal relational data [[Competition](#)]

Upcoming Spring 2019
AutoDL challenge [[SIGN UP](#)]

December 7, 2019
[Competition Workshop at NeurIPS 2019](#)

August 28, 2018
We had a [workshop at PRICAI 2018](#).

July 14-15 2018:
We had a nice [workshop at ICML 2018](#).

March 2018:
**Our next competition on Life Long ML is
accepted to NIPS 2018.**

June 21, 2016: Microsoft published a [BLOG](#)
post on AutoML 1.

Why? Automated ML?

AI and machine learning is still a field with high barriers to entry that requires expertise and resources that few companies can afford on their own, Today, while AI offers countless benefits to businesses, developing a custom model often requires rare expertise and extensive resources.

– Fei-Fei Li, 2018 Google press event



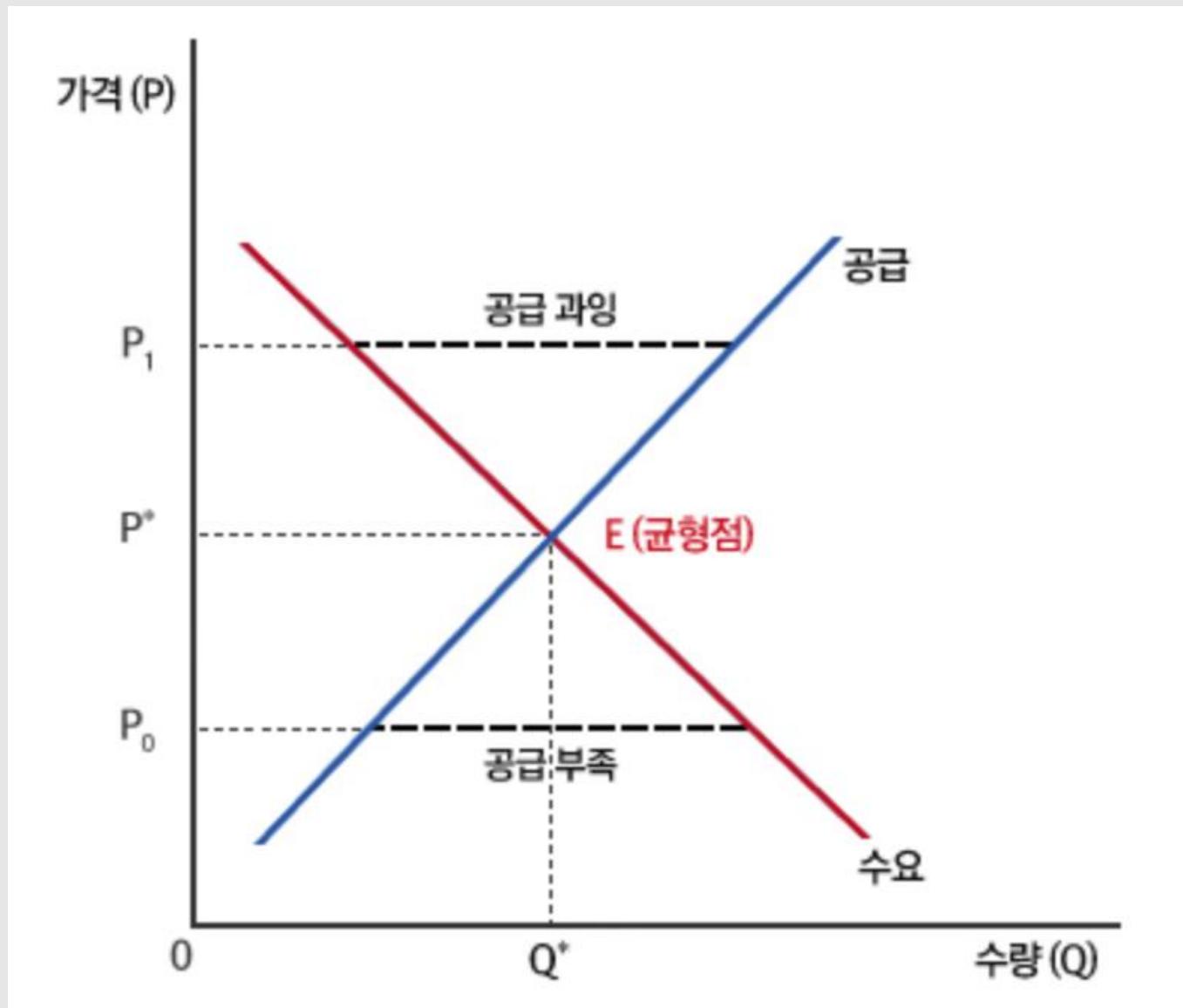
인공 지능과 기계 학습은 여전히 기업이 스스로 제공 할 수 있는 전문 지식과 자원을 필요로 하는 진입 장벽이 매우 높습니다. 오늘날 AI는 기업에 수많은 이점을 제공하지만 커스텀 모델을 개발하려면 종종 전문 지식과 광범위한 자원이 필요합니다.

– 리페이페이, 2018 구글 프레스 이벤트

Google CEO Sundar Pichai [wrote](#), "Today, designing neural nets is extremely time intensive, and requires an expertise that limits its use to a smaller community of scientists and engineers. That's why we've created an approach called AutoML, showing that it's possible for neural nets to design neural nets. We hope AutoML will take an ability that a few PhDs have today and **will make it possible in three to five years for hundreds of thousands of developers to design new neural nets for their particular needs.**" (emphasis mine)



왜? Automated ML? **수요와 공급의 문제**가 있습니다.



놈! 놈! 놈! 이 필요합니다

프로그래밍을 좀 하는 놈!

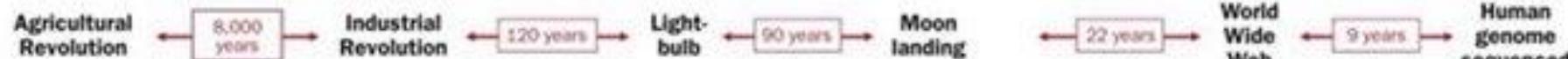
수학 지식이 좀 있는 놈!

도메인을 알고 통계 및 분석을 할 줄 아는 놈!

근데... 있나요? 또 연봉은 얼마인가요?

투자의 문제도

1 The accelerating pace of change ...



2 ... and exponential growth in computing power ...

Computer technology, shown here climbing dramatically by powers of 10, is now progressing more each hour than it did in its entire first 90 years

COMPUTER RANKINGS

By calculations per second per \$1,000



Analytical engine
Never fully built, Charles Babbage's invention was designed to solve computational and logical problems



Colossus

The electronic computer, with 1,500 vacuum tubes, helped the British crack German codes during WW II



UNIVAC I

The first commercially marketed computer, used to tabulate the U.S. Census, occupied 943 cu. ft.



Apple II

At a price of \$1,298, the compact machine was one of the first massively popular personal computers



Power Mac G4

The first personal computer to deliver more than 1 billion floating-point operations per second

3 ... will lead to the Singularity

Surpasses brainpower equivalent to that of all human brains combined

Surpasses brainpower of human in 2023



Surpasses brainpower of mouse in 2015

1900 1920 1940 1960 1980 2000 2020 2045

ELECTROMECHANICAL

RELAYS

VACUUM TUBES

TRANSISTORS

INTEGRATED CIRCUITS

0.00001

1

$100,000$

10^{20}

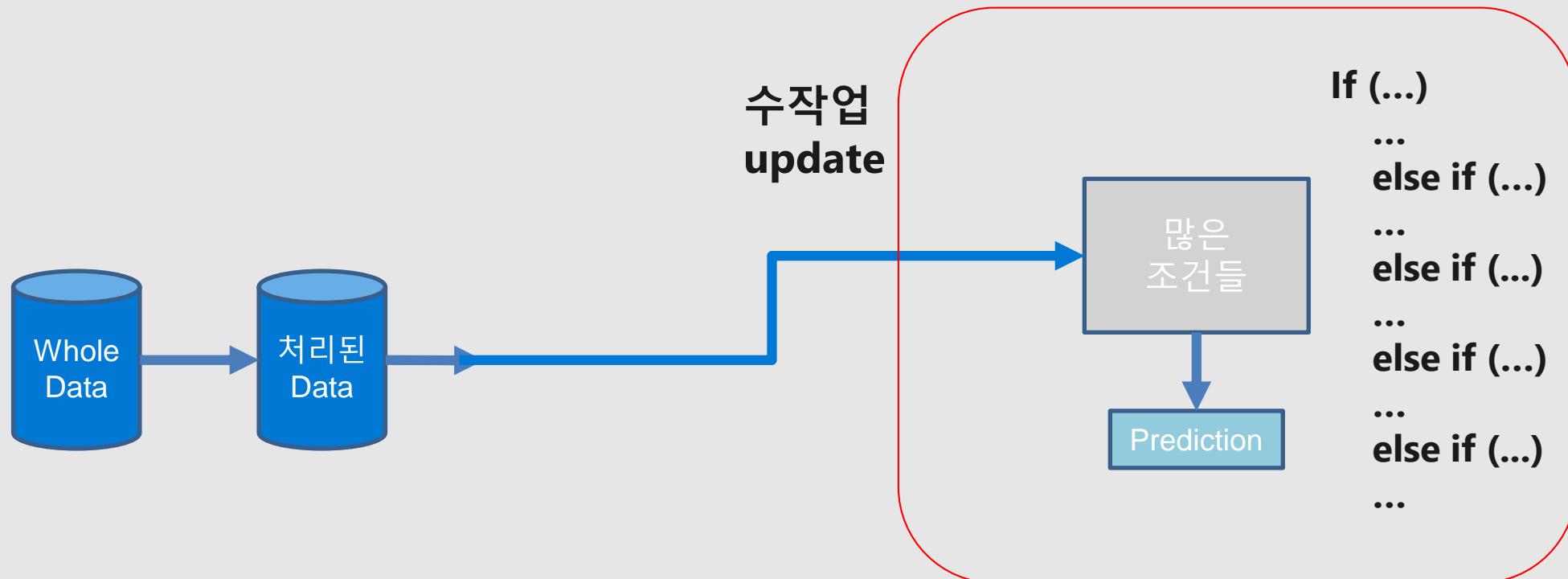
10^{15}

10^{26}

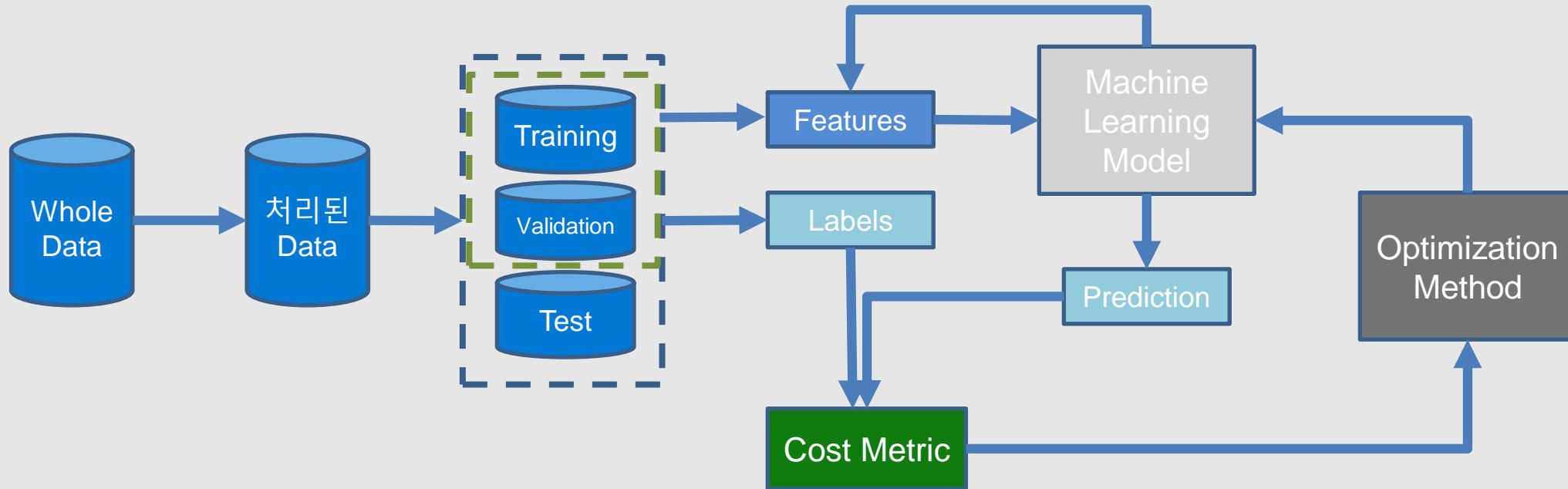


문제 해결 방식도

기존 방식의 데이터 분석



Machine Learning이라면



SBS

그런데 말입니다

3-5500

넓은 의미의 관점에서

누구나 할 수 없는 머신러닝 – 과연?

월 오후 12:49



○ㅅ○...?

저의 생각으로는..

머신러닝 하는게 재밌다고 생각해야지 되는듯요
수학을 잘해도 머신러닝은 재미있을수 있을듯...

○○ 나조

나도 그런 생각



그리고 수학이라는 것도 매우 거대망상한 학문의 큼직한 분야 인데..
수학을 잘해야지된다 하면 매우 두리뭉술 할듯요

수학을 잘 해야한다는 것보다 데이터를 이해하고 생각하는 것을 난 강조
하는데

아직 저분처럼 다르게 생각하시는 분도 있는 것 같음 ㅠㅠ

7월 세미나때는 이런 관점들도 반영하면 좋을듯! ㅎㅎ



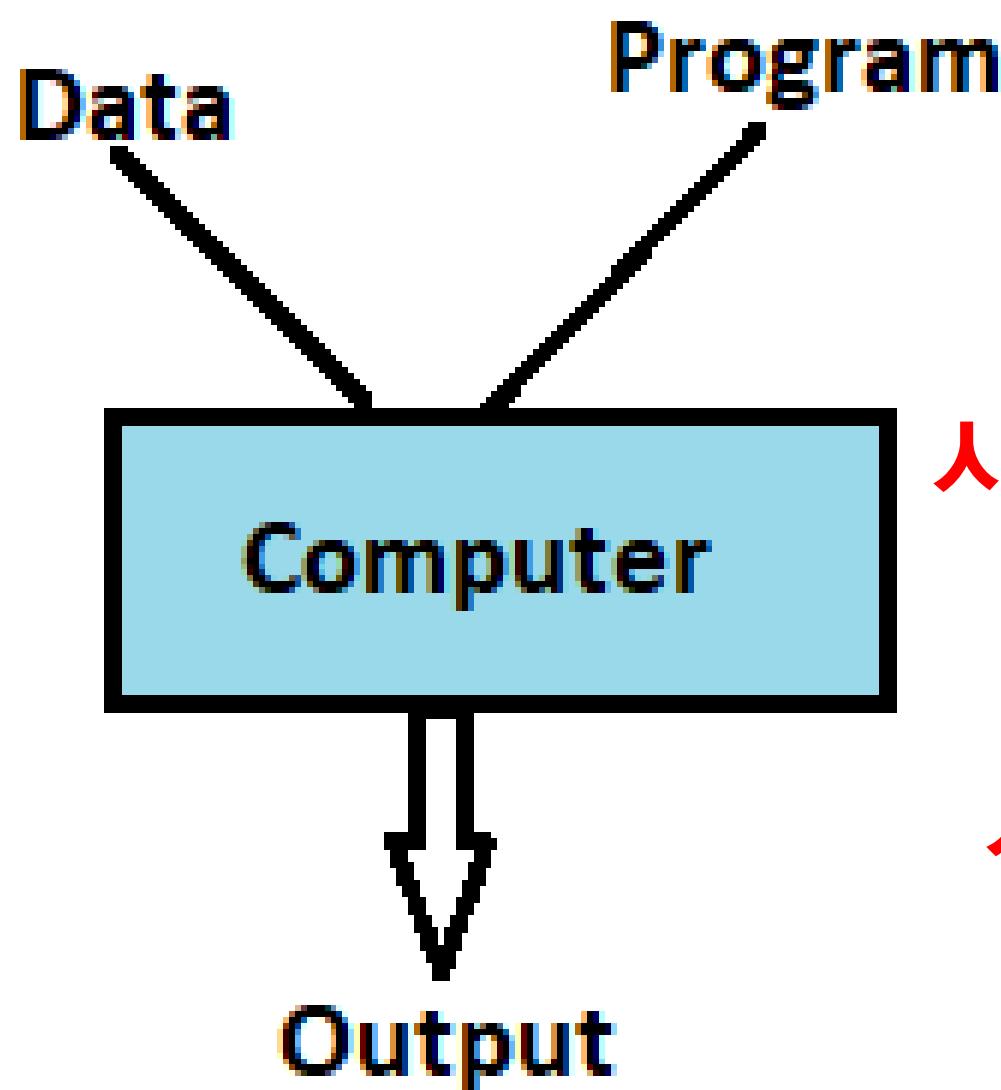
○ㅅ○....

해석학 한다고 머신러닝 잘하는 것은 아니므로..

○○ 저러한 관점을 고심 해서 만들어볼게요 ㅎㅎ

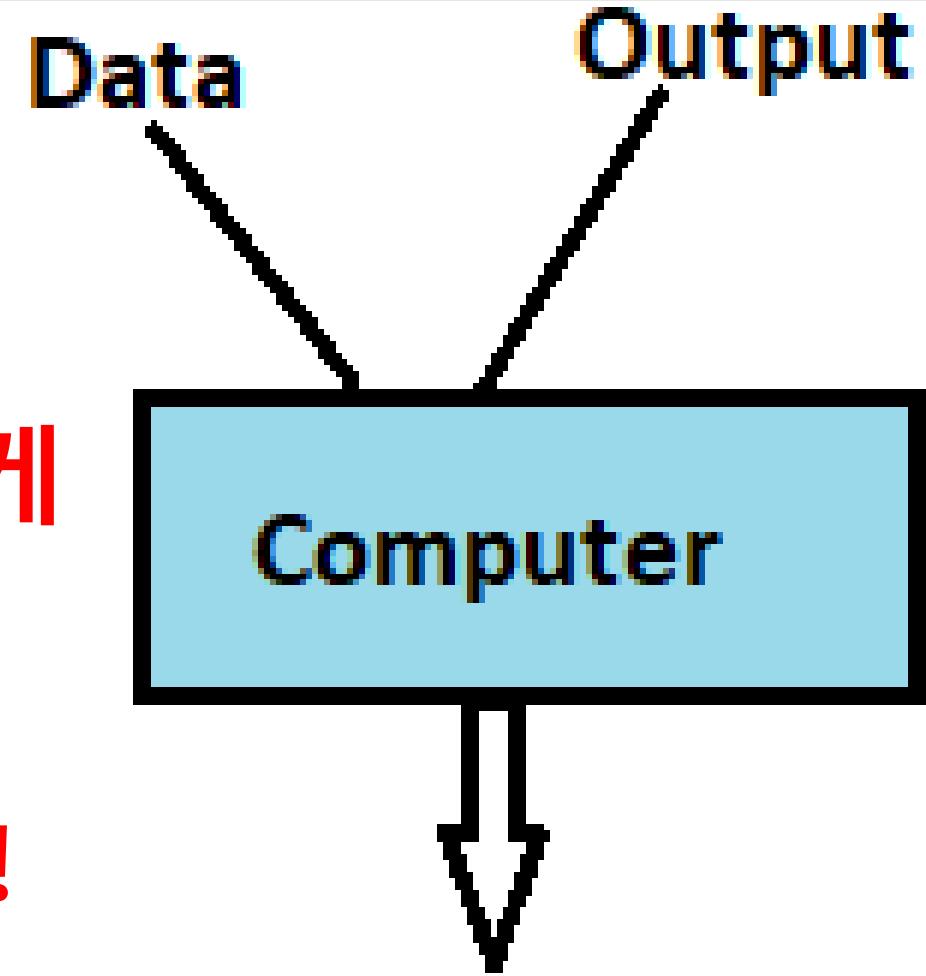
당근! ㅎㅎ

오늘 C 트랙에서
발표하시는
홍원의 님과의 대화



Simple Algorithm

사실 이렇게
간단하게
하고
싶은건데!



Machine Learning



AI의 구조를 쉽게 이해할 수 있는 딥러닝 초(超)입문

엑셀로 배우는 딥러닝

와쿠이 요시유키 · 와쿠이 사다미 저 · 관기태 역



첫 AI 학습에 최적!

어려운 수학 계산은 엑셀(Excel)에 맡기고
딥러닝의 구조를 엑셀을 동작시키면서 이해
할 수 있다!

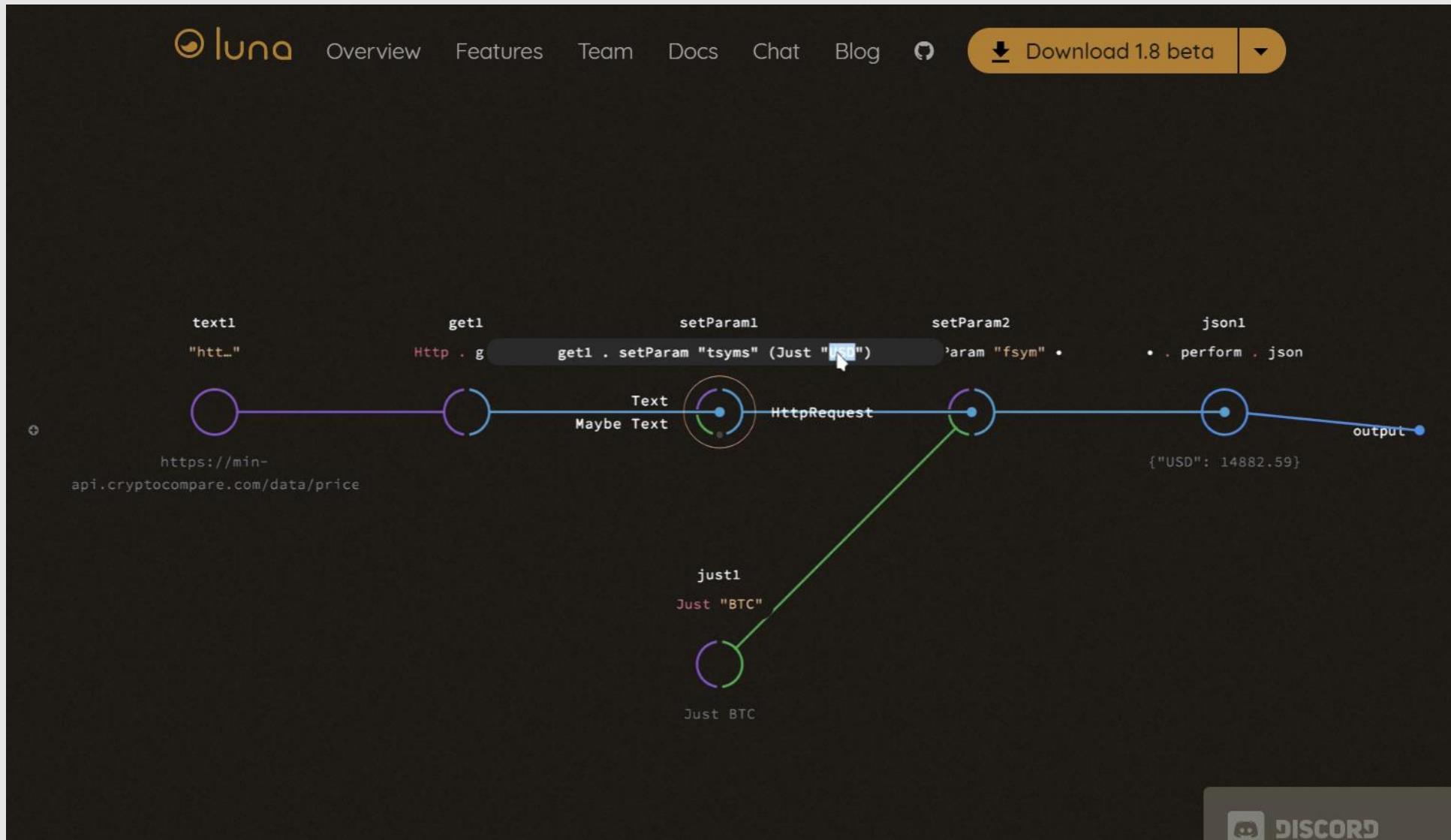


BM 성안당
www.cyber.co.kr

Margaret Hamilton; lead software engineer of the Apollo Project, stands next to the code she wrote by hand that was used to take humanity to the moon. [1969]



Visual Programming Language / luna <https://luna-lang.org>



Visual Programming Language / Ballerina <https://ballerina.io>

Cloud Native Programming Language

Learn Philosophy Central Community Help Blog 

Ballerina

Ballerina makes it easy to write microservices that integrate APIs.

[SEE WHY >](#)



Download Ballerina

Distributions available for Linux, OS X, and Windows

INTEGRATION SYNTAX:

A compiled, transactional, statically and strongly typed programming language with textual and graphical syntaxes

NETWORKED TYPE SYSTEM:

A type system that embraces network payload variability with primitive, object, union, and tuple types

Example : <hello-service/hello_service.bal> 



SOURCE

INTERACTION

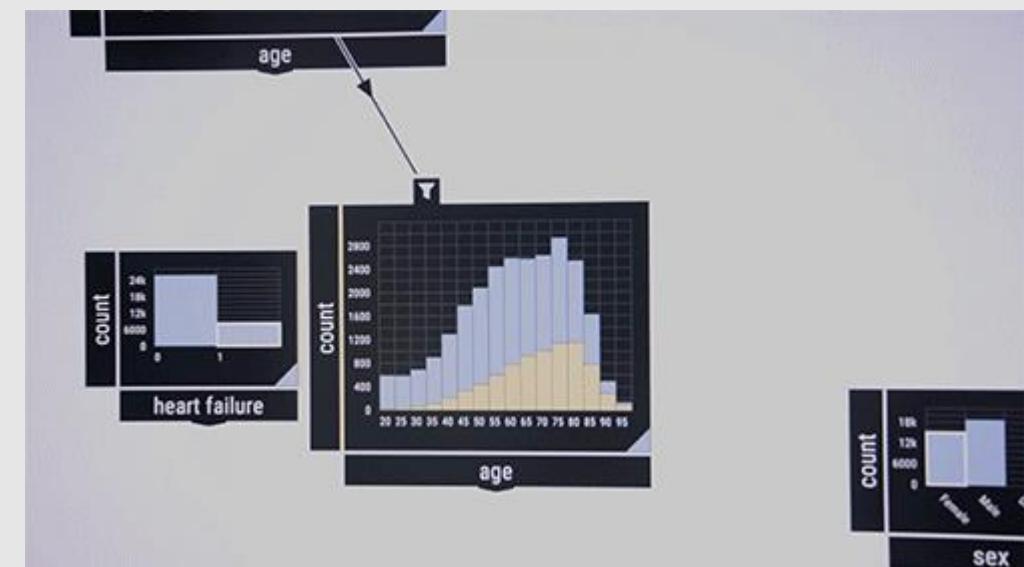
DEPLOYMENT

```
1 // The HTTP module provides implementations for connecting and
2 // interacting with HTTP, HTTP2, and WebSocket endpoints. This
3 // package is referenced with 'http' namespace in the code body.
4 import ballerina/http;
5 import ballerina/io;
6
7 // A service is a network-accessible entry point. This service
8 // is accessed at '/hello', and bound to a listener on port 9090.
9 service hello on new http:Listener(9090) {
10
11    // A resource is an API method which can be called by a listener.
12    // It is always visible to the listener to which the service is
13    // attached. This resource is accessed at '/hello/sayHello' and
14    // 'caller' is the client calling us.
15    resource function sayHello(http:Caller caller,
16                             http:Request request) {
17
18        // Create an object to carry data back to the caller.
19        http:Response response = new;
20
21        response.setBody("Hello " + caller.getName());
22
23        return response;
24    }
25}
```

```
curl -s http://playground.localhost/hello/sayHello
```

토니 스타크처럼 했으면 좋겠는데... 서비스?





<http://news.mit.edu/2019/drag-drop-data-analytics-0627>

<http://www.einblick.ai/>

Gen

A general-purpose probabilistic programming system with programmable inference.

Overview

Tutorials

Docs

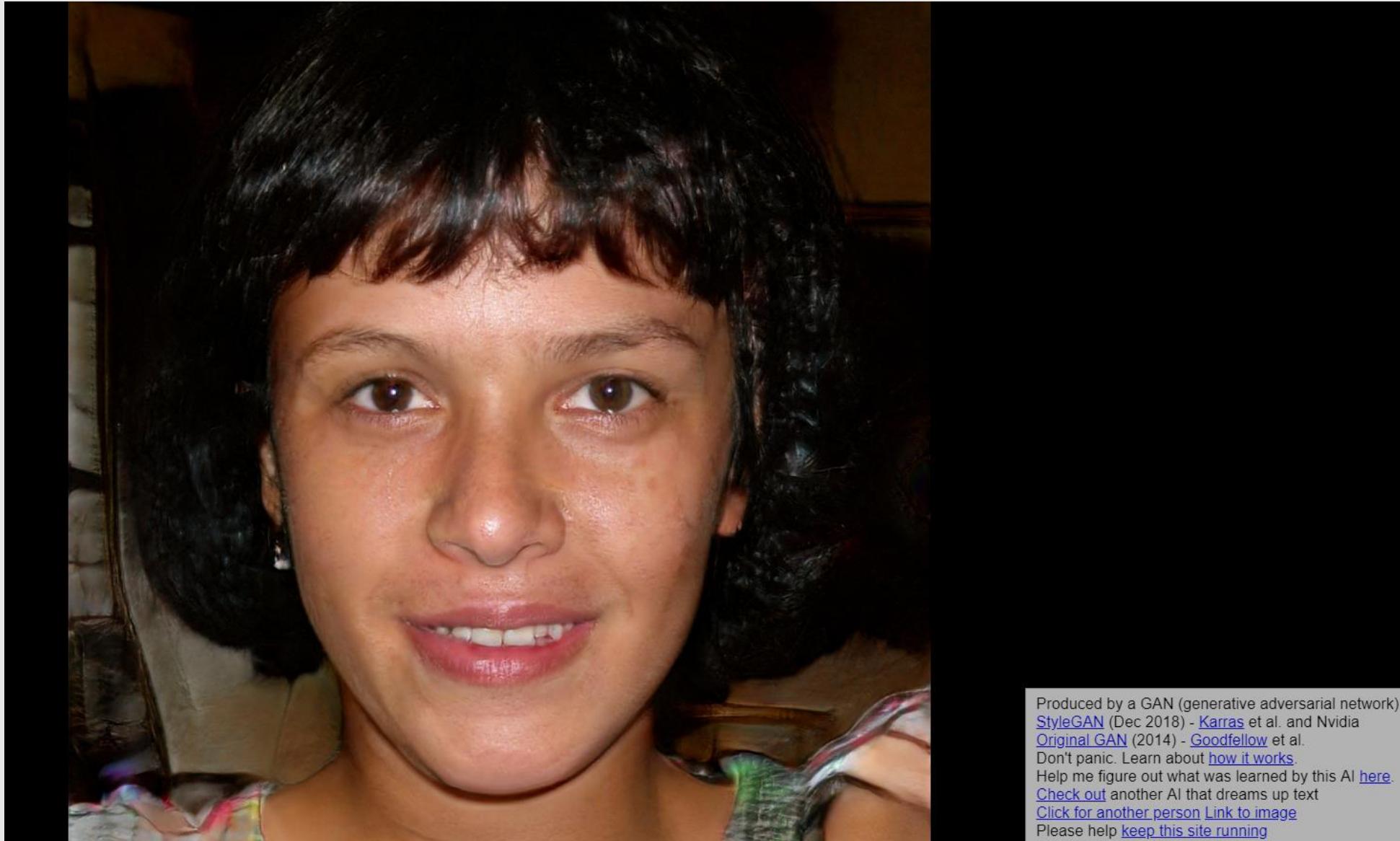
Source

Introduction

Probabilistic modeling and inference are core tools in diverse fields including statistics, machine learning, computer vision, cognitive science, robotics, natural language processing, and artificial intelligence. To meet the functional requirements of applications, practitioners use a broad range of modeling techniques and approximate inference algorithms. However, implementing inference algorithms is often difficult and error prone. Gen simplifies the use of probabilistic modeling and inference, by providing *modeling languages* in which users express models, and high-level programming constructs that automate aspects of inference.

<https://probcomp.github.io/Gen/>

<https://thispersondoesnotexist.com/>



Produced by a GAN (generative adversarial network)

[StyleGAN](#) (Dec 2018) - [Karras](#) et al. and Nvidia

[Original GAN](#) (2014) - [Goodfellow](#) et al.

Don't panic. Learn about [how it works](#).

Help me figure out what was learned by this AI [here](#).

[Check out](#) another AI that dreams up text

[Click for another person](#) [Link to image](#)

Please help [keep this site running](#)

<http://www.whichfaceisreal.com/index.php>

PLAY

ABOUT

METHODS

LEARN

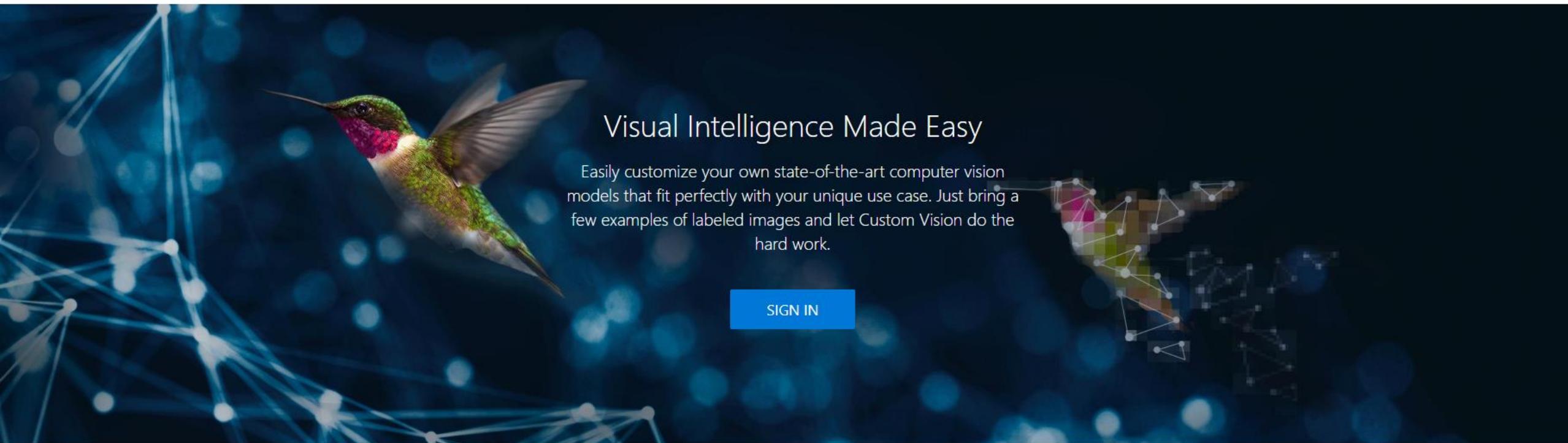
PRESS

CONTACT

CALLING BS

Click on the person who is real.

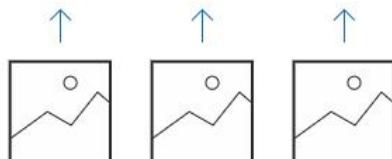




Visual Intelligence Made Easy

Easily customize your own state-of-the-art computer vision models that fit perfectly with your unique use case. Just bring a few examples of labeled images and let Custom Vision do the hard work.

SIGN IN



Upload Images

Bring your own labeled images, or use Custom Vision to quickly add tags to any unlabeled images.



Train

Use your labeled images to teach Custom Vision the concepts you care about.



Evaluate

Use simple REST API calls to quickly tag images with your new custom computer vision model.



[GitHub](#) [Docs](#) [Blog](#)

LUMINOTH

Open source Computer Vision toolkit

[GET STARTED!](#)



2,071



```
$ pip install luminoth
Successfully installed luminoth-0.2.0.

$ lumi train -c custom.yml
[INFO] Starting training for FasterRCNN.
[INFO] Step: 1, file: '524320.jpg', train_loss: 10.42, in 1.50s
[INFO] Step: 2, file: '524291.jpg', train_loss: 10.22, in 1.49s

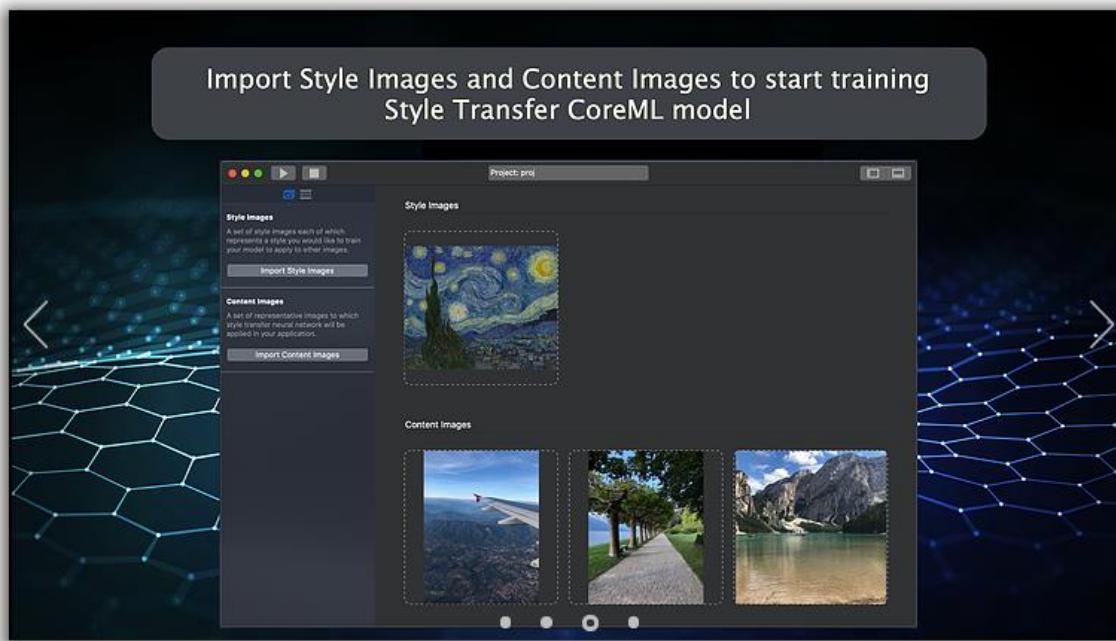
$ lumi predict -c custom.yml bicycles.jpg
{
  "objects": [
    {"bbox": [294, 231, 468, 536], "label": "person", "prob": 0.9997},
    {"bbox": [494, 289, 578, 439], "label": "person", "prob": 0.9971},
    [...]
  ]
}
```





MakeML <https://www.makeml.app/>

Create Object Detection CoreML model on Mac not writing a line of code



Wow! I want the app

👤 Name

✉ Email

Send me link

인공지능 게임 을 만들어봐요.

1

먼저 여러 데이터를 모아보세요

2

데이터를 사용하여 인공지
련시켜보세요

시작해봅시다

더 알아볼까요?

3

인공지능을 사용하여 스크
립트 만들어보세요

IBM Watson를 활용했어요!

The screenshot shows the main interface of the ML website. At the top, there's a navigation bar with the ML logo, '소개', '프로젝트', '워크시트', '뉴스', '도움말', and '로그아웃'. Below the title '인공지능 게임을 만들어봐요.' are three numbered steps:

- 먼저 여러 데이터를 모아보세요
- 데이터를 사용하여 인공지련시켜보세요
- 인공지능을 사용하여 스크립트 만들어보세요

Below step 3, there are three large callout boxes:

- 훈련**: '컴퓨터가 훈련할 수 있도록 다양한 데이터를 준비하세요.' with a '훈련' button.
- 학습 & 평가**: '데이터를 사용하여 컴퓨터를 학습시키세요. numbers' with a '학습 & 평가' button.
- 만들기**: '스크래치나 파이썬을 사용하여 여러분이 만든 머신러닝 모델로 게임이나 프로그램을 만들어보세요.' with a '만들기' button.

In the top right corner of the main area, it says '"Titanic survivors"'.

<https://machinelearningforkids.co.uk>

The screenshot shows the 'ML' website again, specifically a section for the 'numbers' dataset. It features two tables side-by-side:

survived		did_not_survive	
ticket class 3 gender male age 0.42 sibl. sp. 0 par. ch. 1 ticket fare 8.5167 embarked Cherbourg	ticket class 3 gender female age 0.75 sibl. sp. 2 par. ch. 1 ticket fare 19.2583 embarked Cherbourg	ticket class 3 gender female age 0.75 sibl. sp. 2 par. ch. 1 ticket fare 19.2583 embarked Cherbourg	ticket class 3 gender female age 9 sibl. sp. 1 par. ch. 1 ticket fare 15.2458 embarked Cherbourg
ticket class 3 gender female age 1 sibl. sp. 0 par. ch. 2 ticket fare 15.7417 embarked Cherbourg	ticket class 2 gender male age 1 sibl. sp. 2 par. ch. 2 ticket fare 37.0042 embarked Cherbourg	ticket class 2 gender female age 3 sibl. sp. 1 par. ch. 2 ticket fare 41.5792 embarked Cherbourg	ticket class 3 gender male age 11 sibl. sp. 0 par. ch. 0 ticket fare 18.7875 embarked Cherbourg
ticket class 3 gender female age 15 sibl. sp. 1 par. ch. 1 ticket fare 7.2292 embarked Cherbourg	ticket class 3 gender male age 17 sibl. sp. 0 par. ch. 1 ticket fare 14.4583 embarked Cherbourg	ticket class 3 gender male age 14.5 sibl. sp. 1 par. ch. 0 ticket fare 14.4542 embarked Cherbourg	ticket class 3 gender female age 17 sibl. sp. 1 par. ch. 1 ticket fare 7.2292 embarked Cherbourg

At the bottom of each table are buttons for '데이터 추가' (Add Data) and page numbers: 288 and 424 respectively.

To the right, a Scratch project window titled 'Titanic survivors' is open. It shows a script for 'recognise numbers ticket class' with various conditions for gender, age, sibl. sp., par. ch., ticket fare, and embarked. The script also includes 'sunweld' and 'did_not_survive' blocks. A cat sprite is visible on the stage.

Automated ML의 명암



0



null



undefined



Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II.

<http://www.jstatsoft.org/>

Tidy Data

Hadley Wickham
RStudio

Abstract

A huge amount of effort is spent cleaning data to get it ready for analysis, but there has been little research on how to make data cleaning as easy and effective as possible. This paper tackles a small, but important, component of data cleaning: data tidying. Tidy datasets are easy to manipulate, model and visualise, and have a specific structure: each variable is a column, each observation is a row, and each type of observational unit is a table. This framework makes it easy to tidy messy datasets because only a small set of tools are needed to deal with a wide range of un-tidy datasets. This structure also makes it easier to develop tidy tools for data analysis, tools that both input and output tidy datasets. The advantages of a consistent data structure and matching tools are demonstrated with a case study free from mundane data manipulation chores.

<https://vita.had.co.nz/papers/tidy-data.pdf>

추적 / 이력관리

공유

배포

이 부분을 Automated ML에게 맡길 수 있을까요?

비즈니스 문제 인식 및 정의

데이터 수집 및 정제

논리적인 해석 및 배포

실무 담당자와의 의사소통

겸손한 이메일 발신

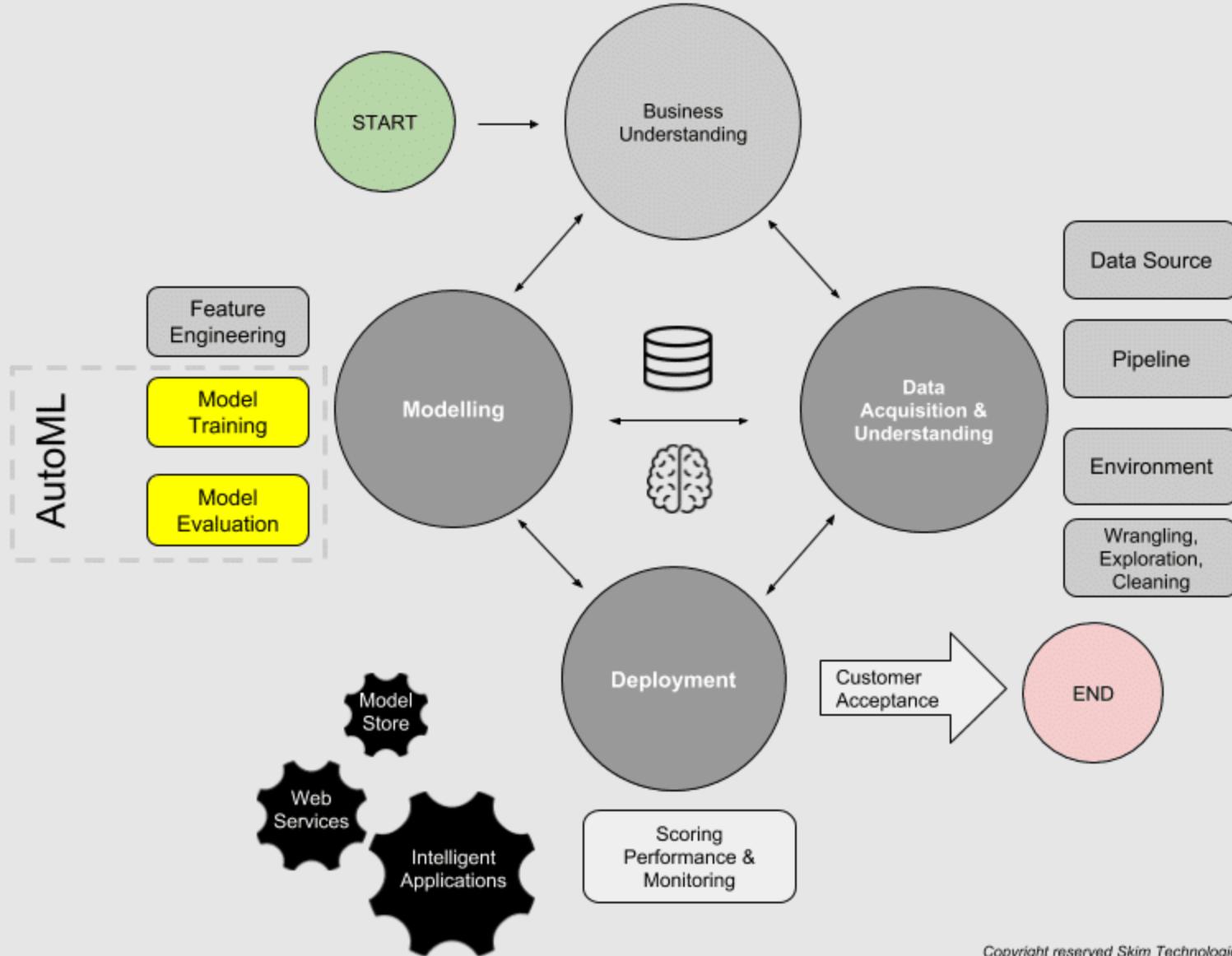
프레젠테이션 기술

이 부분을 Automated ML에게 맡길 수 있을까요?

Automated ML의 3가지 한계

1. 비지도 학습 및 강화학습에 적용
2. 복잡한 데이터 형태
3. 도메인 지식이 필요한 분야

AutoML Applications Within TDSP Workflow



Copyright reserved Skim Technologies

<https://www.skimtechnologies.com/3-reasons-why-automl-wont-replace-data-scientists-yet/>

저는 지금까지 머신러닝
이야기를 좀 자제했습니다

다음 이야기를 함
보겠습니다



Data Science Central

SIGN UP HERE
TO RECEIVE OUR
NEWSLETTER!



[Subscribe to DSC Newsletter](#)

[All Blog Posts](#) [My Blog](#)

+ Add



In Japan, "Artificial Intelligence" comes to be a super star while "Data Scientist" is fading away

Posted by Takashi J. OZAKI on January 13, 2017 at 6:30am [View Blog](#)

I published a post about the current status of "Data Scientist" in Japan, as a periodic follow-up analysis since two years ago. Its trend still remains, but it's beyond my anticipation at that time.

データサイエンティスト
Search term
"Data Scientist"

人工知能
Search term
"Artificial Intelligence"

+ Add comparison

Worldwide ▾

Past 5 years ▾

All categories ▾

Web Search ▾

Welcome to
AnalyticBridge

[Sign Up](#)
[or Sign In](#)

Or sign in with:



<https://www.analyticbridge.datasciencecentral.com/profiles/blogs/in-japan-now-artificial-intelligence-comes-to-be-a-super-star>

데이터 과학자
검색어

인공지능
검색어

+ 비교 추가

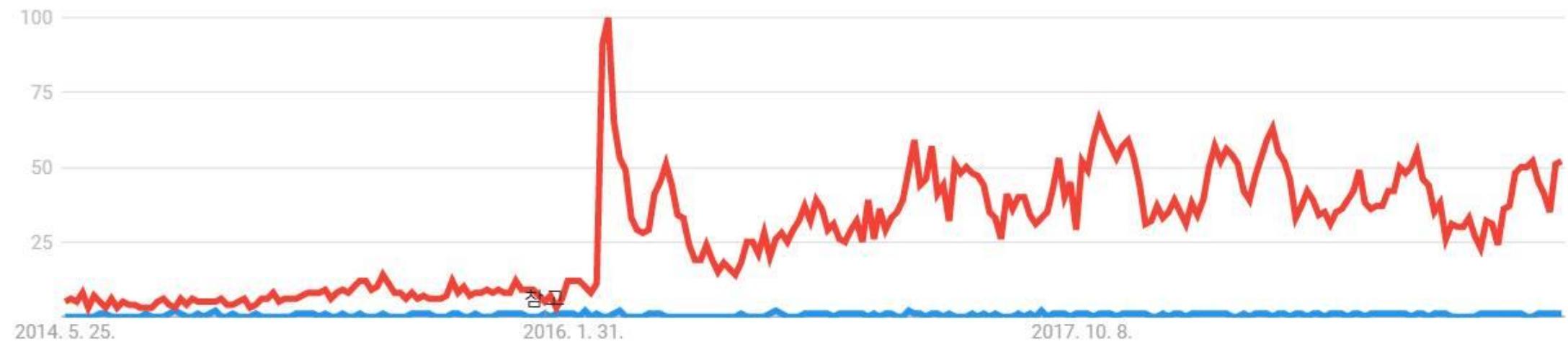
대한민국 ▾

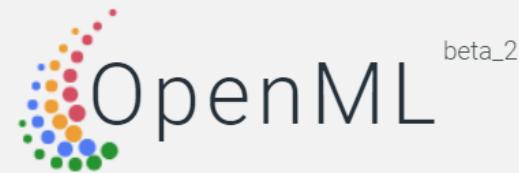
지난 5년 ▾

모든 카테고리 ▾

웹 검색 ▾

시간 흐름에 따른 관심도 변화 ?





Machine learning, better, together

20642
data sets

Find or add **data** to analyse

89904
tasks

Download or create scientific
tasks

10522
flows

Find or add data analysis **flows**

9817856
runs

Upload and explore all **results**
online.



HACKATHON

<https://www.openml.org/home>

데이터가...

뭐가

어디에

어떻게

데이터 주인정신을 가집시다!

1. 문제 의식 – 데이터가 보이기 시작함
2. Garbage In, Garbage Out – 정확도와 신뢰성의 문제
3. 수집량 – 비용 및 공간의 문제
4. 수집주기 – 정확한 문제 분석을 위한 문제
5. 민감한 데이터 보안 및 백업 정책

Autoamted ML의 장점

- 일단 컴퓨터 지식, 도메인 지식, 수학 지식이 없어도 시작해 볼 수 있습니다.
- 좀더 객관적으로 데이터를 분석하고 모델링하려고 노력을 할 것입니다.
- IT, 도메인, 수학 전문가 고용 및 인프라 투자 비용을 줄일 수 있습니다.
- 실제 서비스 구현을 위한 시간을 줄일 수 있습니다.

Automated ML의 단점

- 참고는 하겠지만 일단 결과에 대해서 의심을 해봅시다! 아직 개발중이기도 하고 만능은 아닙니다.
- 아무리 좋은 Automated ML이라도 계산을 위한 자원이 필요합니다. 이 비용도 생각하셔야지요.... 그리고 시간도...

감사합니다!

김영하

youngha@dplus.com

010-9369-1310