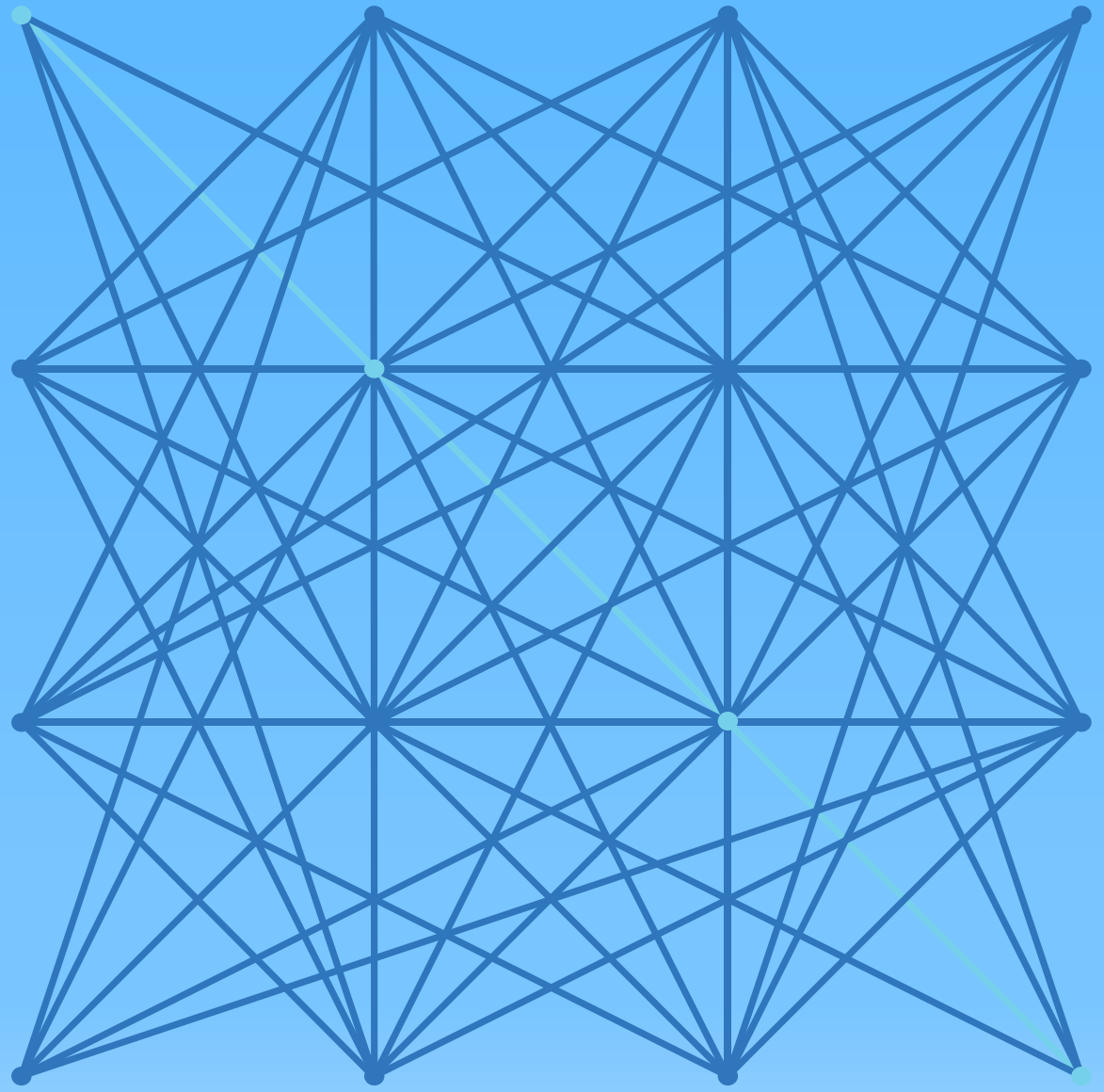
2nd 함께하는 딥러닝 컨퍼런스

Automated ML 동향 알아보기

2019.7.4 (목) 15:00 사이언스홀

디플러스 김영하

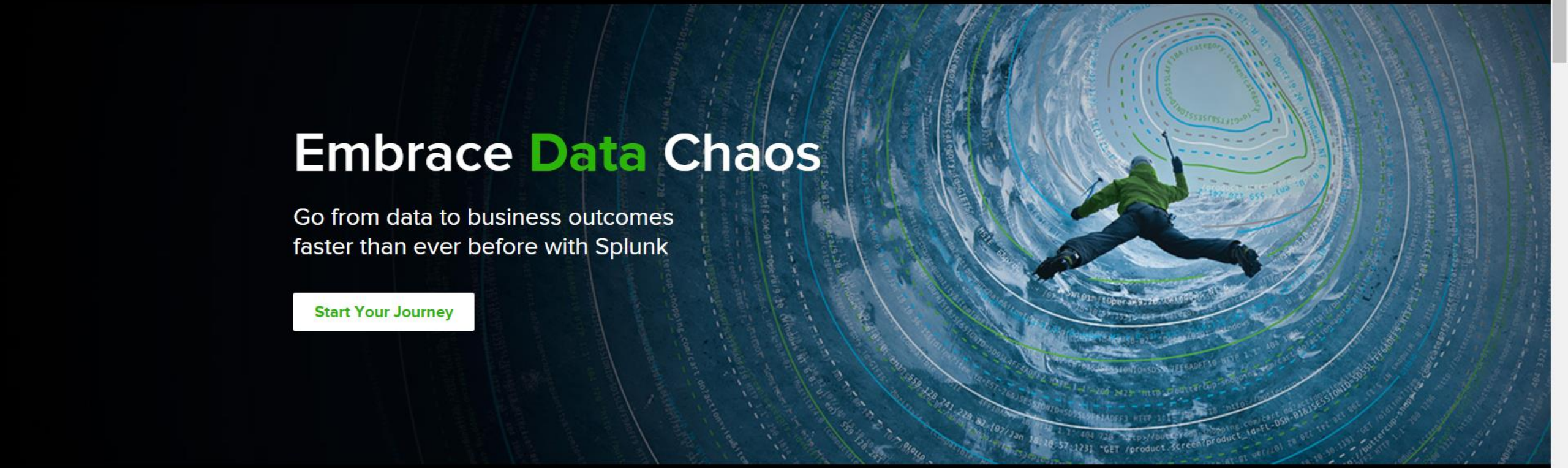


6년동안
직업군인
하다가...

Training Support Pricing Splunk Sites

splunk> PRODUCTS SOLUTIONS CUSTOMERS PARTNERS RESOURCES ABOUT US

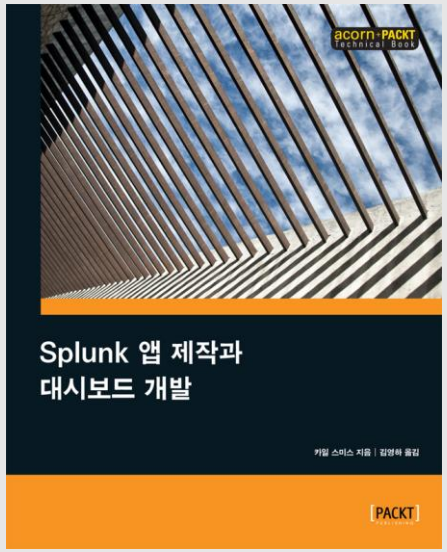
Free Splunk

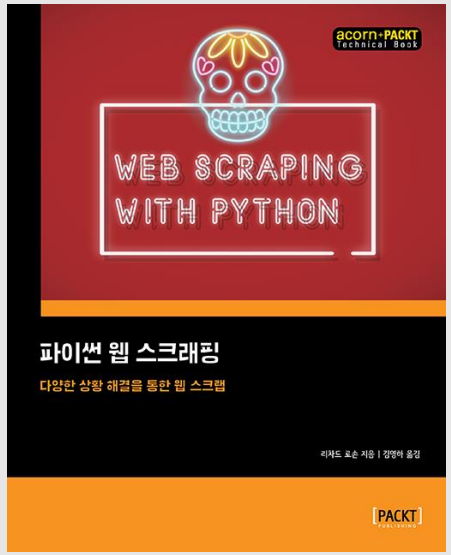


.conf19 Call For Papers is now open! Share Your Story

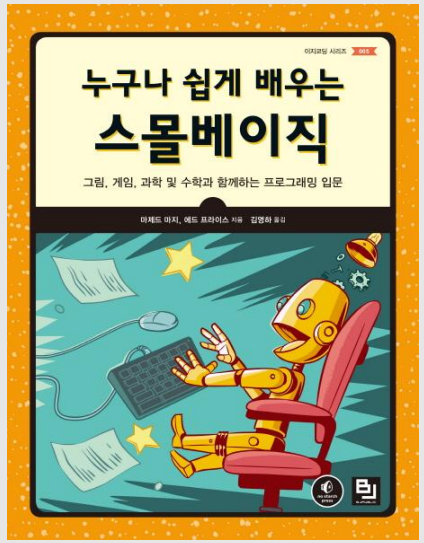
Turn Data Into Action

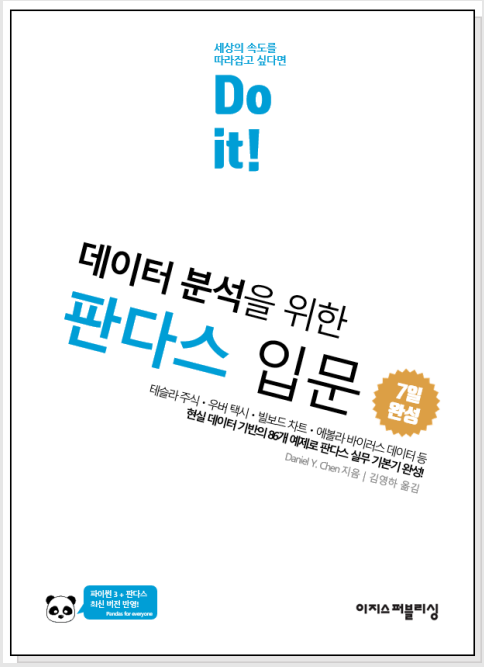
어쩌다 운좋게 빅데이터를 만지게 되어


acorn+PACKT
Technical Book
Splunk 앱 제작과
대시보드 개발
PACKT

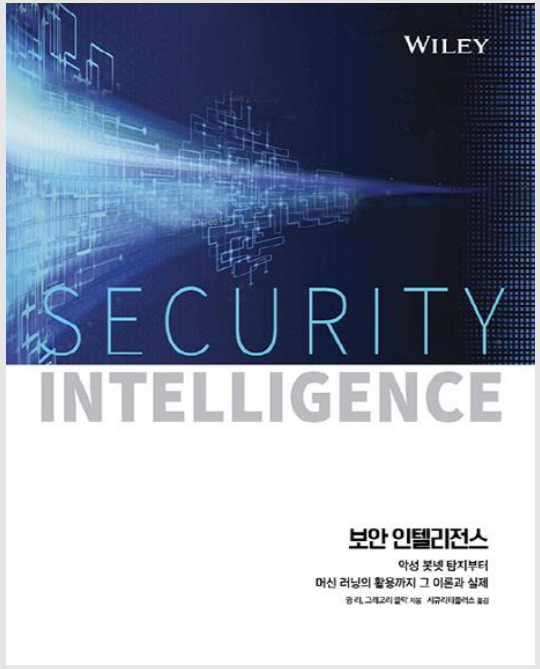

acorn+PACKT
Technical Book
WEB SCRAPING
WITH PYTHON
파이썬 웹 스크래핑
다양한 상황 해결을 통한 웹 스크랩
PACKT


O'REILLY
뷰티풀
자바스크립트
15명의 리딩 프로그래머와 함께하는 자바스크립트

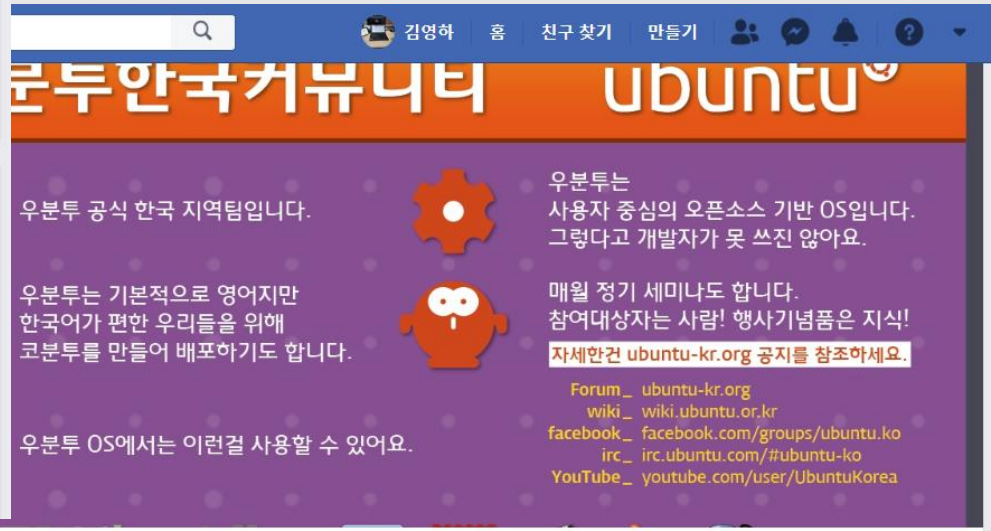

누구나 쉽게 배우는
스몰베이직
그림, 게임, 과학 및 수학과 함께하는 프로그래밍 입문


세상의 속도를
따라잡고 싶다면
Do
it!
데이터 분석을 위한
판다스 입문
7일
완성
이지스퍼블리싱


MICRO
SOFTWARE
LEARN & TRY & SHARE
2019 VOL.395
Data Adventure


WILEY
SECURITY
INTELLIGENCE
보안 인텔리전스
악성 봇넷 탐지부터
머신 러닝의 활용까지 그 이론과 실제


Korea ML.NET User Group
김영하
홈
친구 찾기
만들기
Korea ML.NET User Group
공개 그룹
정보
토론
채팅 1
공지
멤버
이벤트
사진
그룹 인사이트
그룹 관리
ML.NET
KOREA
가입함
알림
공유하기
더 보기
이 그룹 검색
바로가기
따빠마드레 Tapa M... 20+
Korea ML.NET User Gr...
TensorFlow KR 20+
OpenLab.AI
책쓰는 프로그래머 ... 1
게시물 작성
사진/동영상 추가
라이브 방송
더 보기
글쓰기...
사진/동영상
함께 시청하기
친구 태그하기
멤버 초대
초대 퍼기기
이름이나 이메일 주소를 입력하세요...
멤버
멤버 124명
이번 주에 새로운 멤버가 17명 있습니다. 새 멤버를 환영하는 게시물을 작성해보세요.
게시물 작성
추천 멤버
숨기기
친구


김영하
홈
친구 찾기
만들기
ubuntu
우분투는
사용자 중심의 오픈소스 기반 OS입니다.
그렇다고 개발자가 못 쓰진 않아요.
우분투 공식 한국 지역팀입니다.
우분투는 기본적으로 영어지만
한국어가 편한 우리들을 위해
코분투를 만들어 배포하기도 합니다.
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
우분투 OS에서는 이런걸 사용할 수 있어요.
가입함
알림
공유하기
더 보기
사진
파일
이 그룹 검색
바로가기
따빠마드레 Tapa M... 20+
Korea ML.NET User Gr...
TensorFlow KR 20+
OpenLab.AI
책쓰는 프로그래머 ... 1
게시물 작성
사진/동영상 추가
라이브 방송
더 보기
글쓰기...
사진/동영상
함께 시청하기
친구 태그하기
관리 페이지
Ubuntu Korea Community
564명이 좋아합니다
문의하기
멤버 초대
이름이나 이메일 주소를 입력하세요...
멤버
멤버 9,435명
공지
모두 보기(6개)

D Plus

데이터 분석이 중요한 이유 D PLUS는 MEMBERS CONTACT

MEMBERS

데이터 분석을 돕는

모두의 희망 재택근무!

정 원 혁

대 표

김 영 하

연 구 원

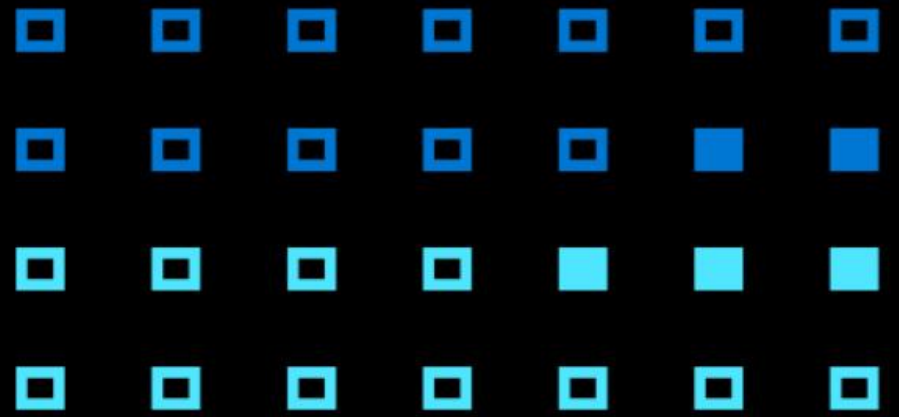
김 정 훈

데이터 엔지니어

Microsoft

Azure

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.

이 교육 과정은 Microsoft 클라우드 플랫폼에 대한 입문 과정으로 IT 전문가와 엔지니어 분들에게 추천합니다

Bokeh

1.0.4 ▾ INSTALLATION USER GUIDE GALLERY TUTORIAL REFERENCE RELEASES DEVELOPER GUIDE

Welcome to Bokeh ¶

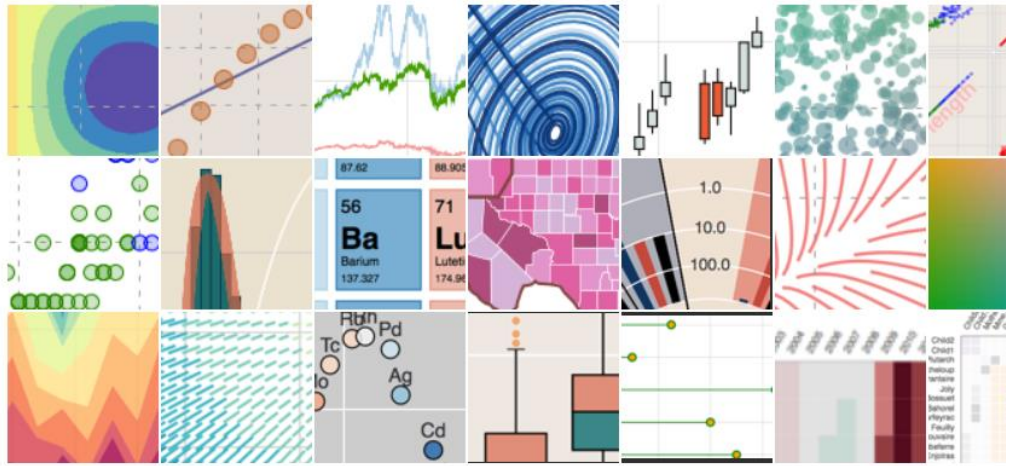
Bokeh is an interactive visualization library that targets modern web browsers for presentation. Its goal is to provide elegant, concise construction of versatile graphics, and to extend this capability with high-performance interactivity over very large or streaming datasets. Bokeh can help anyone who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

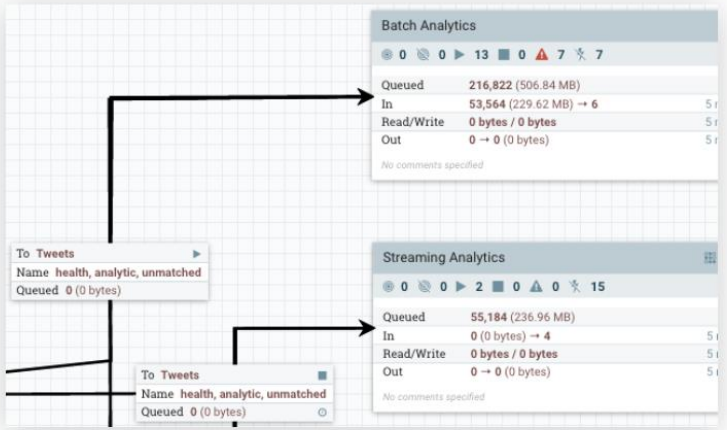
If you are interested in contributing to Bokeh, or extending the library, see the Developer Guide.



Project Documentation Downloads Community Development ASF Links Subprojects

APACHE nifi

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

3개월째
무시무시한 제조데이터를 만지고 있습니다!

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."

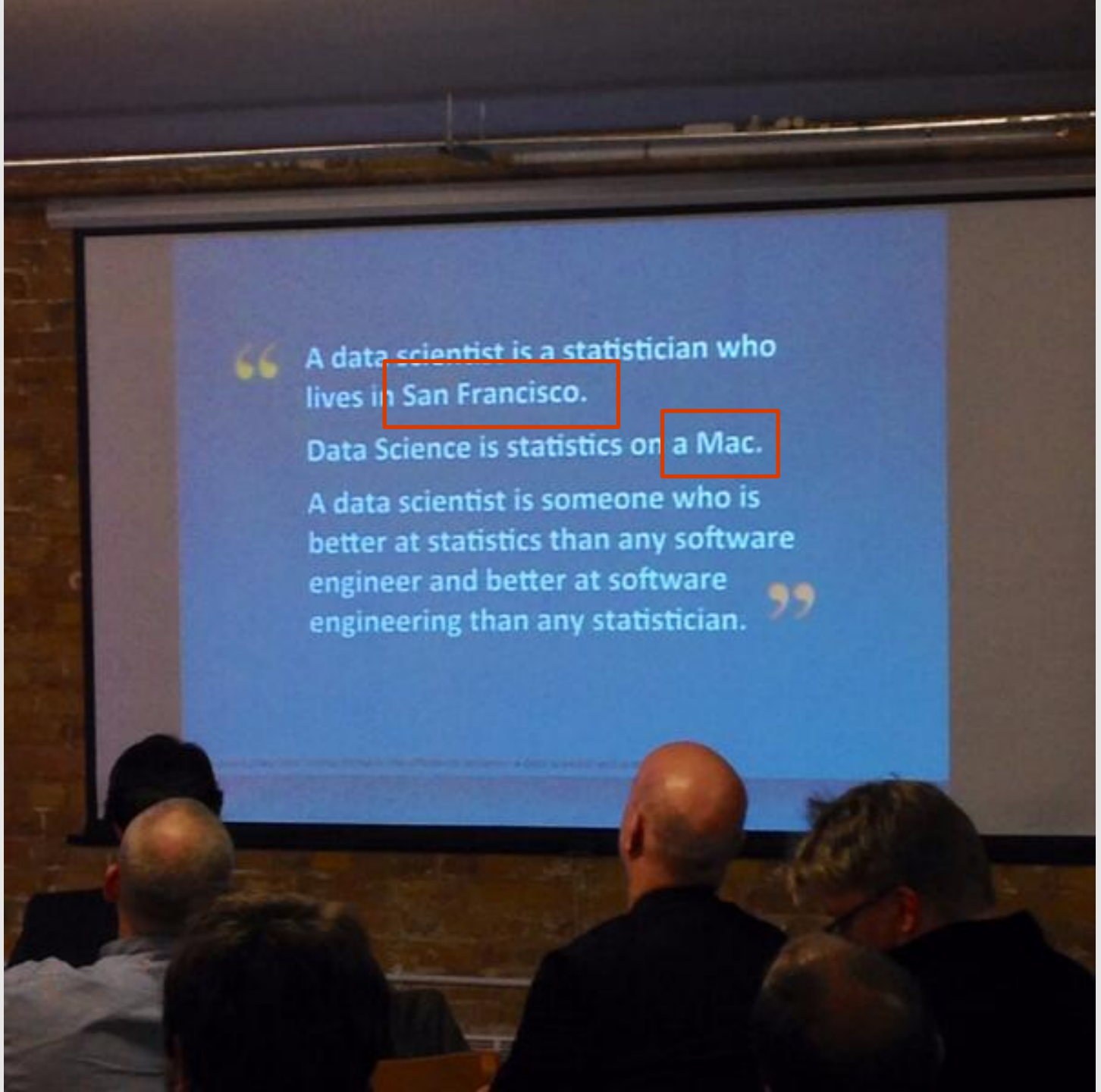
-- Josh Wills, Director of Data Science at Cloudera
(formerly Google)

긍정적인 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

부정적인 마음`

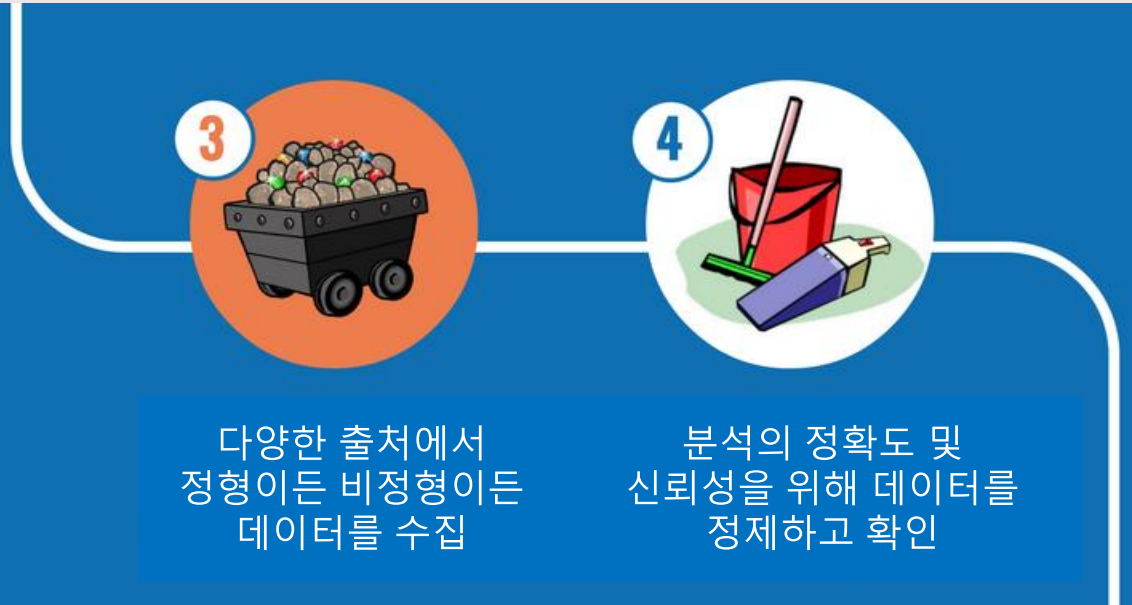


자! 이 정의에 의하면

한국에는 진정한
데이터과학자가
없습니다!

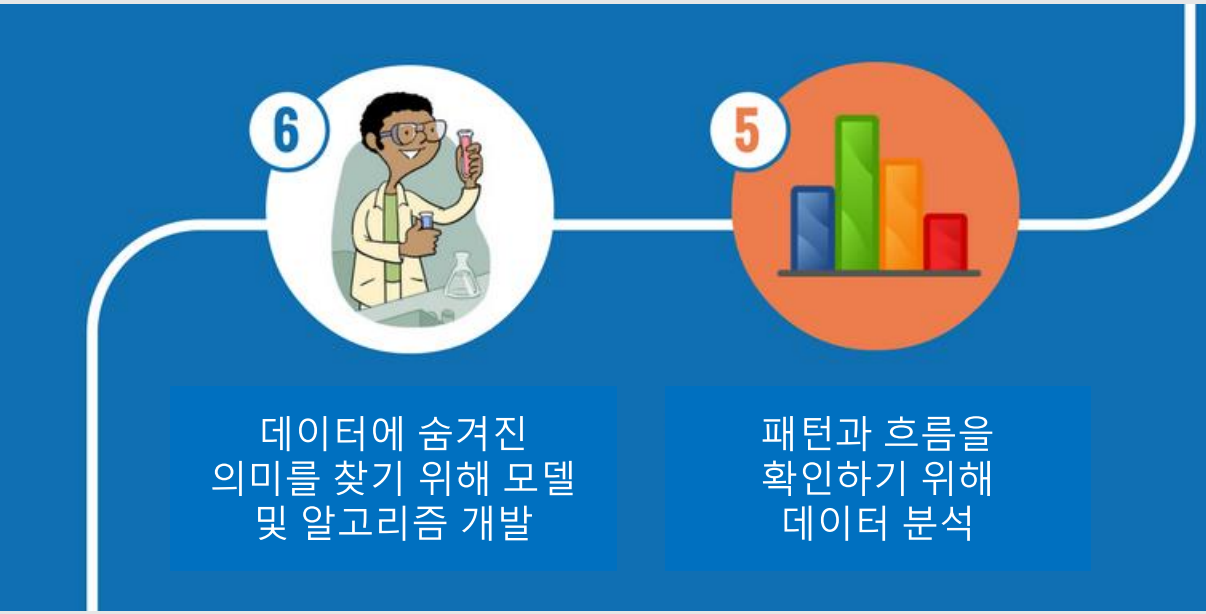
용기를 가지세요!

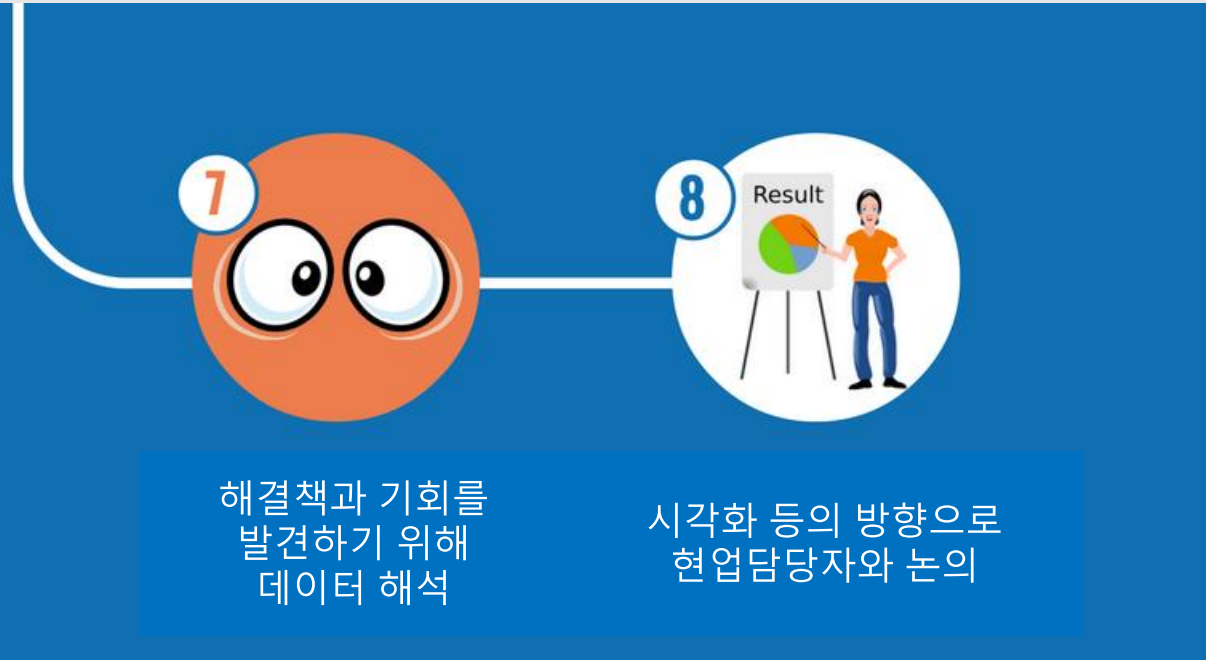
데이터 과학자의 일상 1



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

데이터 과학자의 일상 2





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

데이터 과학자들은

데이터를 분석하기에도 **시간이 부족**합니다!

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
Exa	E	10^{18}	= 1 000 000 000 000 000 000
Peta	P	10^{15}	= 1 000 000 000 000 000
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

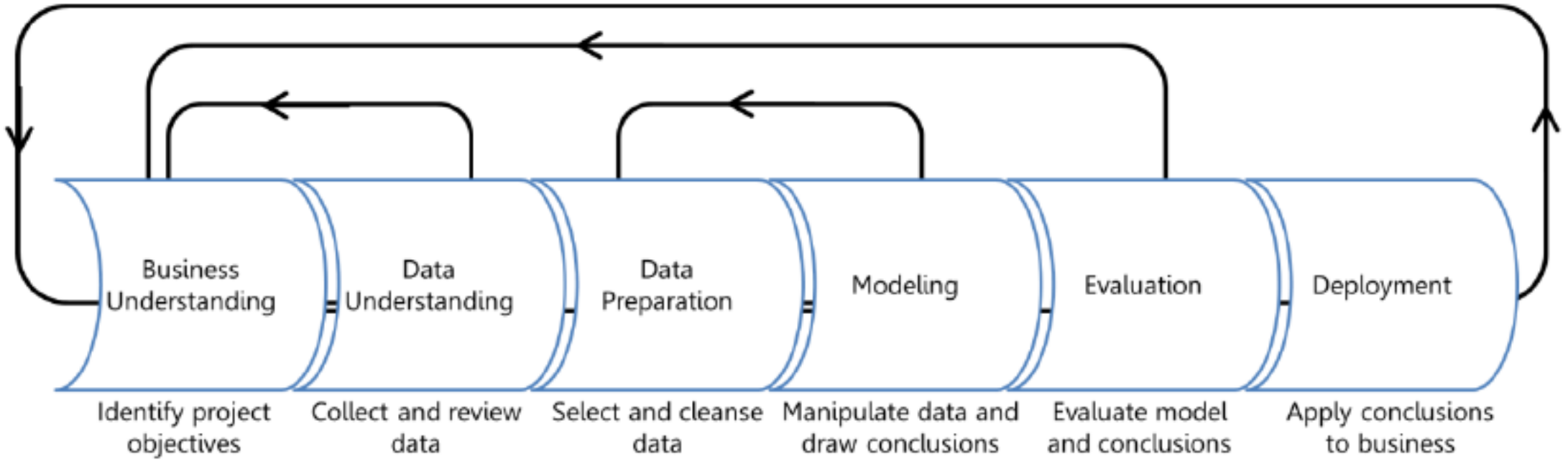
참으로
데이터가
많군요!

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!

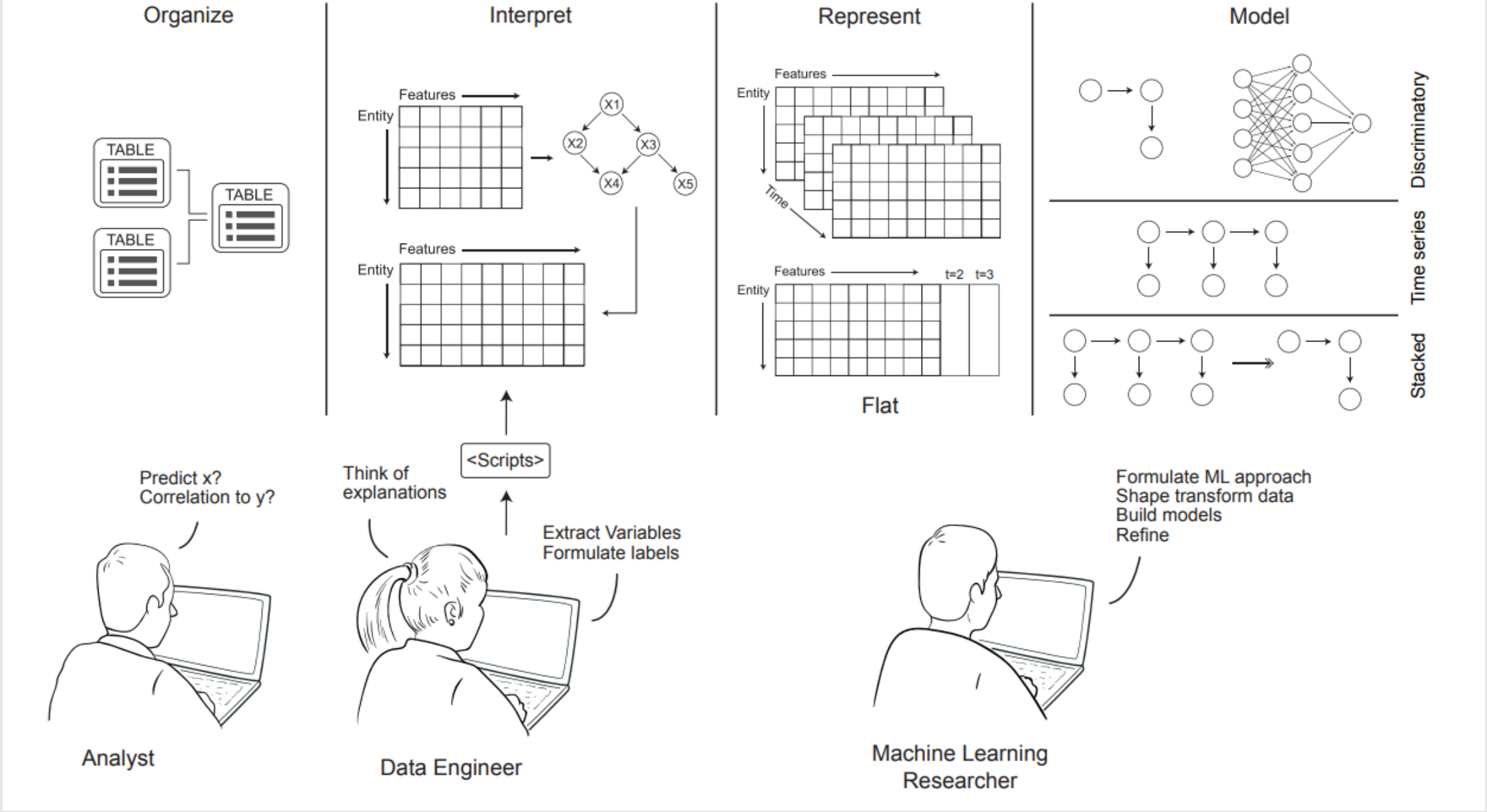
DataOps - https://www.dataopsmanifesto.org/

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



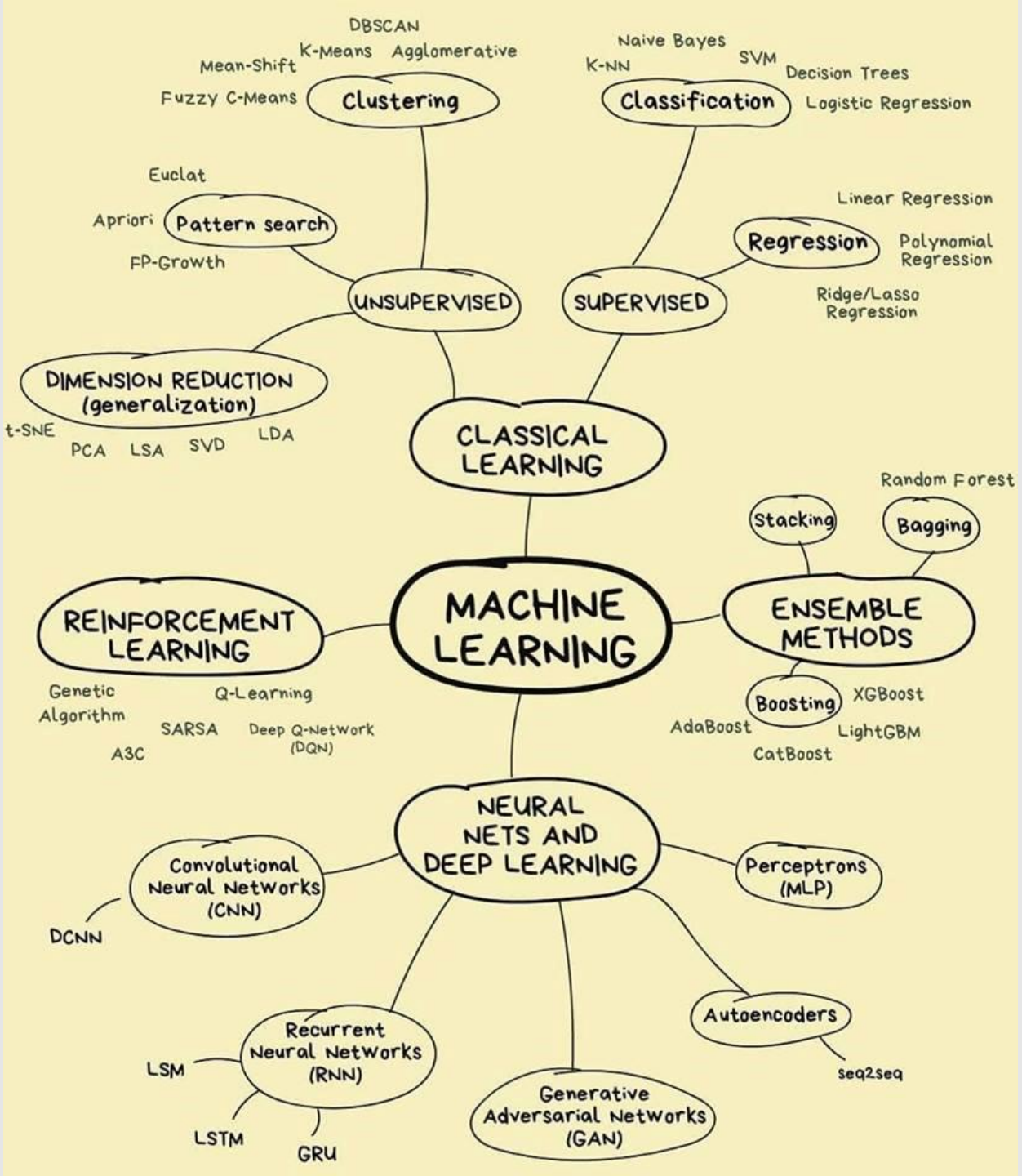
http://groups.csail.mit.edu/EVO-DesignOpt/groupWebSite/uploads/Site/DSAA_DSM_2015.pdf

이쁘게 잘

정리해주셨는데...

왠지 겁나고

내 일이 아닌 것 같습니다



너도 해보고
나도 해보는

데이터 분석의 민주화!

Data Analysis Democracy!

Automated ML

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

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

Automated machine learning

From Wikipedia, the free encyclopedia

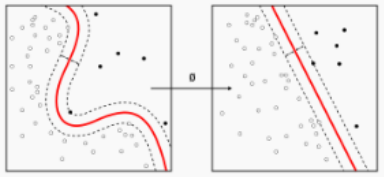
This article has multiple issues. Please help **improve it** or discuss these issues on the **talk page**. *(Learn how and when to remove these template messages)* [hide]

- 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.

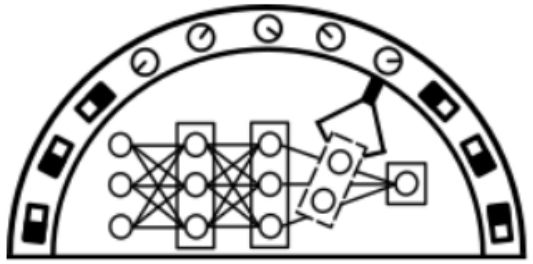
Contents [hide]

Machine learning and data mining



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

https://en.wikipedia.org/wiki/Automated_machine_learning



AutoML.org Freiburg

Search

Follow Us

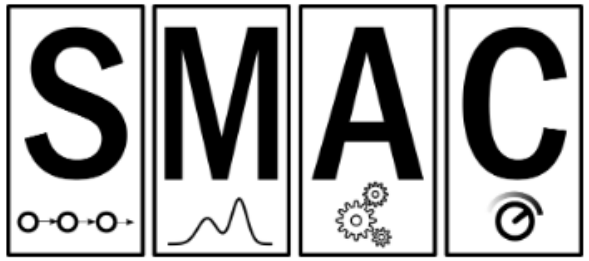
AutoML Freiburg

Home | Blog | AutoML | AAD | Analysis | Book | Events | Team & Partners

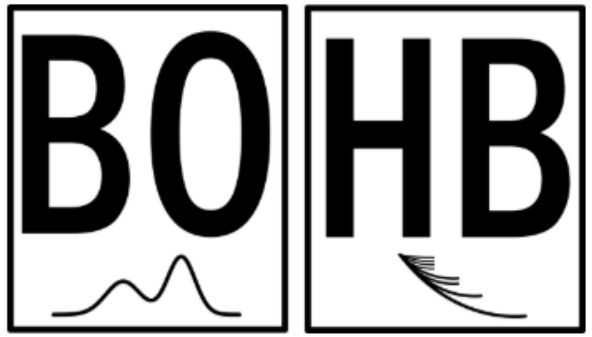
https://www.automl.org/

AutoML ...

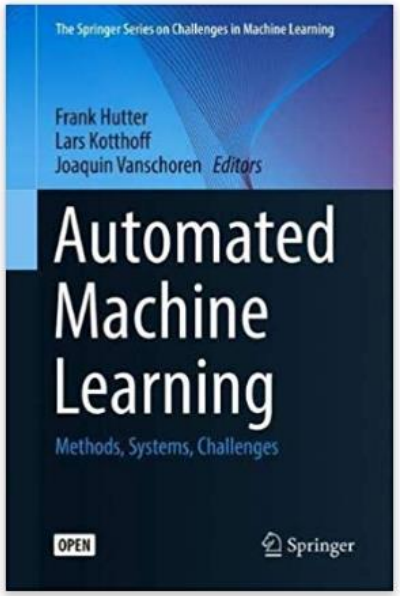
provides methods and processes to make Machine Learning available for non-Machine Learning experts, to improve efficiency of Machine Learning and to accelerate research on Machine Learning. Machine learning (ML) has achieved considerable successes in recent years and an ever-growing number of disciplines 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



amazon
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
Cart
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

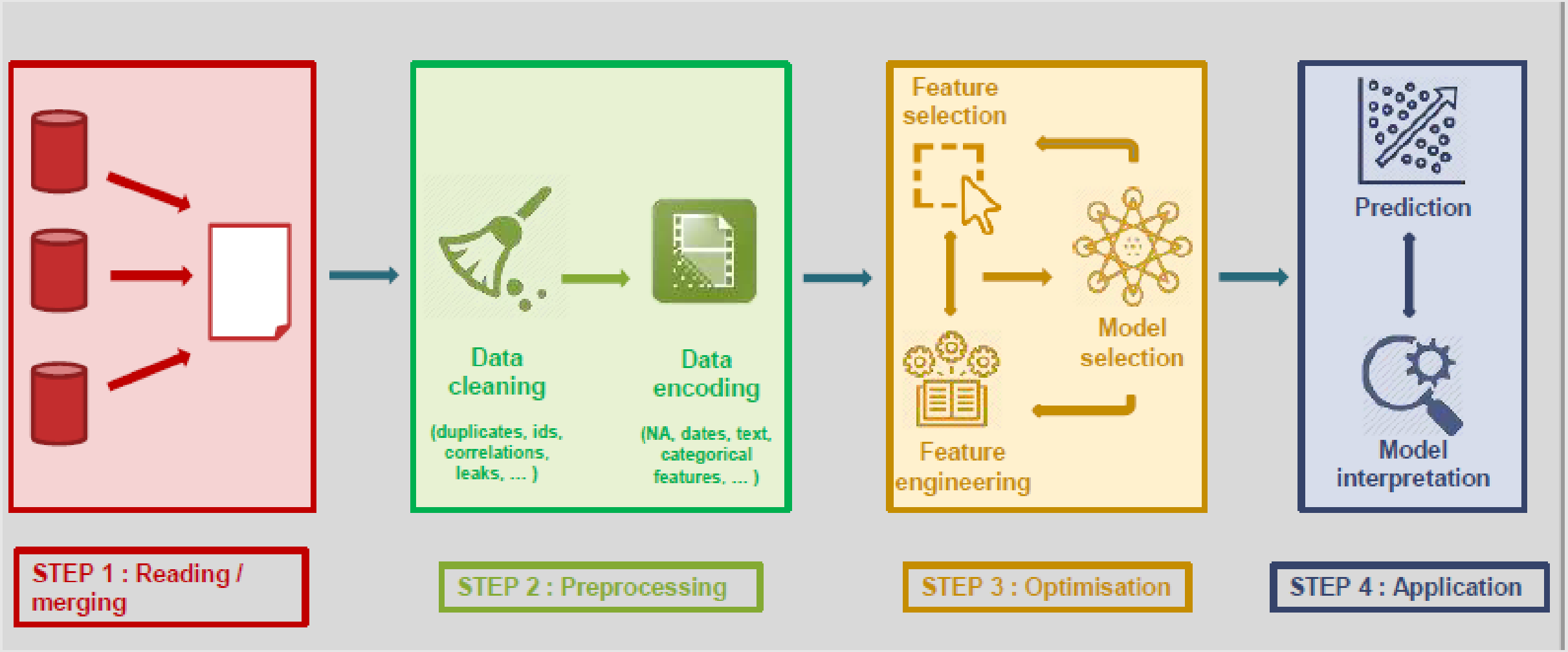
Add to List

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

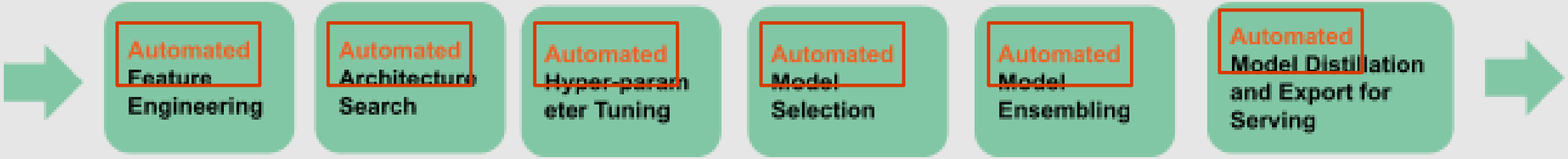
무료 다운로드 가능 : https://link.springer.com/book/10.1007%2F978-3-030-05318-5

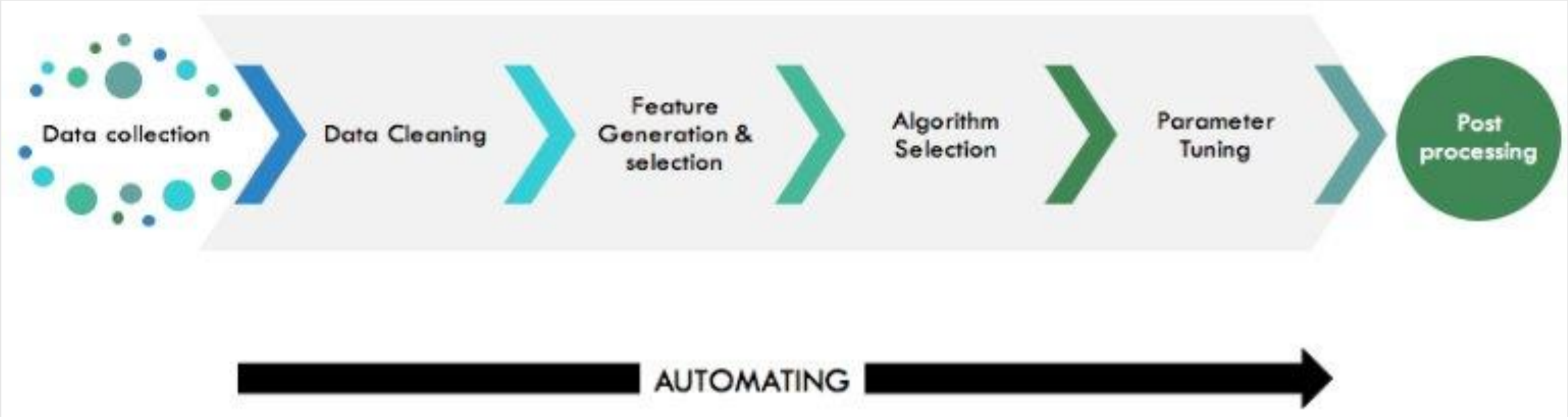
일반적인 ML 파이프라인



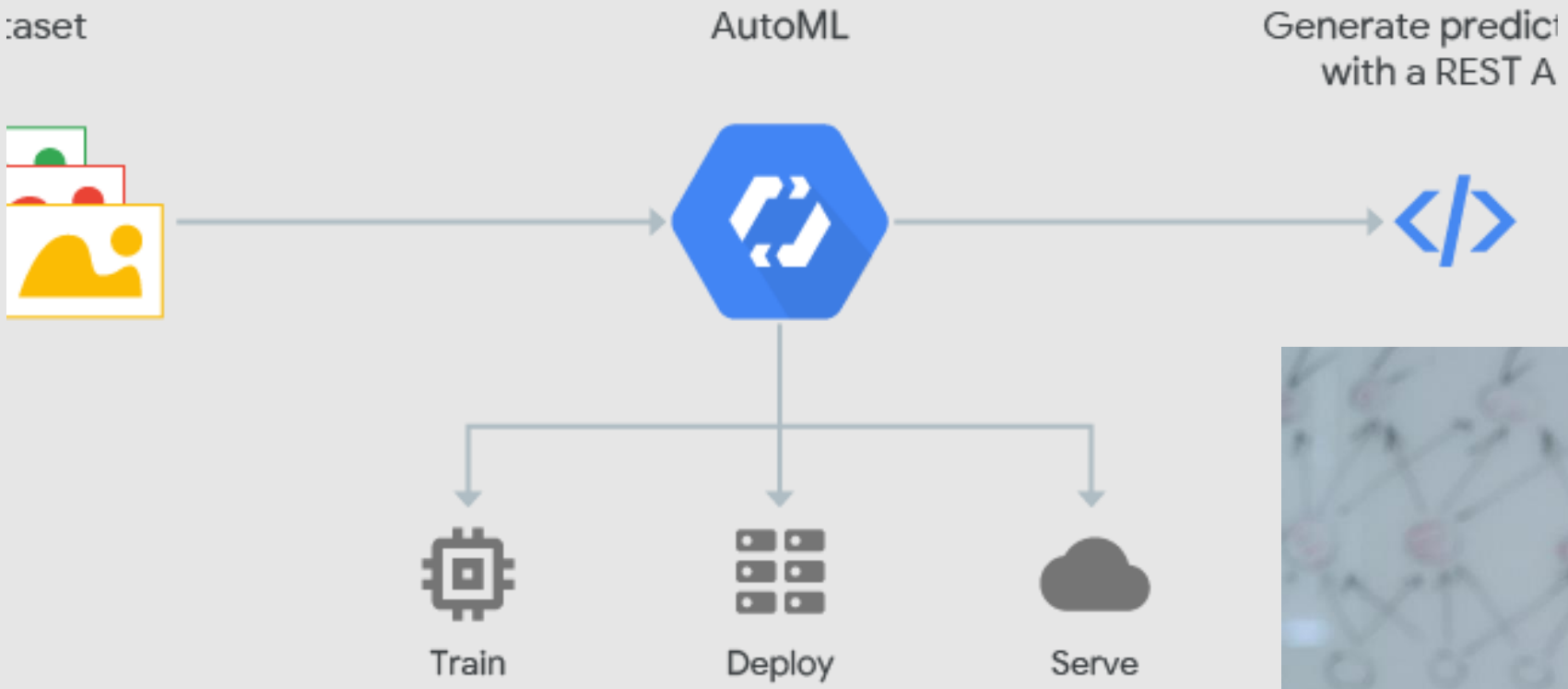
출처 : AUTOMATED MACHINE LEARNING – Axel de Romblay
https://www.slideshare.net/AxeldeRomblay

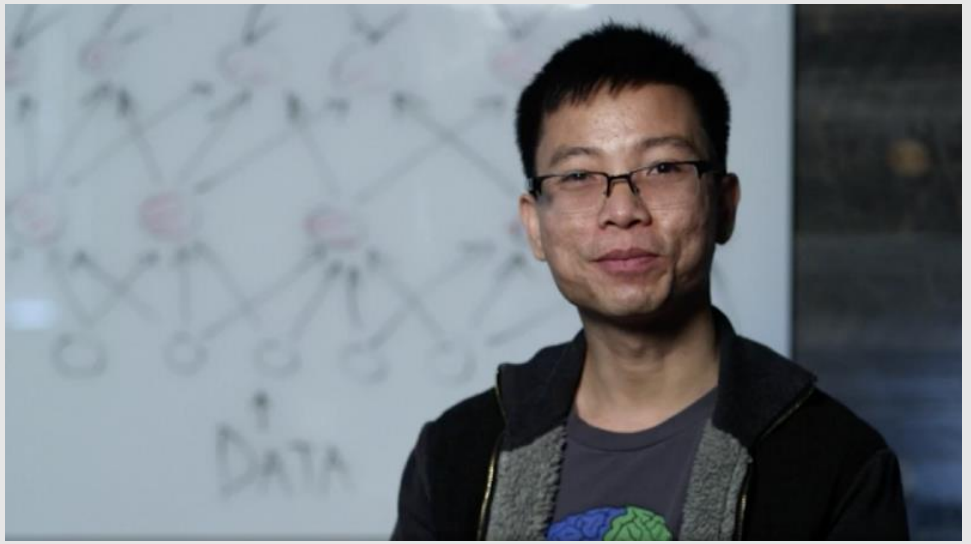
이렇게 하고자하고 실현하고 있습니다!





구글에서는 AutoML로





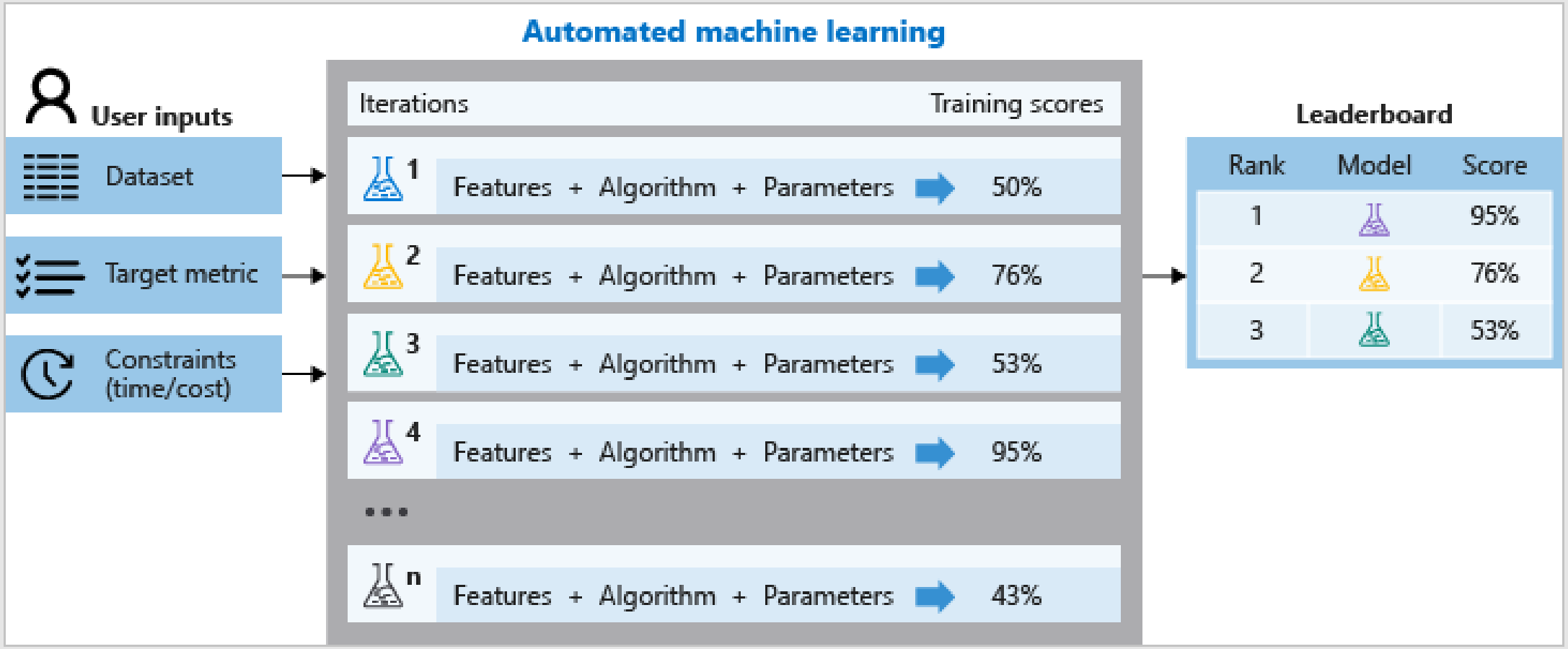
Quoc Le

https://arxiv.org/abs/1112.6209

MS Azure에서는

Azure Machien Learning service로!

오늘 F 트랙에서 13~15시에 전미정님께서 실습해 주셨습니다!



mlflow

DOCS COMMUNITY CODE

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 &

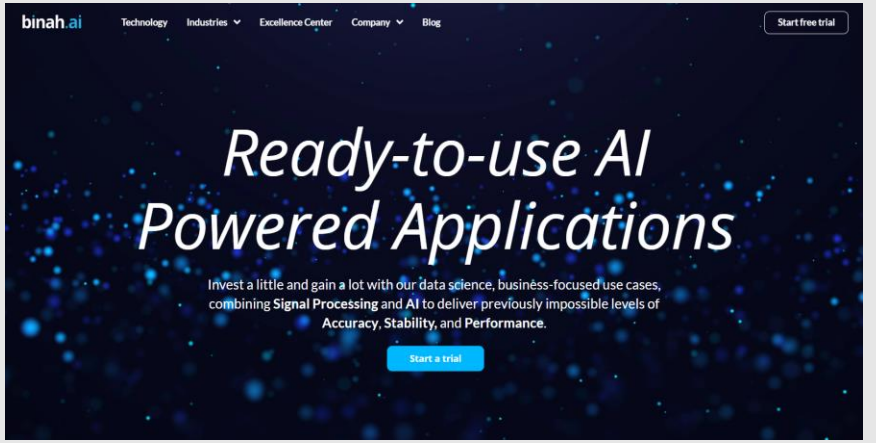
RUNS THE SAME WAY IN ANY CLOUD

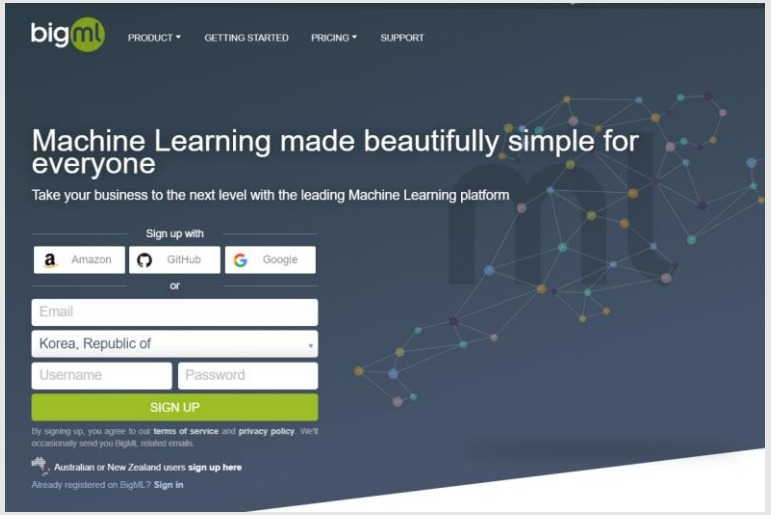
DESIGNED TO SCALE FROM 1 USER TO LARGE ORGS

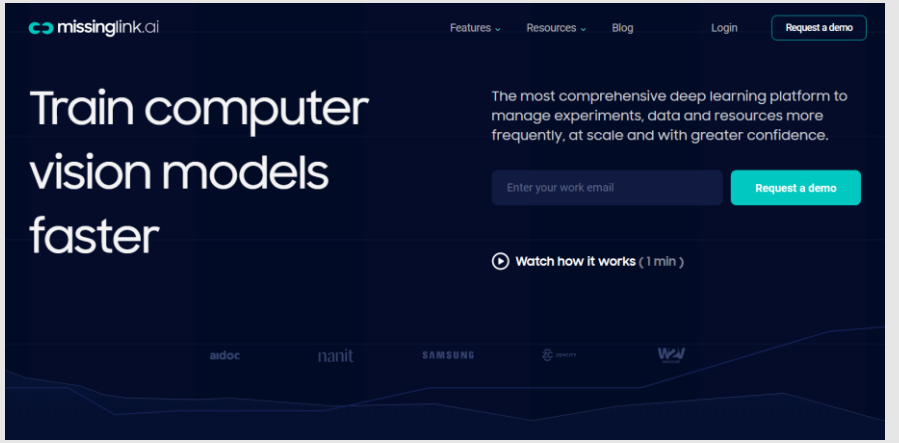
SCALES TO BIG DATA WITH APACHE SPARK™

AutoML 스타트업

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







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/
- H20 : 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))
```

MERCARI

What are you looking for?

Sell now | Sign in

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/

AutoML

AutoML1 (2015-2016) | AutoML2 (2017-2018) | AutoML3 (current)

Navigation

- **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

Home

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.

http://automl.chalearn.org/

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.

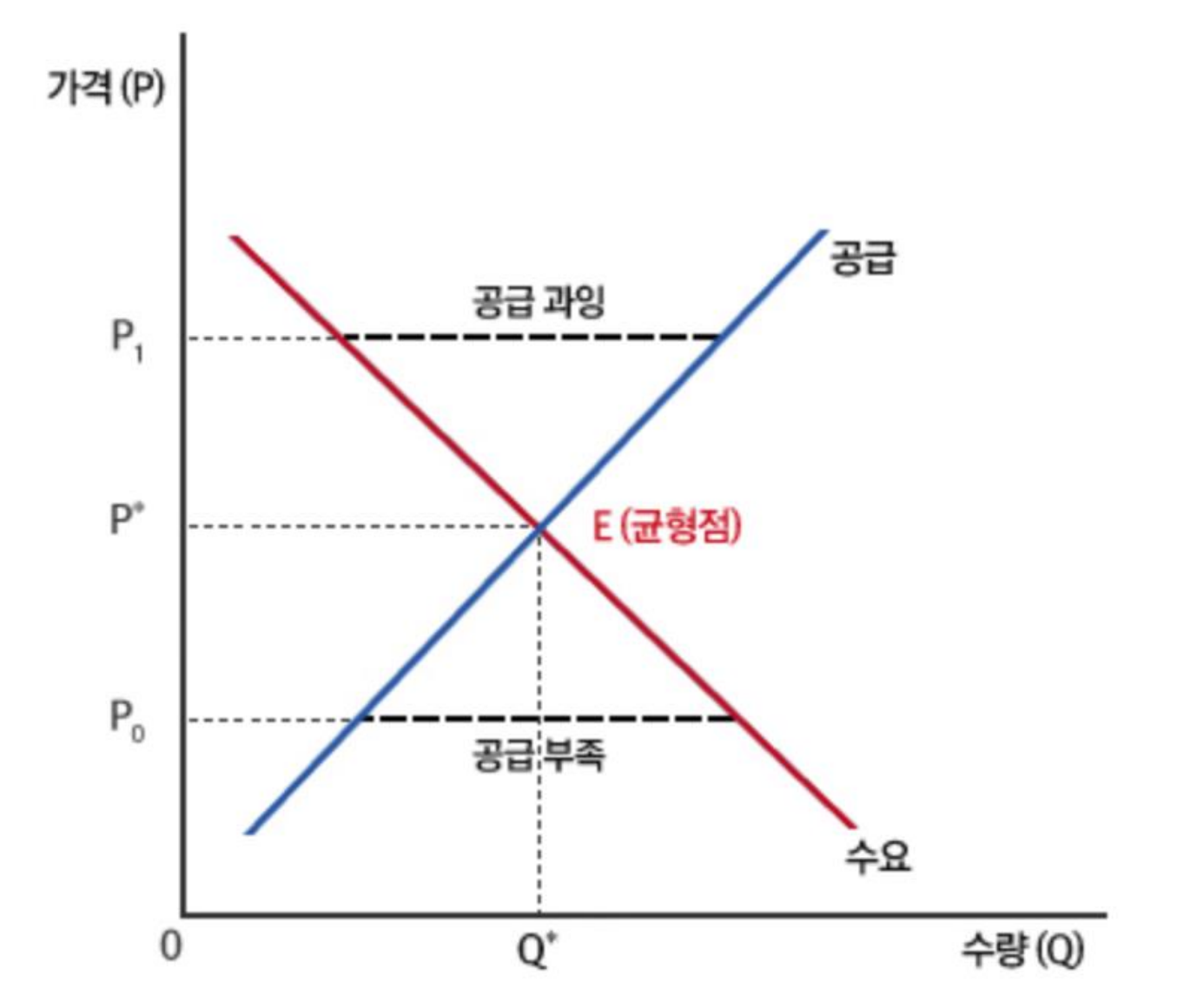
– Fei-Fei Li, 2018 Google press event

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

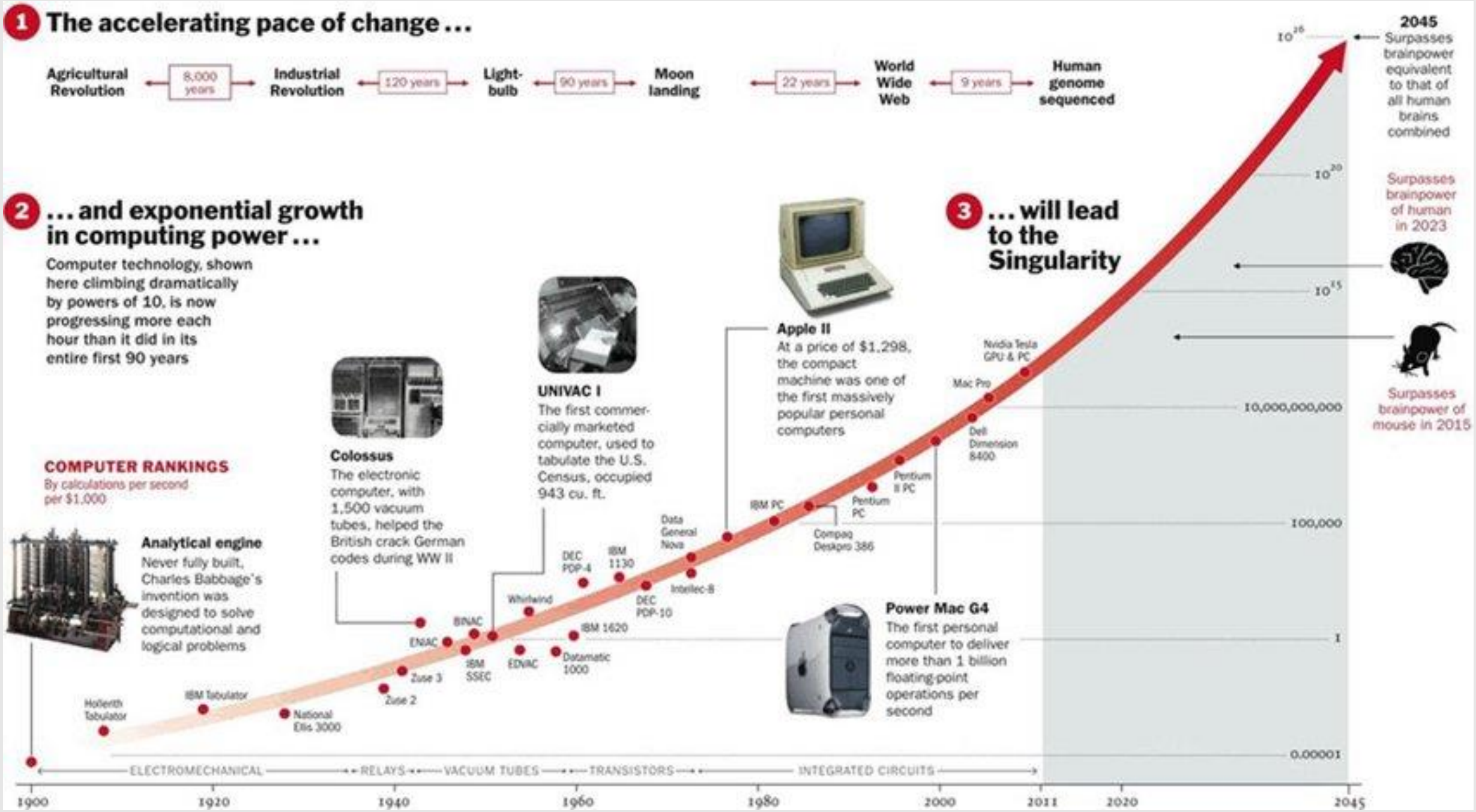
프로그래밍을 좀 하는 놈!

수학 지식이 좀 있는 놈!

도메인을 알고 통계 및 분석을 할 줄 아는 놈!

근데... 있나요? 또 연봉은 얼마인가요?

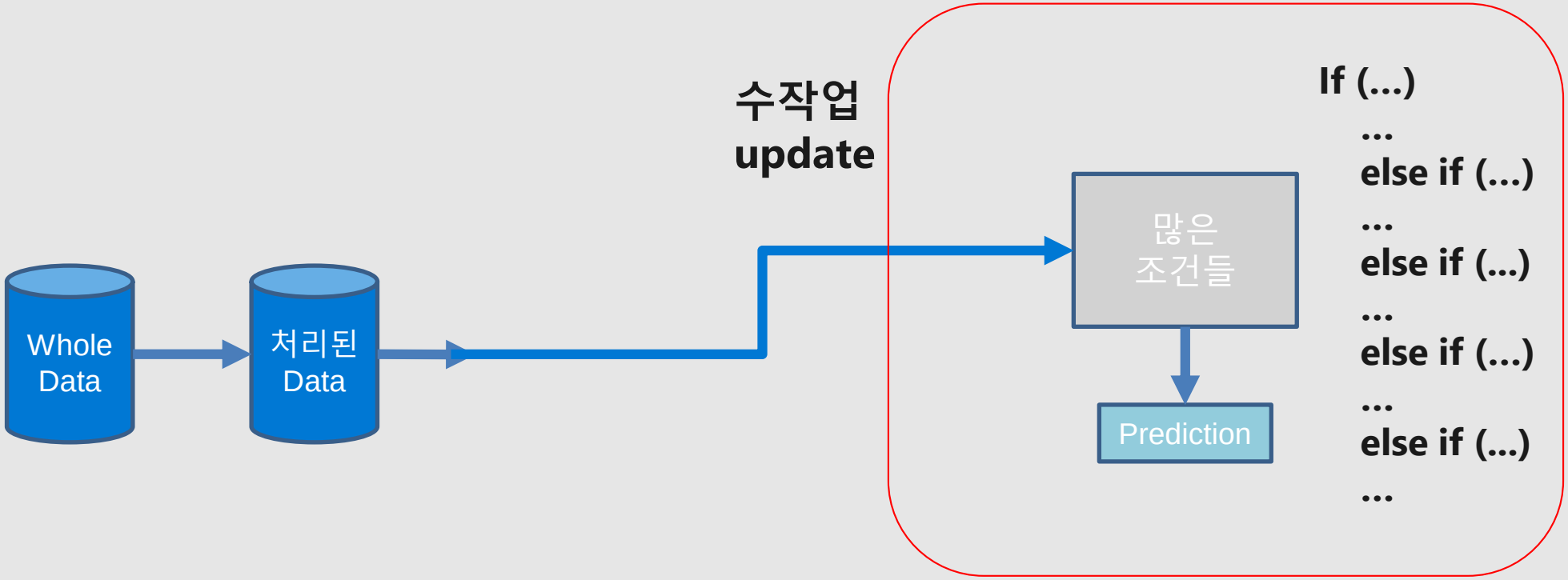
투자의 문제도


1 The accelerating pace of change ...
Agricultural Revolution
8,000 years
Industrial Revolution
120 years
Light-bulb
90 years
Moon landing
22 years
World Wide Web
9 years
Human genome sequenced
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
3 ... will lead to the Singularity
2045
Surpasses brainpower equivalent to that of all human brains combined
Surpasses brainpower of human in 2023
Surpasses brainpower of mouse in 2015
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
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
Hollerith Tabulator
IBM Tabulator
National Ellis 3000
Zuse 2
Zuse 3
ENIAC
BINAC
IBM SSEC
EDVAC
Whirlwind
Datamatic 1000
IBM 1620
DEC PDP-4
IBM 1130
DEC PDP-10
Data General Nova
Intellec-8
IBM PC
Compaq Deskpro 386
Pentium PC
Pentium II PC
Dell Dimension 8400
Mac Pro
Nvidia Tesla GPU & PC
ELECTROMECHANICAL
RELAYS
VACUUM TUBES
TRANSISTORS
INTEGRATED CIRCUITS
1900
1920
1940
1960
1980
2000
2011
2020
2045
0.00001
1
100,000
10,000,000,000

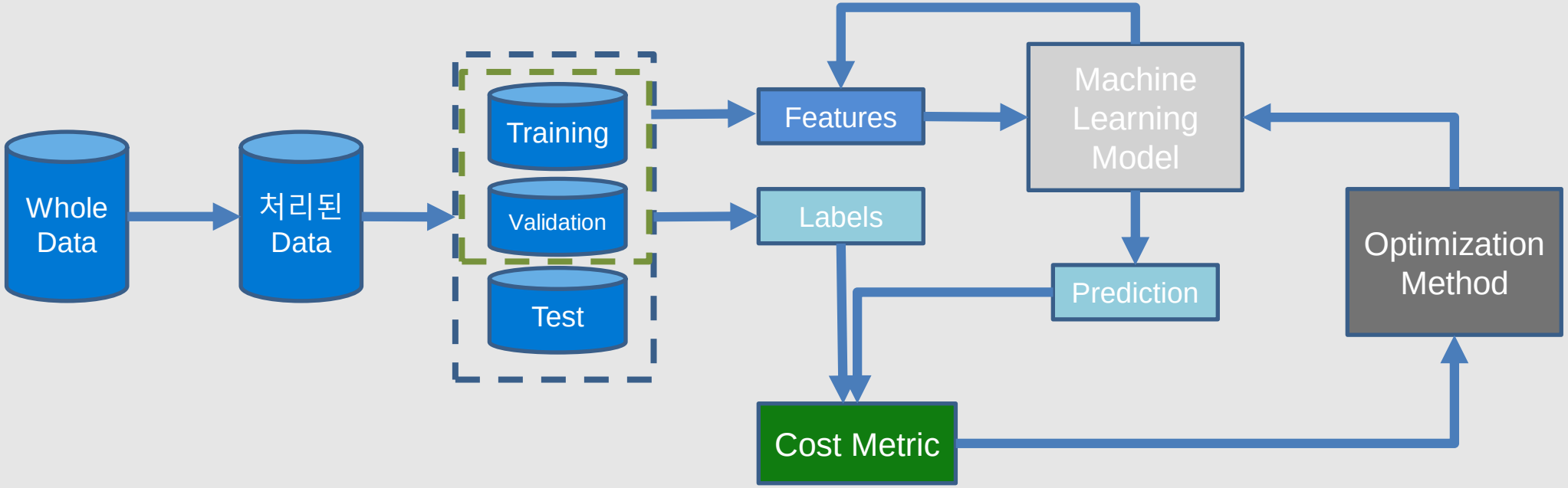

IBM
Data Processing
Division

문제해결 방식도

기존 방식의 데이터 분석


수작업
update
Whole
Data
처리된
Data
많은
조건들
Prediction
If (...)
...
else if (...)
...
else if (...)
...
else if (...)
...
else if (...)
...

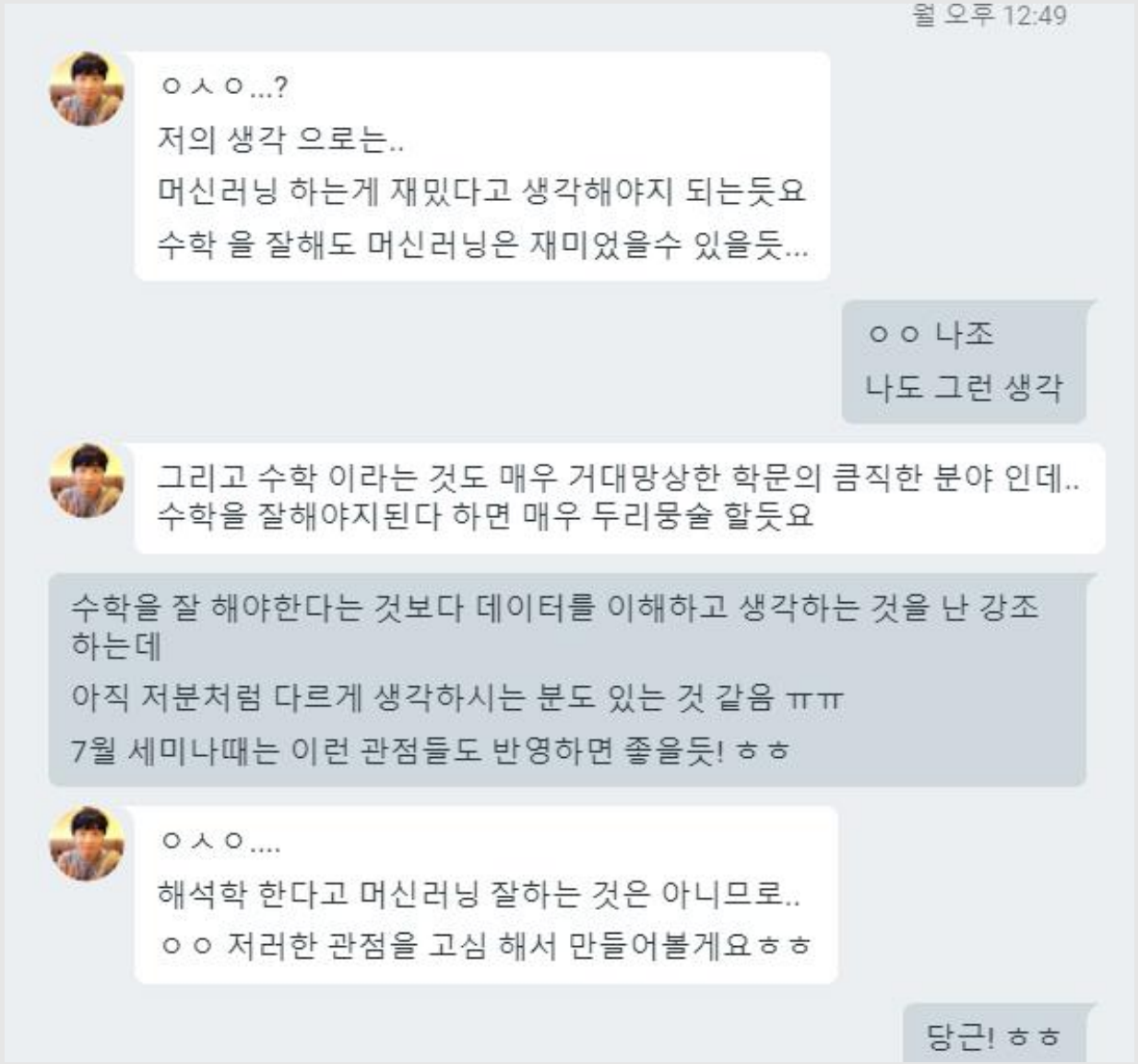
Machine Learning이라면




SBS
그런데 말입니다

넓은 의미의 관점에서

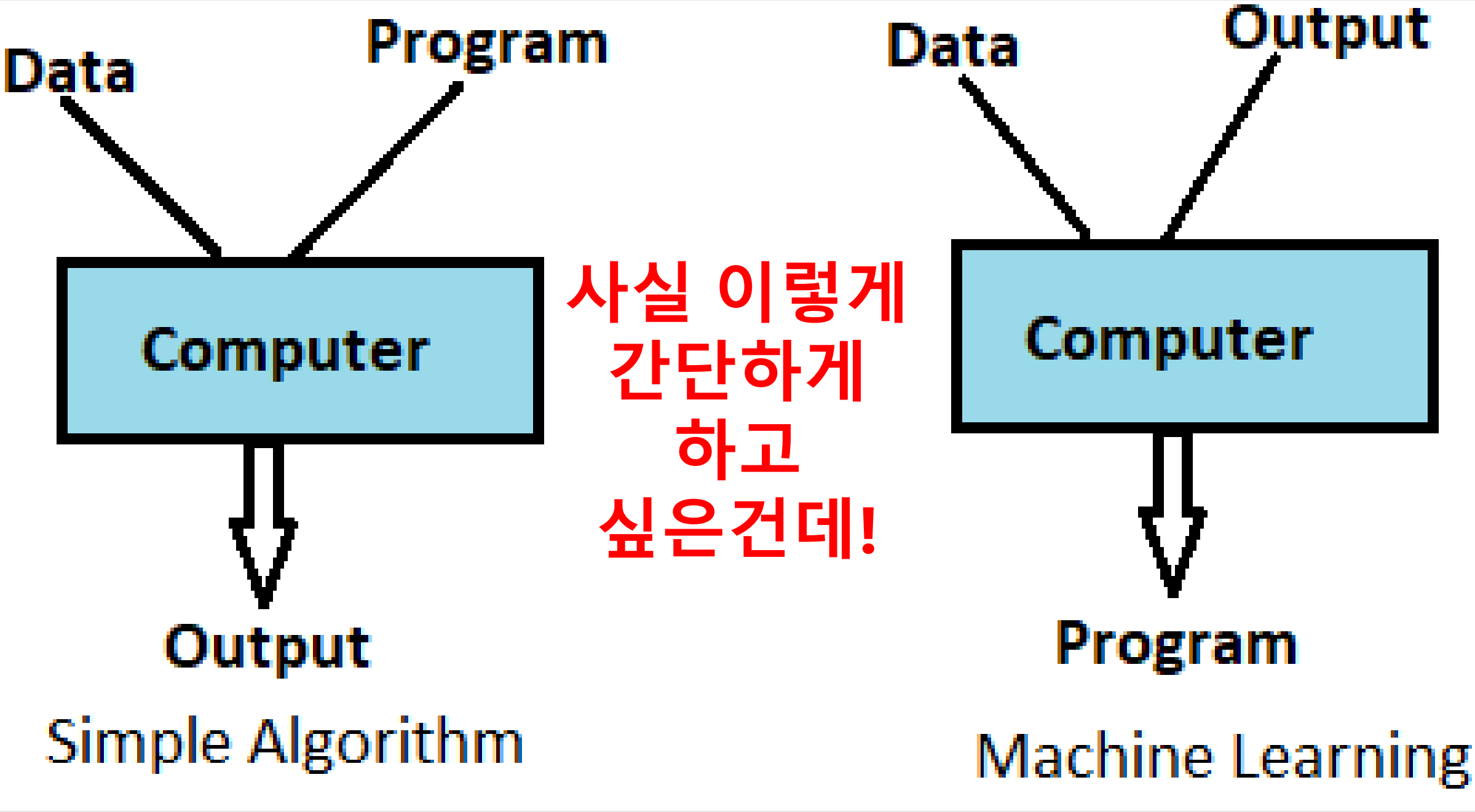
누구나 할 수 없는 머신러닝 – 과연?



오늘 C 트랙에서

발표하시는

홍원의님과의 대화


Data
Program
Computer
Output
Simple Algorithm
사실 이렇게
간단하게
하고
싶은건데!
Data
Output
Computer
Program
Machine Learning


RIDGID

AI의 구조를 쉽게 이해할 수 있는 딥러닝 초(超)입문

엑셀로 배우는 딥러닝

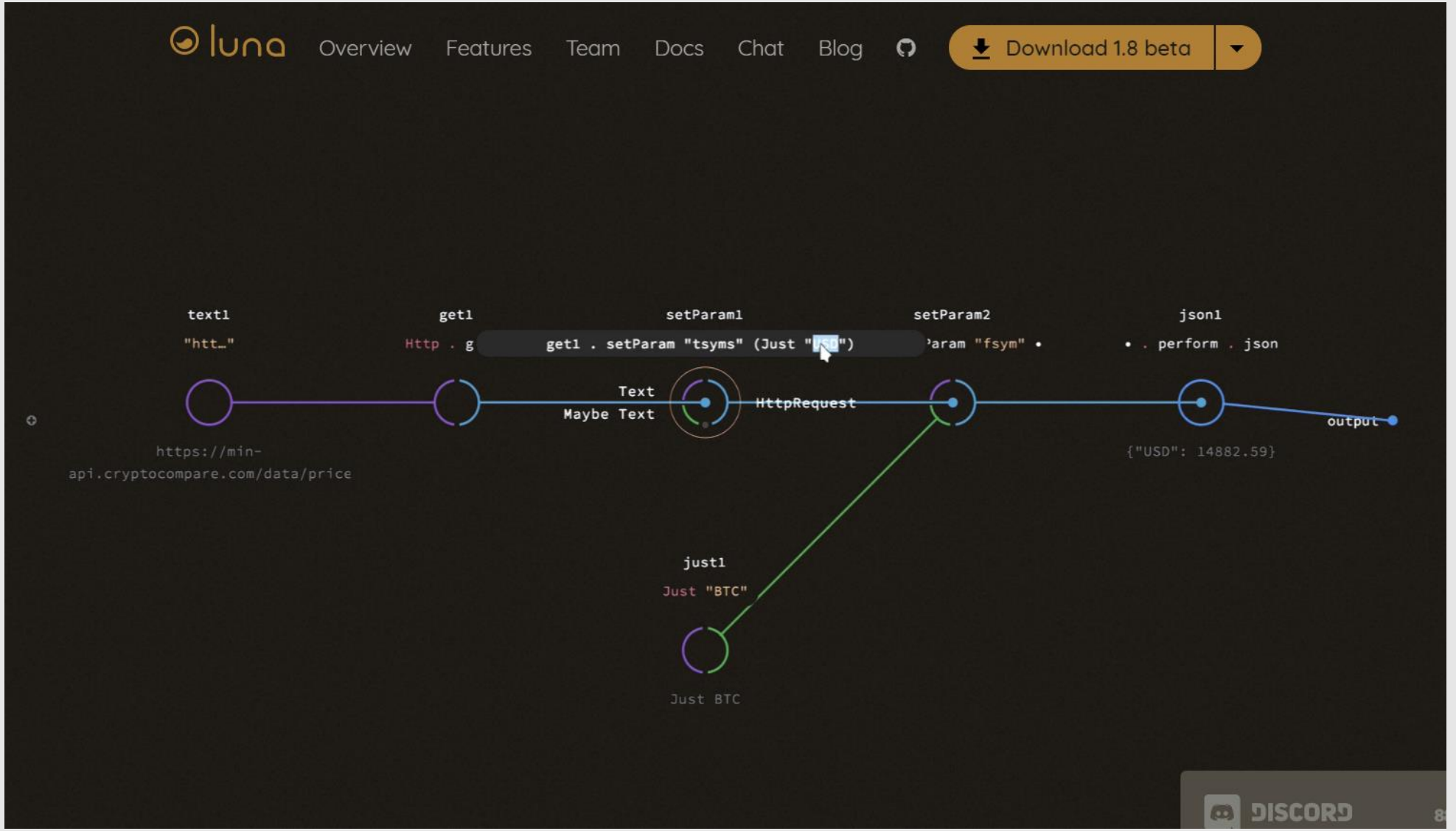
와쿠이 요시유키 · 와쿠이 사다미 저 | 권기태 역

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

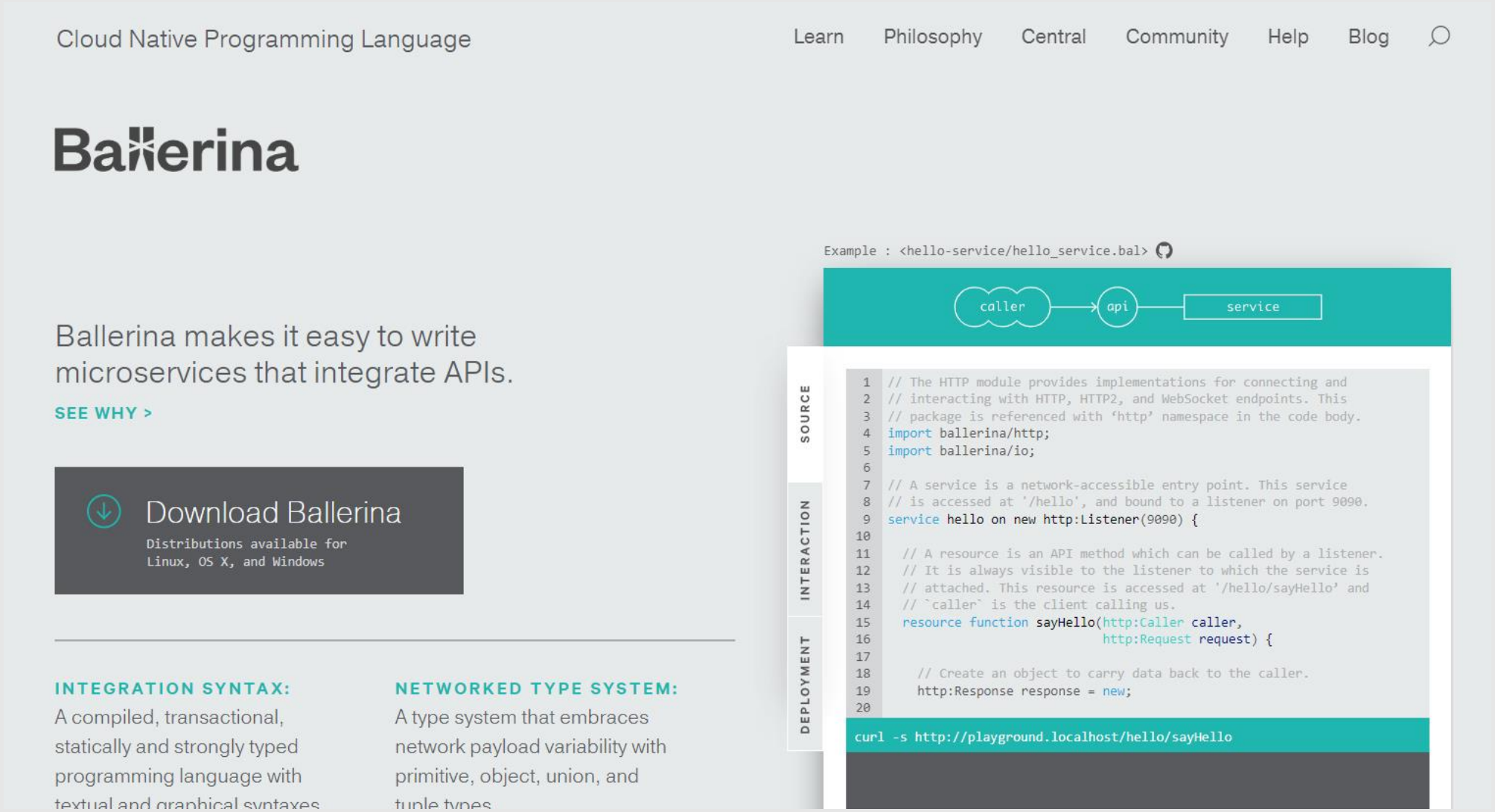
BM 성안당
www.cyber.co.kr

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]

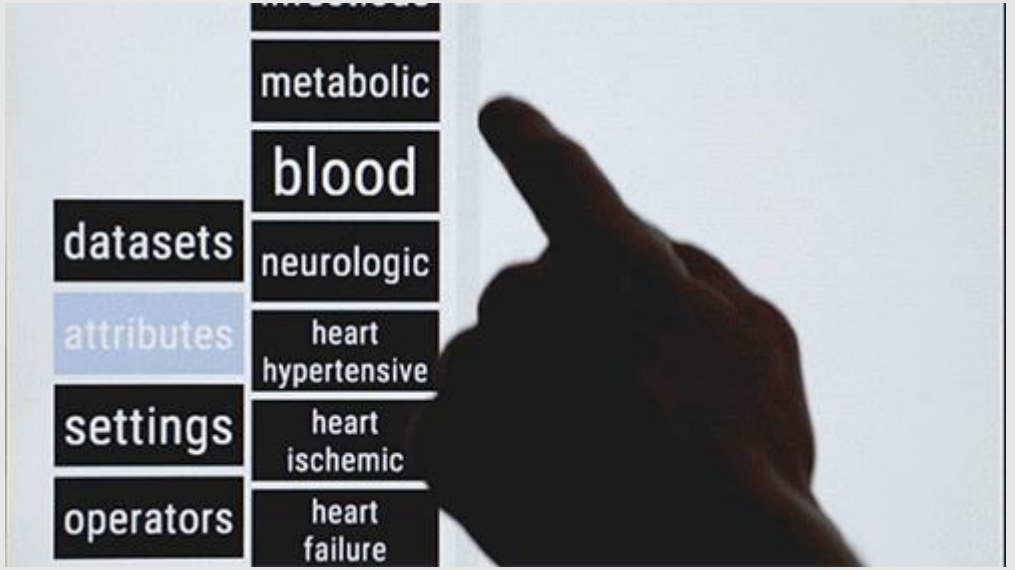
Visual Progamming Language / luna https://luna-lang.org



Visual Progamming Language / Ballerina https://ballerina.io

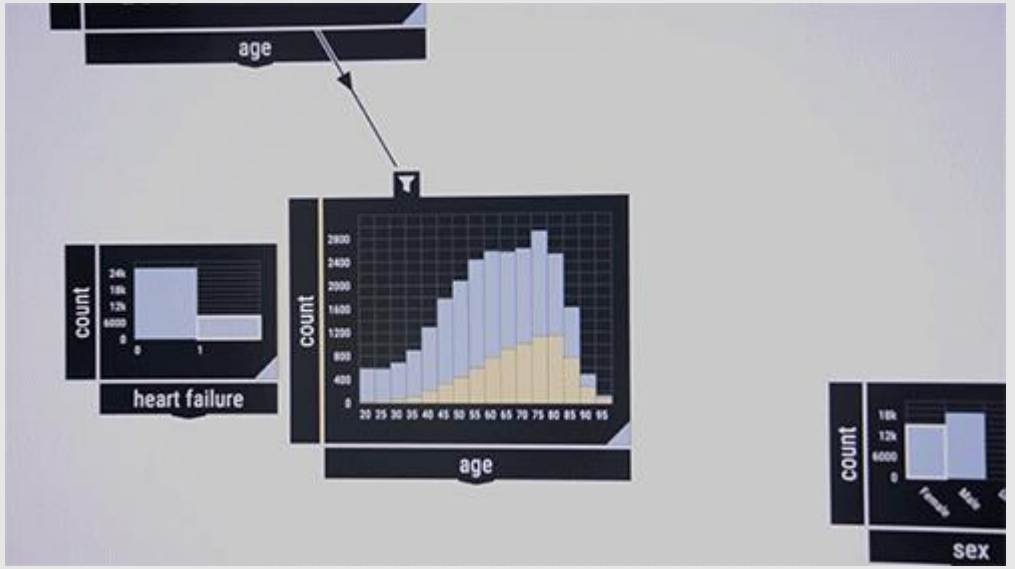


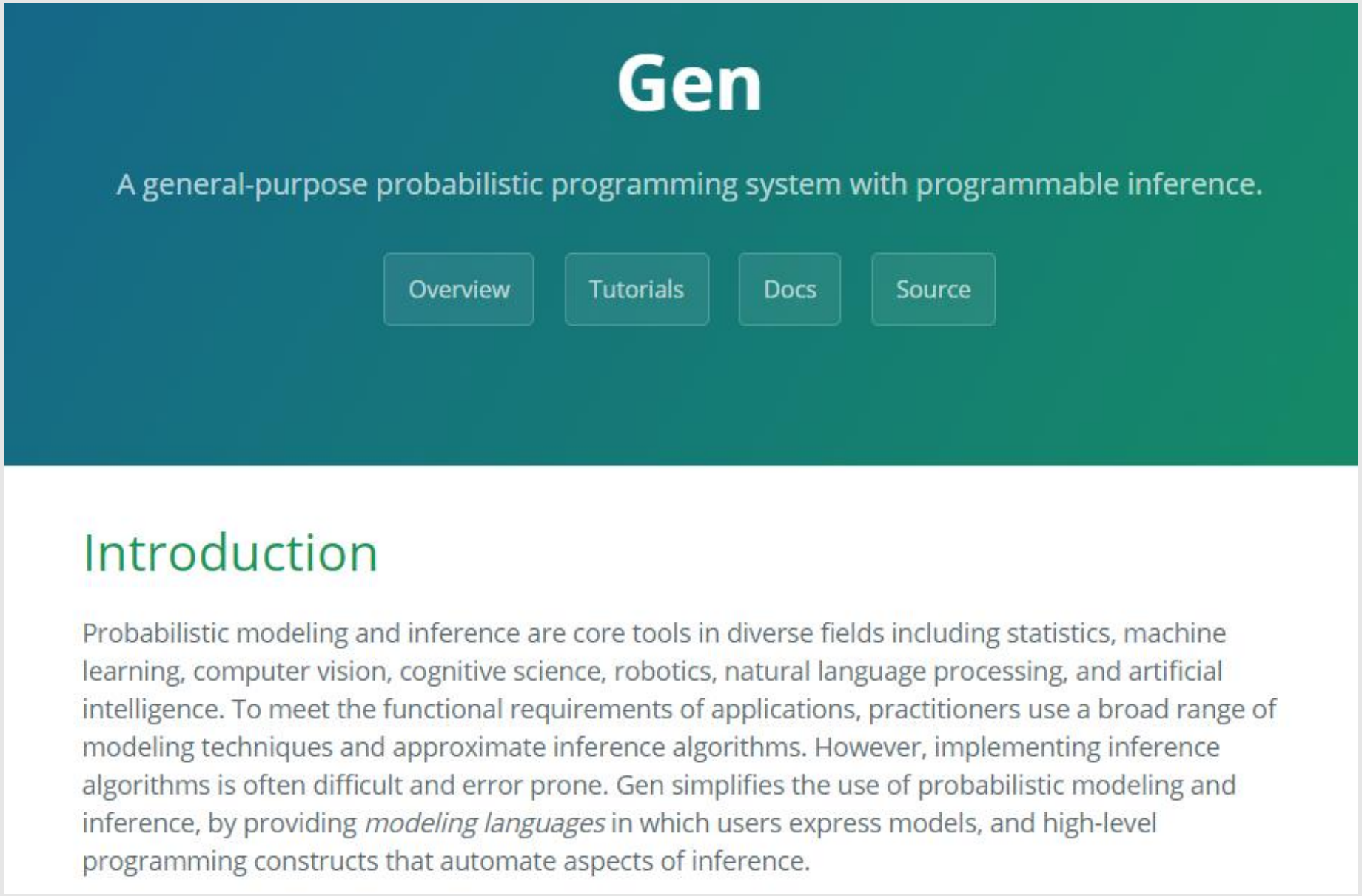
토니 스타크처럼 했으면 좋겠는데... 자비스?



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

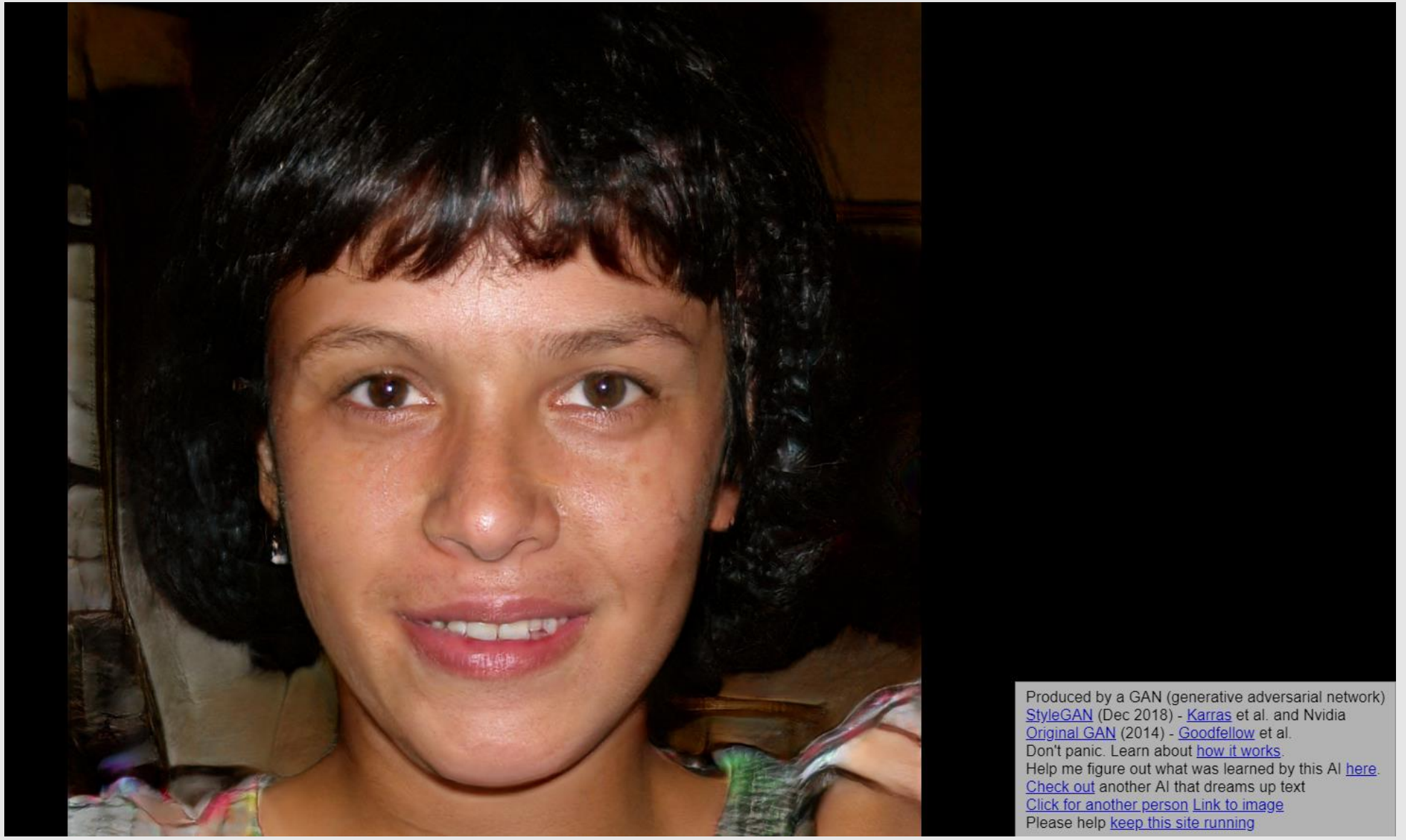
http://www.einblick.ai/





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

https://thispersondoesnotexist.com/

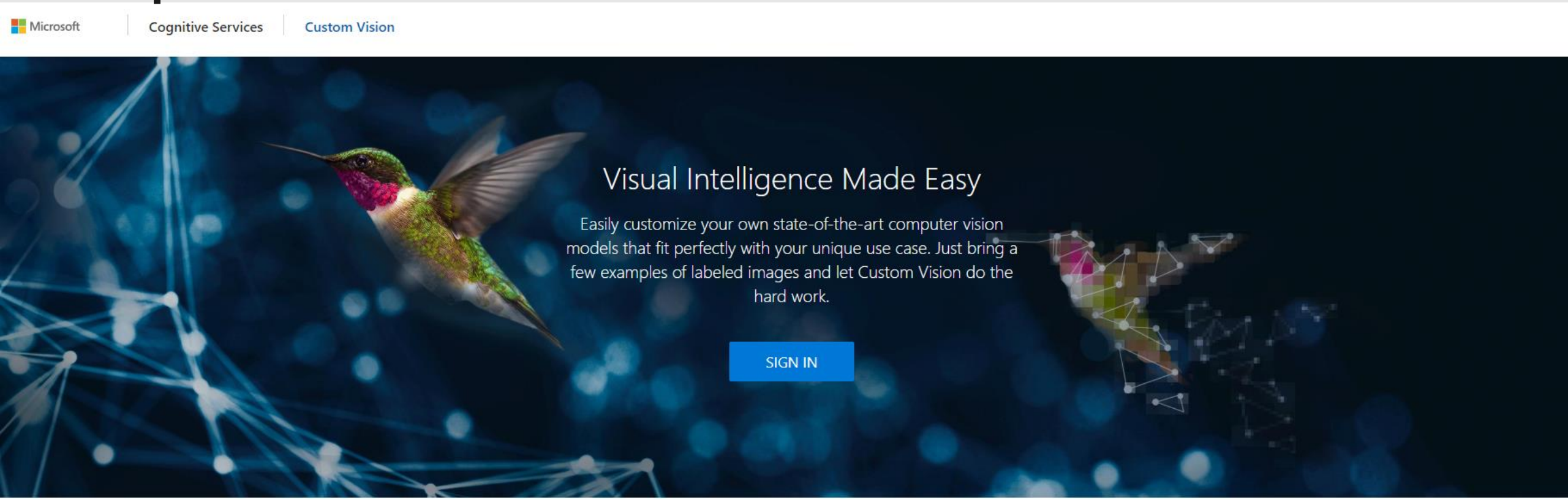


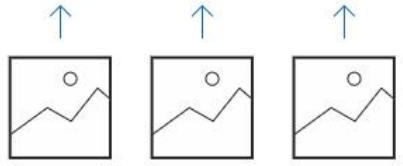
http://www.whichfaceisreal.com/index.php

PLAY ABOUT METHODS LEARN PRESS CONTACT CALLING BS

Click on the person who is real.

https://www.customvision.ai/



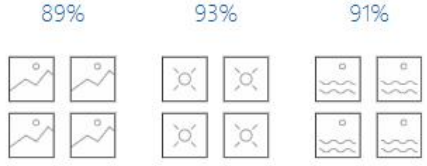


Upload Images

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

https://luminoth.ai/

LUMINOTH

Open source Computer Vision toolkit

GET STARTED!

Star 2,071 Fork

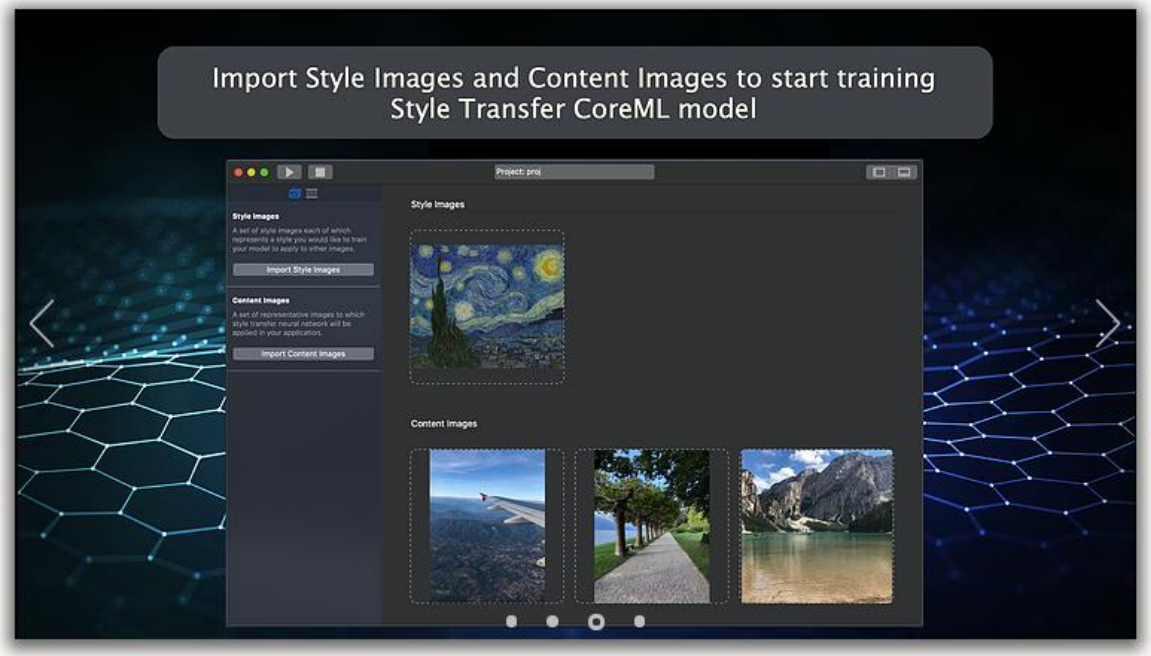
```
$ pip install luminoth
Successfully installed luminoth-0.2.0.

$ 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": 0.999
7},
{"bbox": [494, 289, 578, 439], "label": "person", "prob": 0.997
1},
[...]
]
}
```

MakeML

Home Blog Tutorial

MakeML https://www.makeml.app/

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



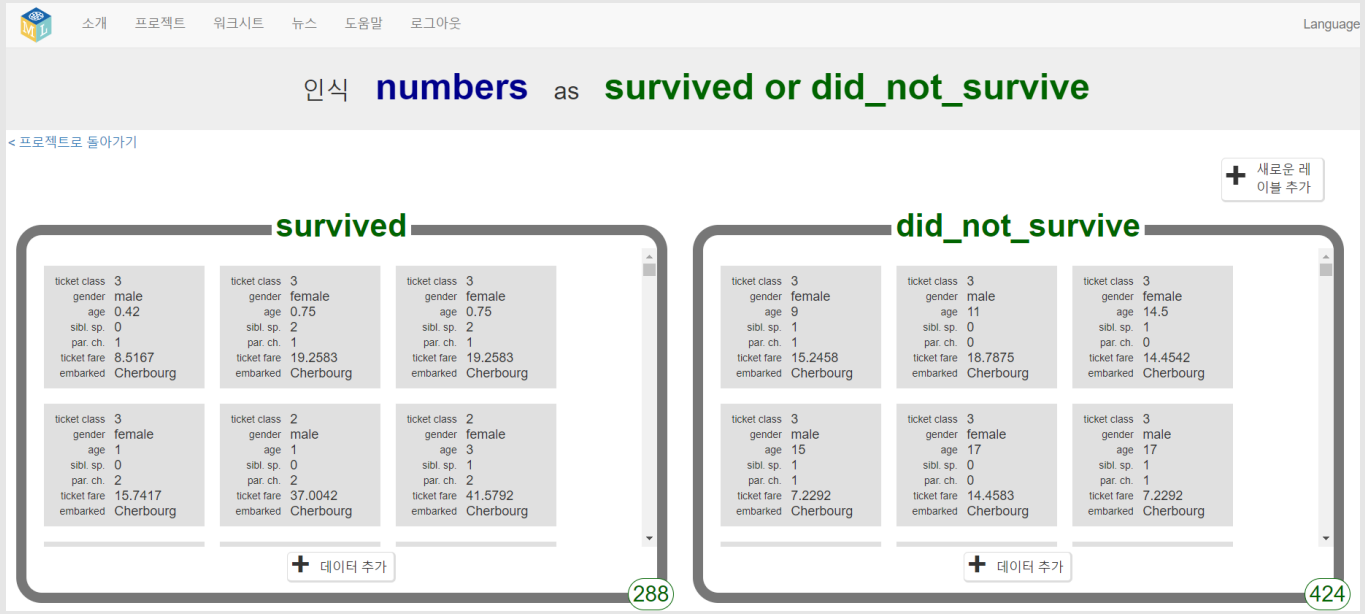
Wow! I want the app

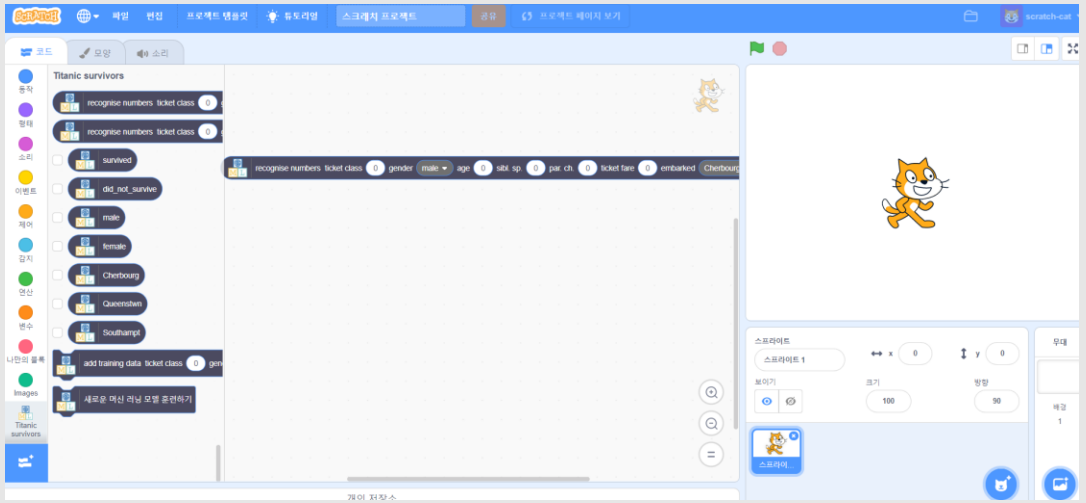
Name

Email

Send me link

https://machinelearningforkids.co.uk





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

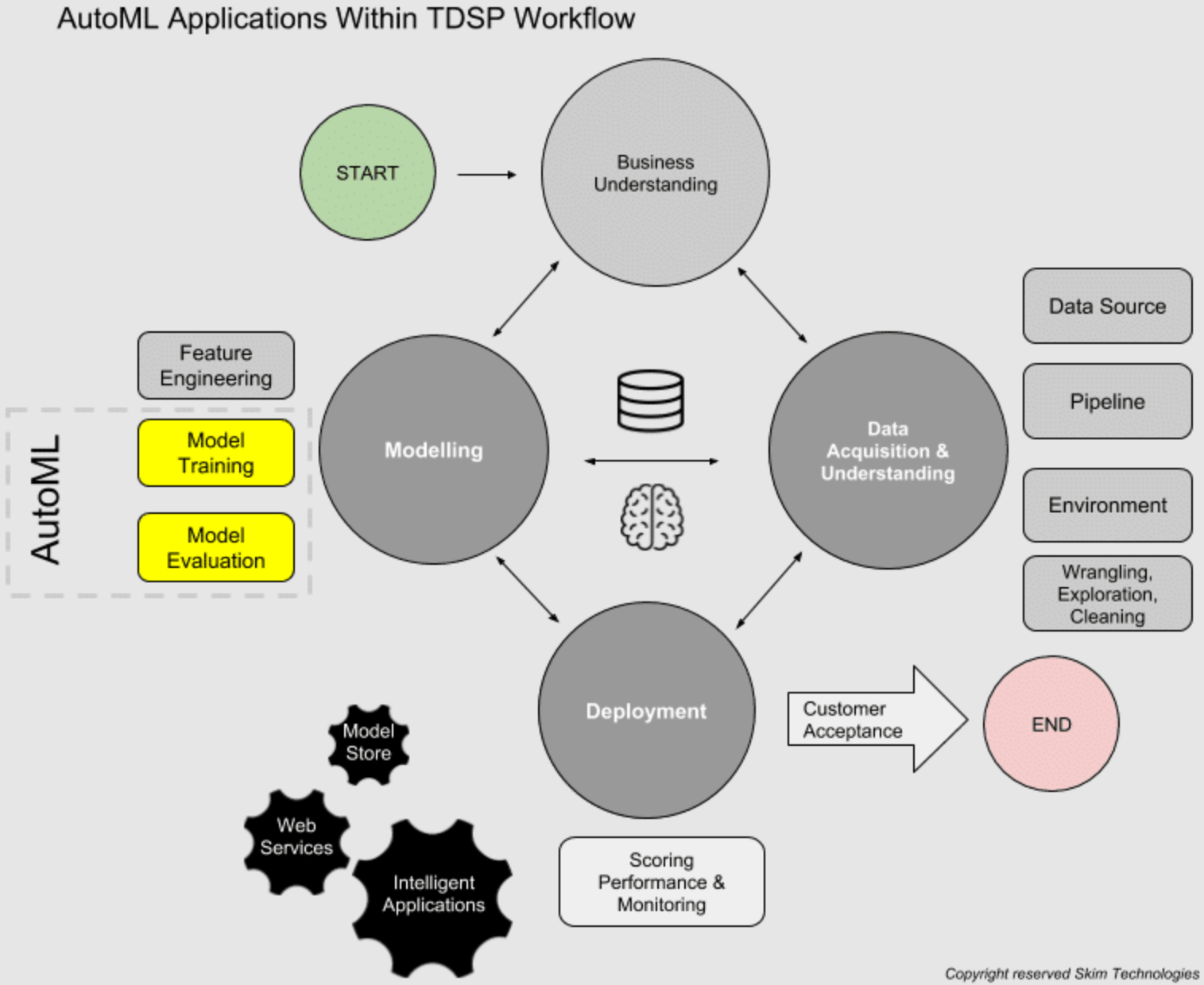
이 부분을 Automated ML에게 맡길 수 있을까요?

Automated ML의 3가지 한계

1. 비지도 학습 및 강화학습에 적용

2. 복잡한 데이터 형태

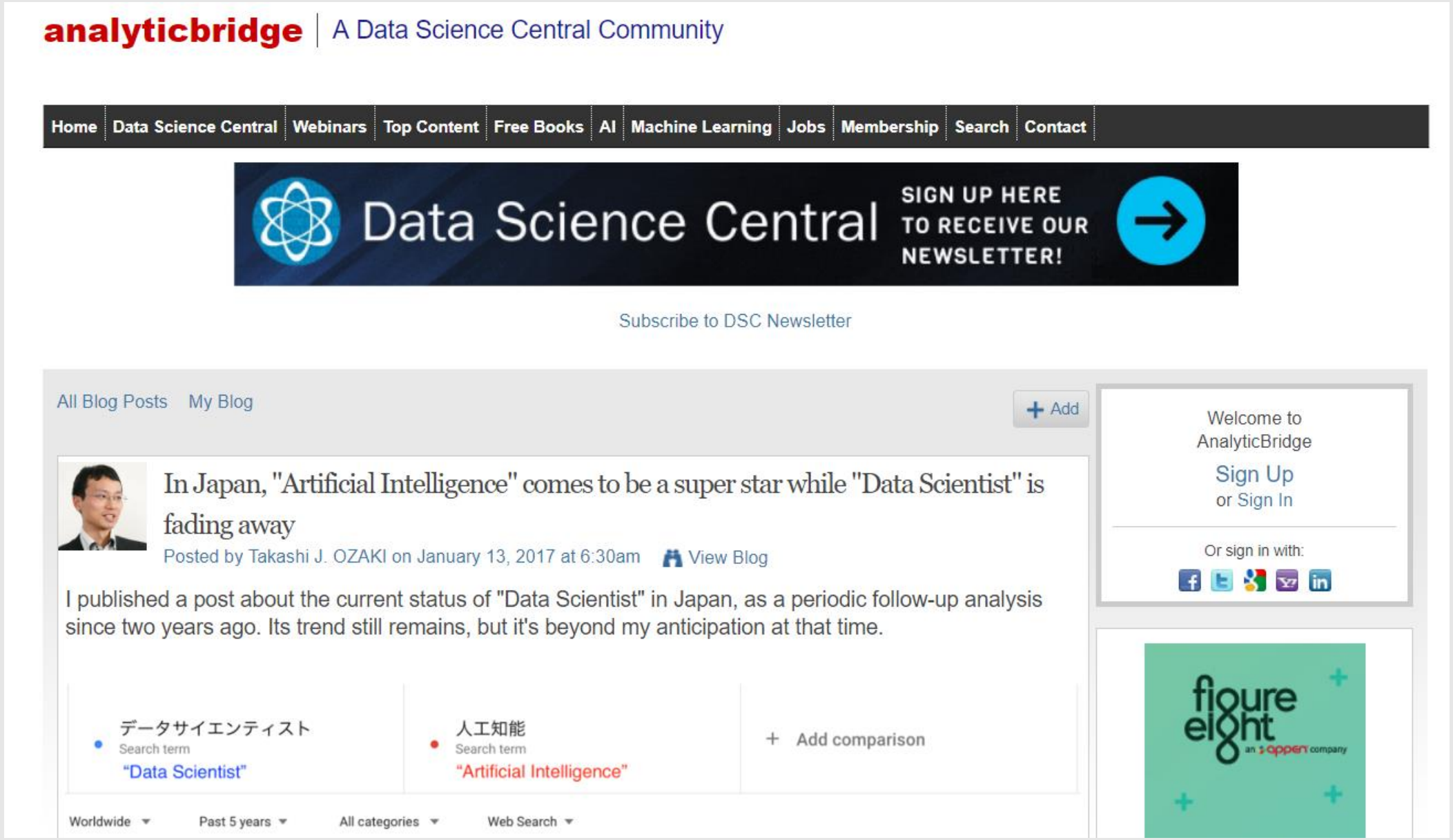
3. 도메인 지식이 필요한 분야



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

저는 지금까지 머신러닝
이야기를 좀 자제했습니다

다음 이야기를 함
보겠습니다



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

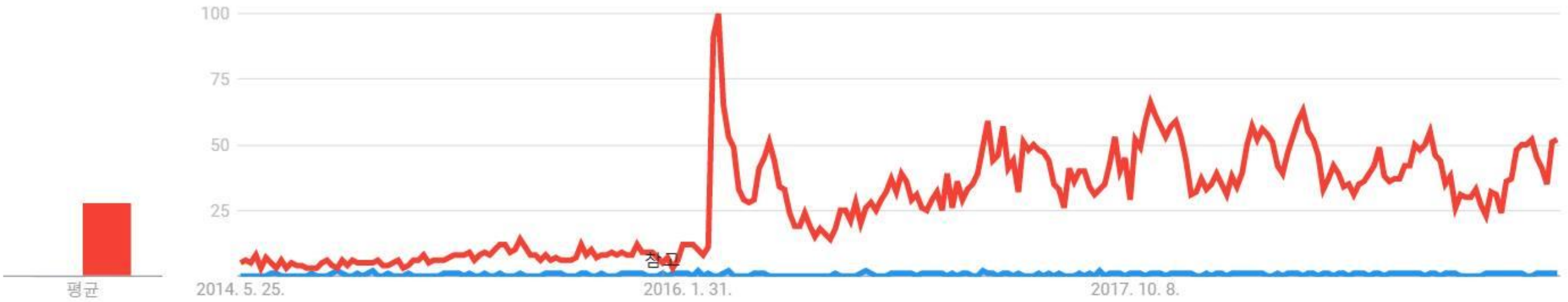
데이터 과학자
검색어

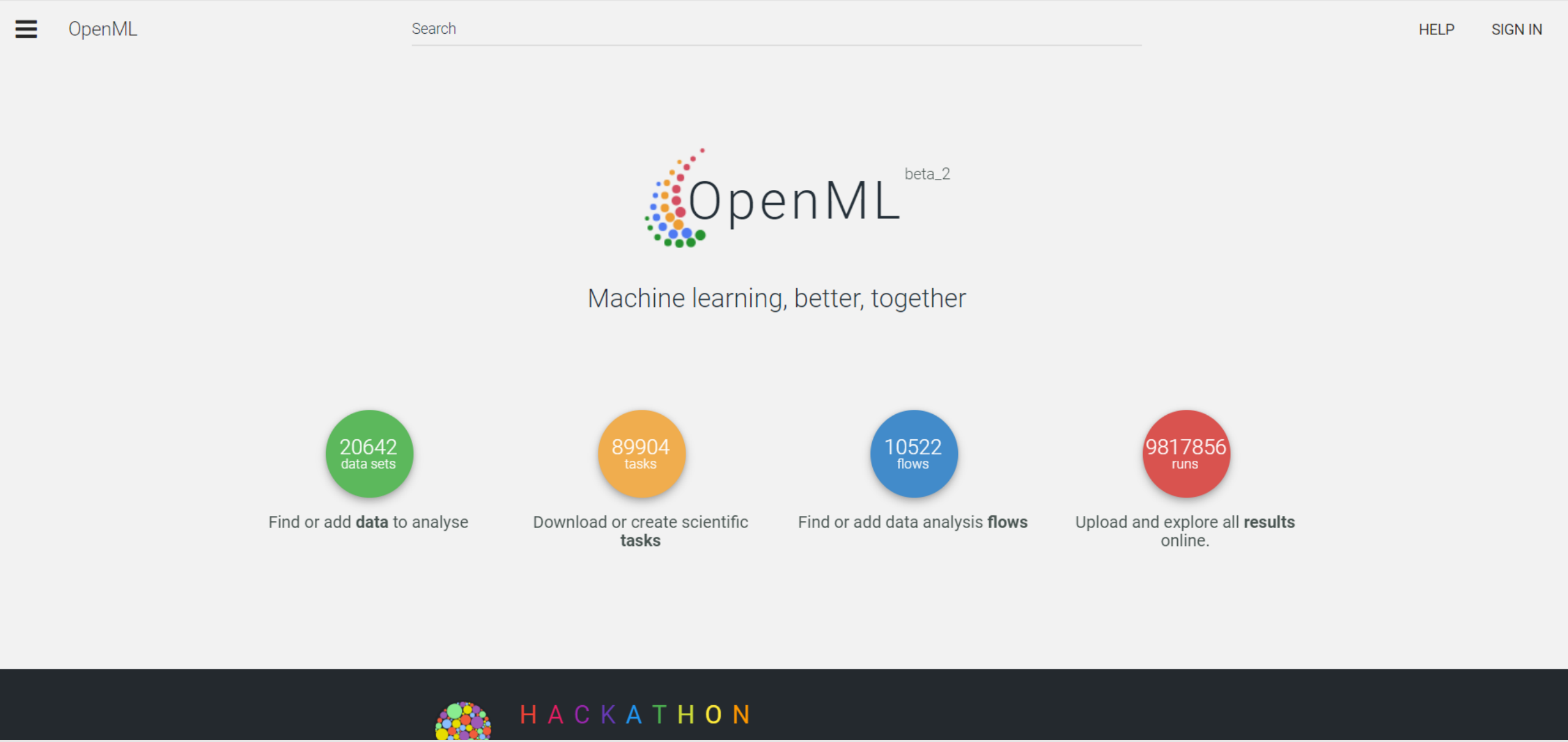
인공지능
검색어

+ 비교 추가

대한민국 지난 5년 모든 카테고리 웹 검색

시간 흐름에 따른 관심도 변화





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