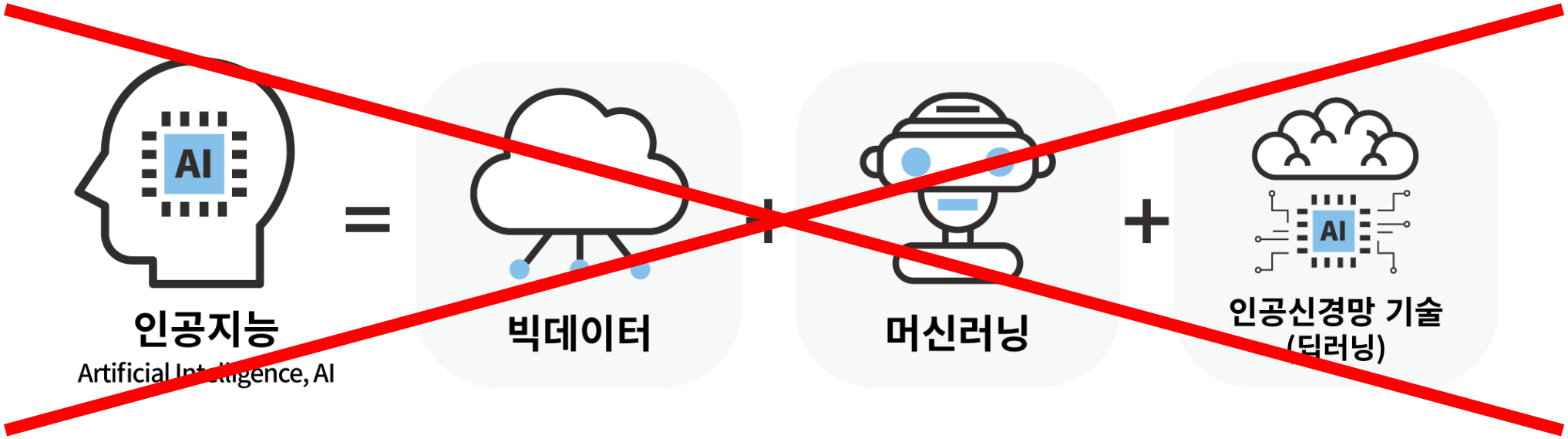
머신 러닝 기본 개념

2021.04.26.

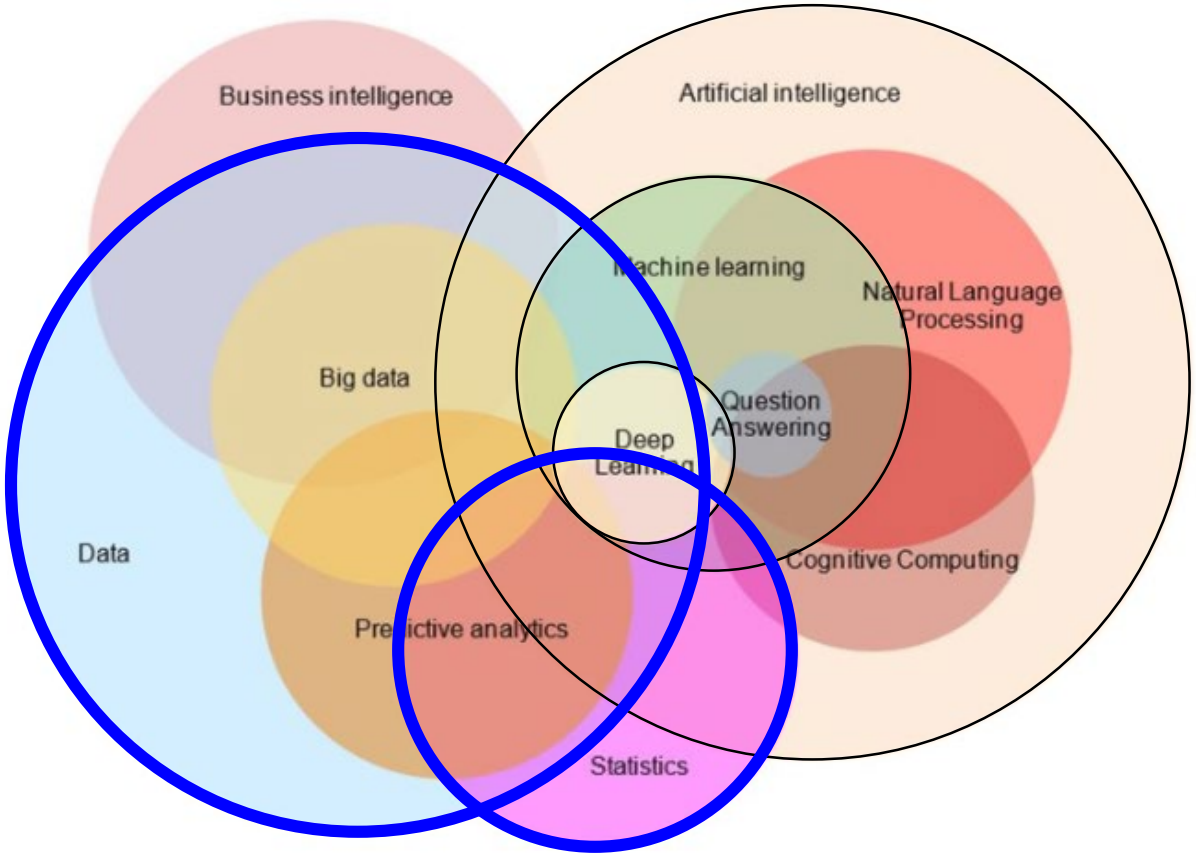
한국에너지기술연구원 계산과학연구실

이제현

인공지능 Artificial Intelligence



인공 지능 vs 머신 러닝



이미지=ibm.com

Artificial Intelligence
The broadest term used to classify machines that mimic human intelligence

Machine Learning
a method of data analysis that automates analytical model building

Deep Learning
automates much of the feature extraction piece of the process, eliminating some of the manual human intervention required and enabling the use of larger data sets.

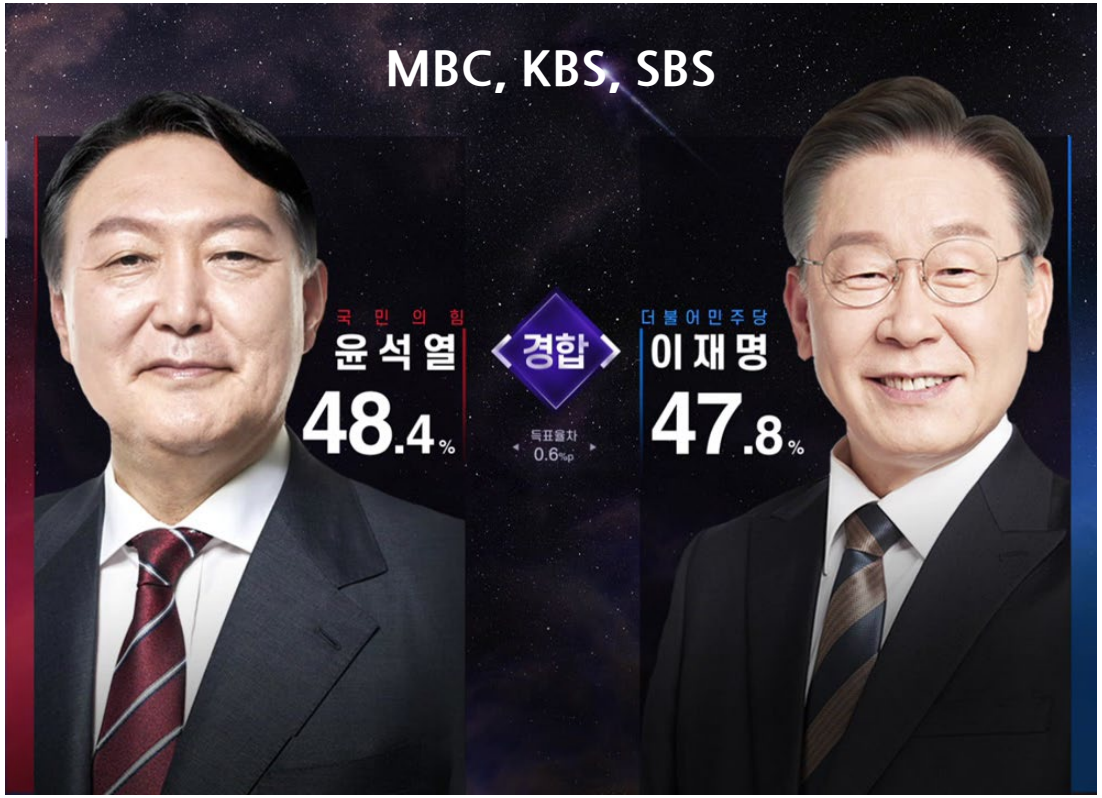
https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks
https://www.sas.com/en_us/insights/analytics/machine-learning.html

일반 연구 vs 머신 러닝

실험 연구	반응물	물리/화학 반응	생성물	장점 : real
전산 모사 연구 "계산과학" @KIER	조건 control	지배 방정식 fix	예측 결과	장점 : 분석력, 속도
머신 러닝 연구 "계산과학" @KIER "데이터 과학"	반응물, 조건 Model, HP control	pattern ???	생성물, 예측 결과	장점 : 비선형, 속도 + coverage

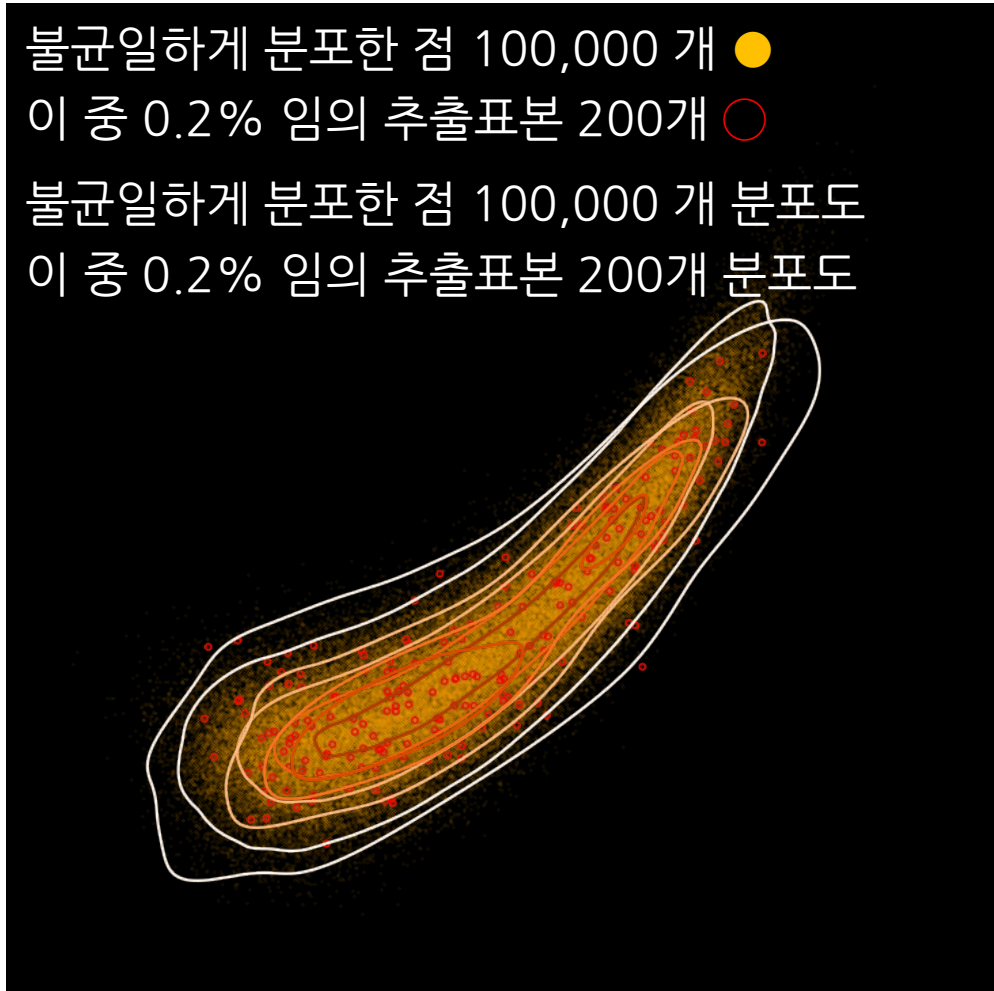
대통령 선거 출구조사 (2022. 3. 9)

- 지상파 3사 vs JTBC
  - 최종 결과: 李 47.83% vs 尹 48.56%



대통령 선거 출구조사 (2022. 3. 9)

- 대한민국 유권자 수 :
-
-
-



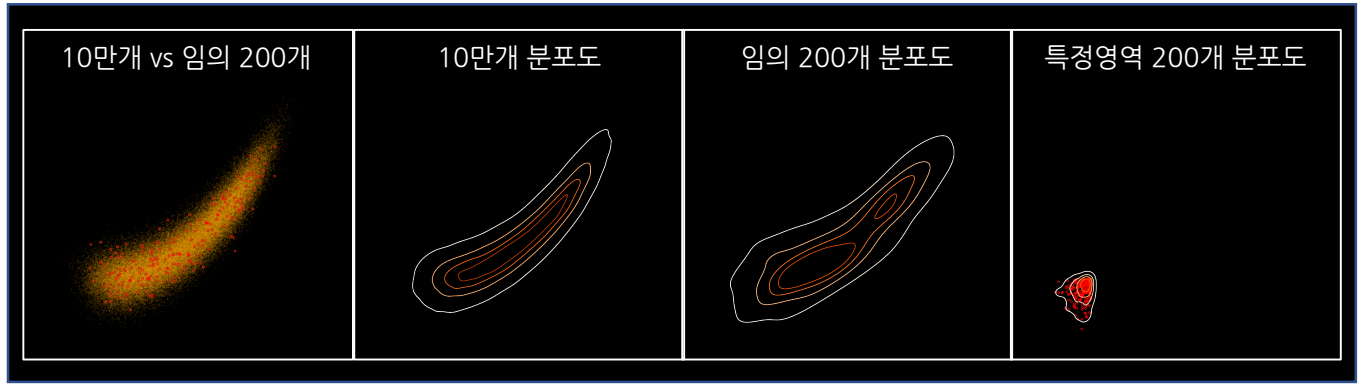
https://www.bbc.com/korean/news-60659235

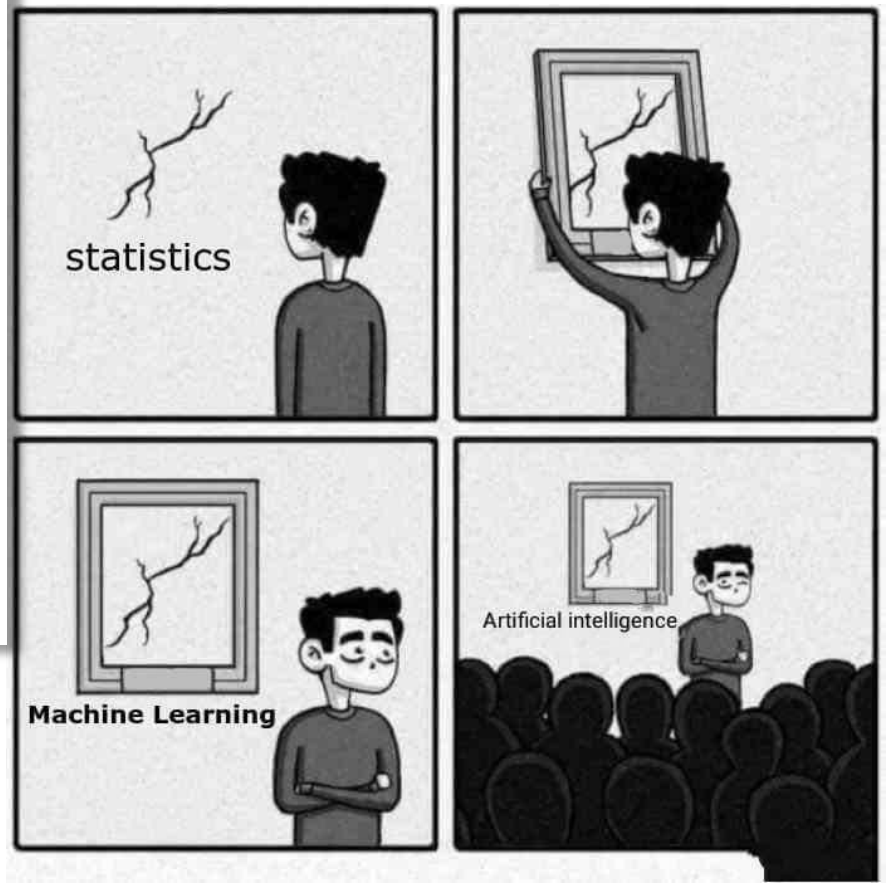
Machine Learning ~ Statistics

- 부분으로 전체를 추정한다.

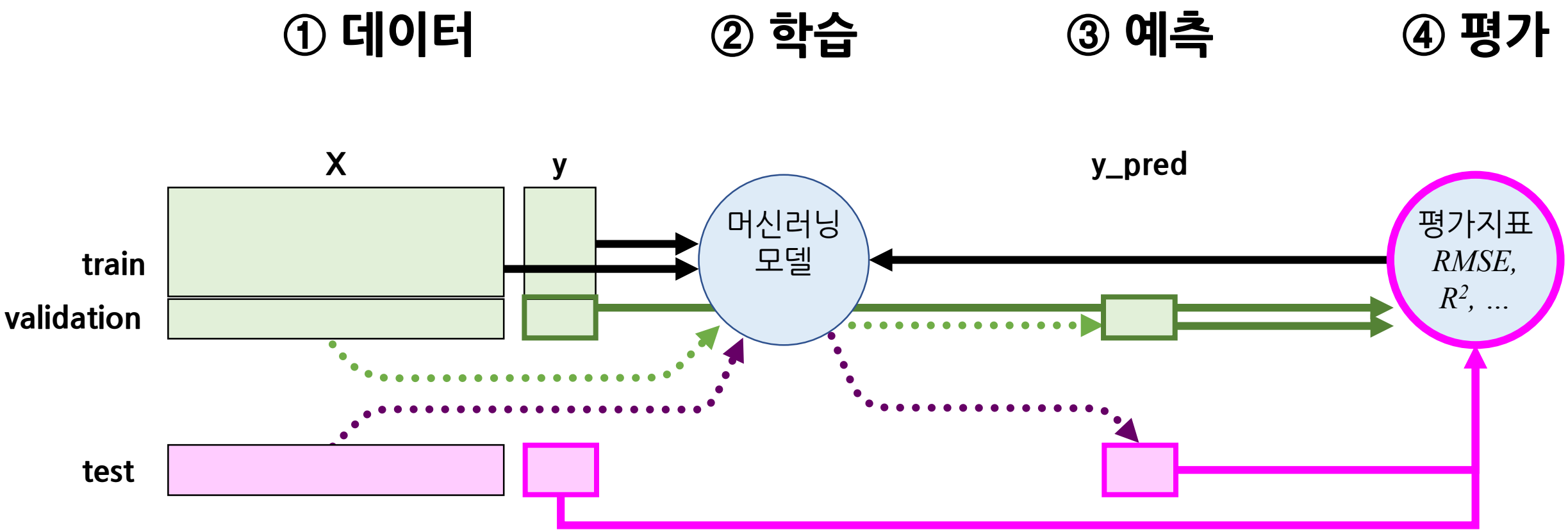
만약 부분이 전체를 대변하지 못한다면?

= 근본부터 폭망

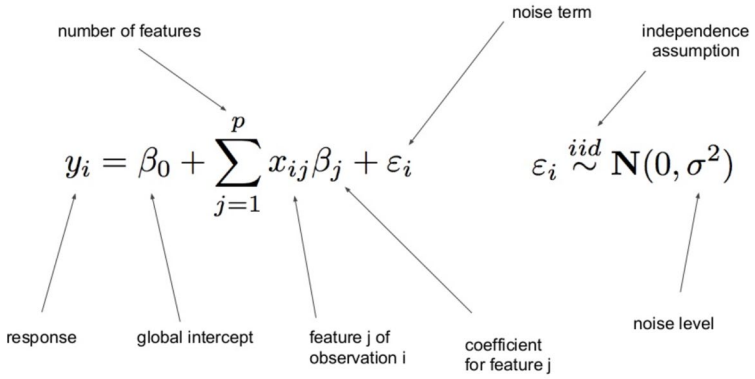


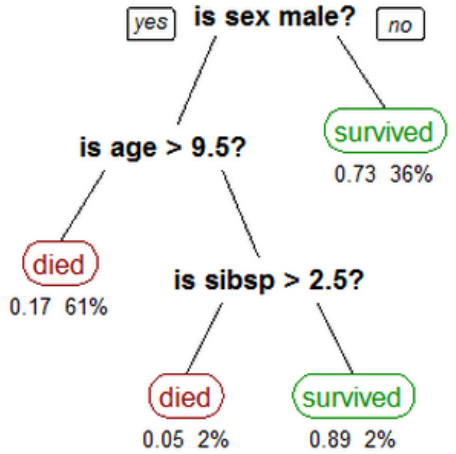


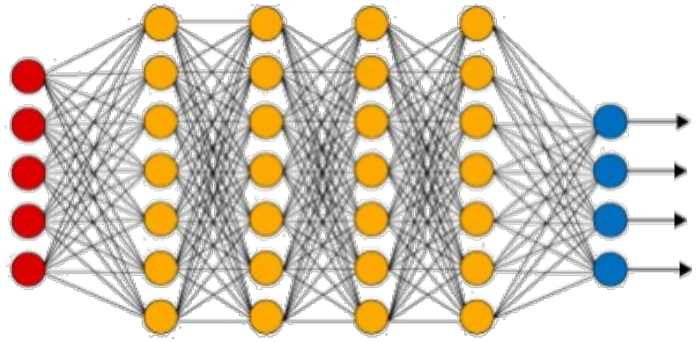
머신 러닝 과정


① 데이터
② 학습
③ 예측
④ 평가
X
y
y_pred
train
validation
test
머신러닝
모델
평가지표
RMSE,
R^2, ...

머신 러닝 Machine Learning

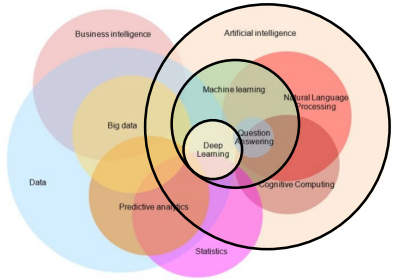




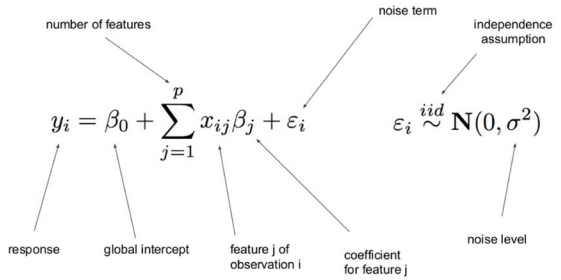


	선형 모델 Linear Model	트리 모델 Tree Model	신경망 모델 Neural Network
비선형성	X	O	◎
설명력	O	O	△
속도	O	△	X
비용	X	X	◎

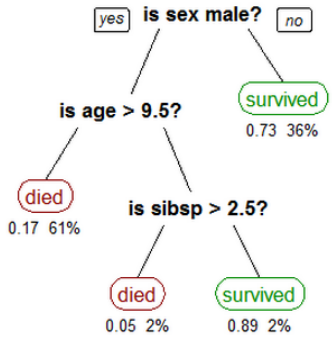
딥 러닝 Deep Learning



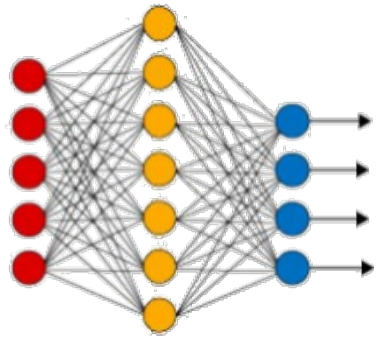
선형 모델



트리 모델



신경망 모델



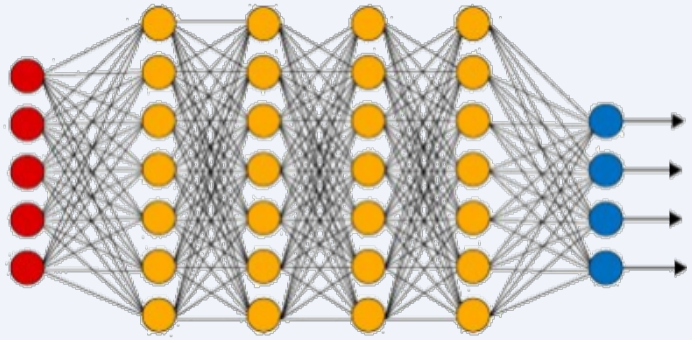
사전적 의미:

머신러닝 ⊃ 딥러닝

최근 통용되는 의미:

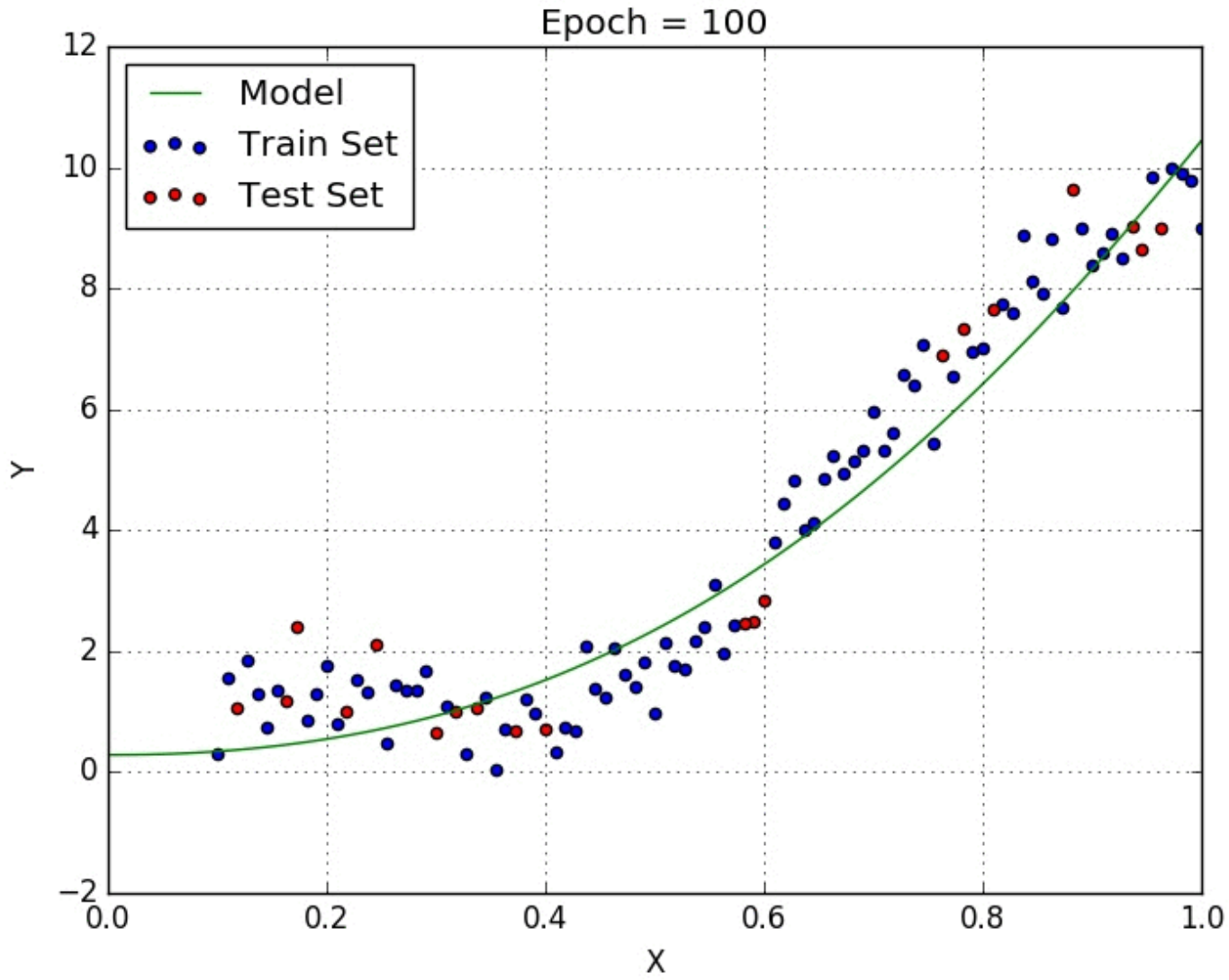
머신러닝 = 딥러닝^C

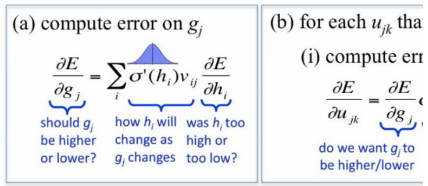
심층 신경망 Deep Neural Network



딥 러닝 Deep Learning

- perceptron



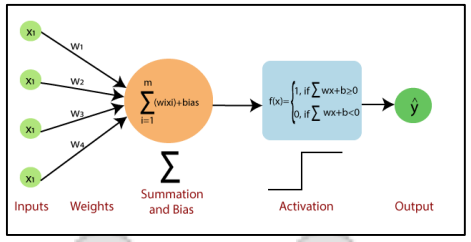


https://www.javatpoint.com/single-layer-

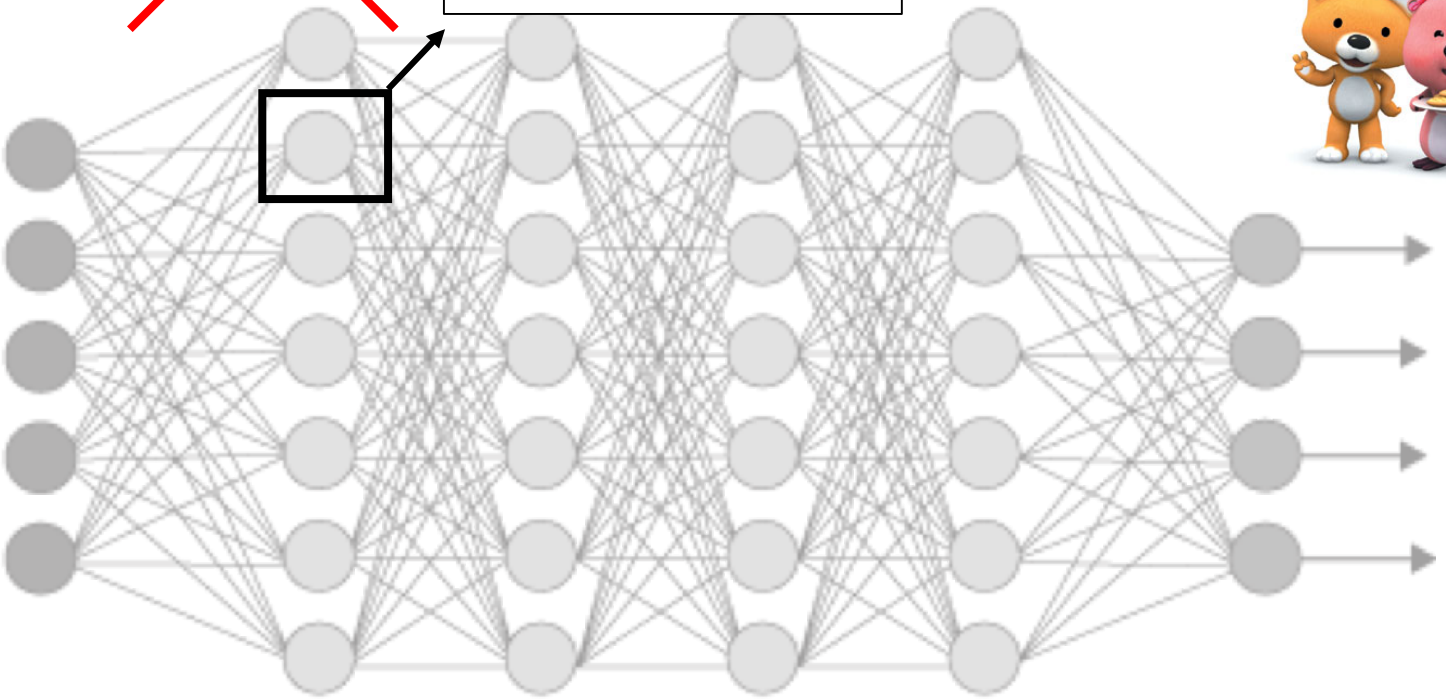
딥 러닝 Deep Learning

- Multilayer Perceptron (MLP) = Deep Neural Network (DNN) = Deep Learning (DL)

~~• 미분~~
~~• 적분~~
~~• 기타 등등~~

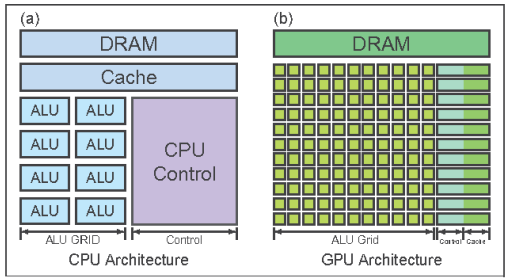


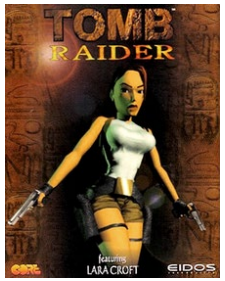
• 곱하고
• 더하고
• 활성화하고



딥 러닝 Deep Learning

- GPU: Graphic Processing Unit
- ALU: Arithmetic Logic Unit
  = 덧셈 곱셈 계산기

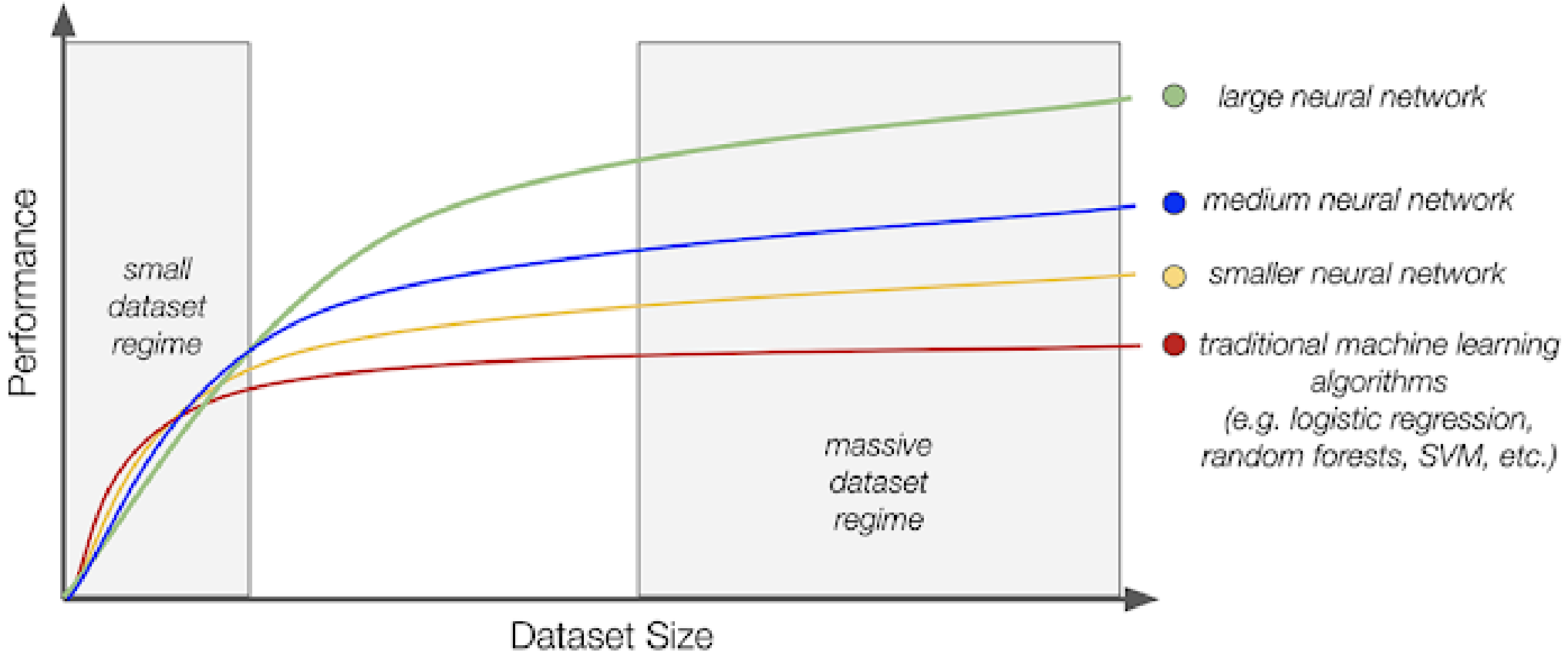




- **Tomb Raider (1996)**

https://www.youtube.com/watch?v=rFyGSxs31Zo

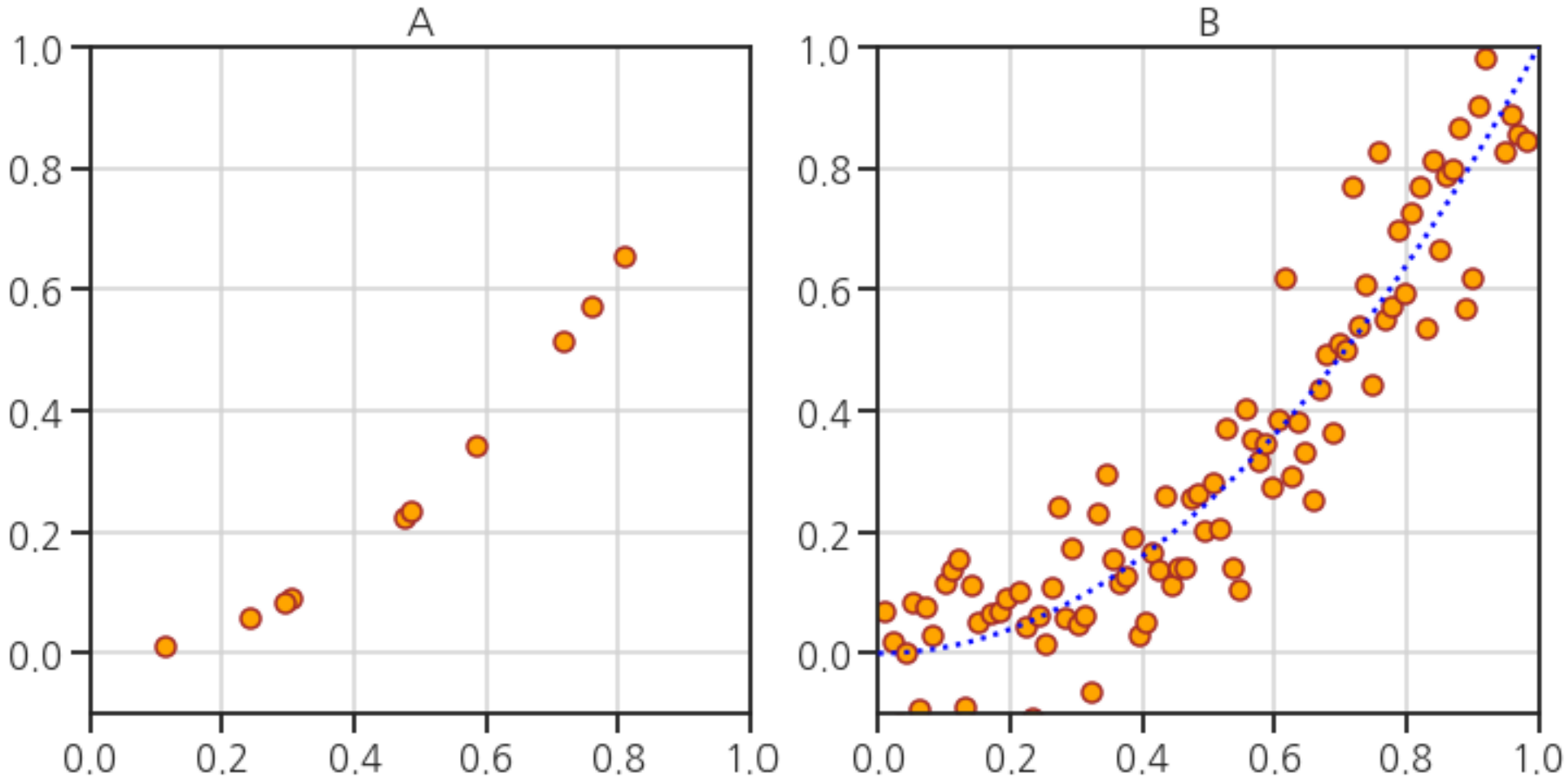
데이터 수 vs 모델 성능



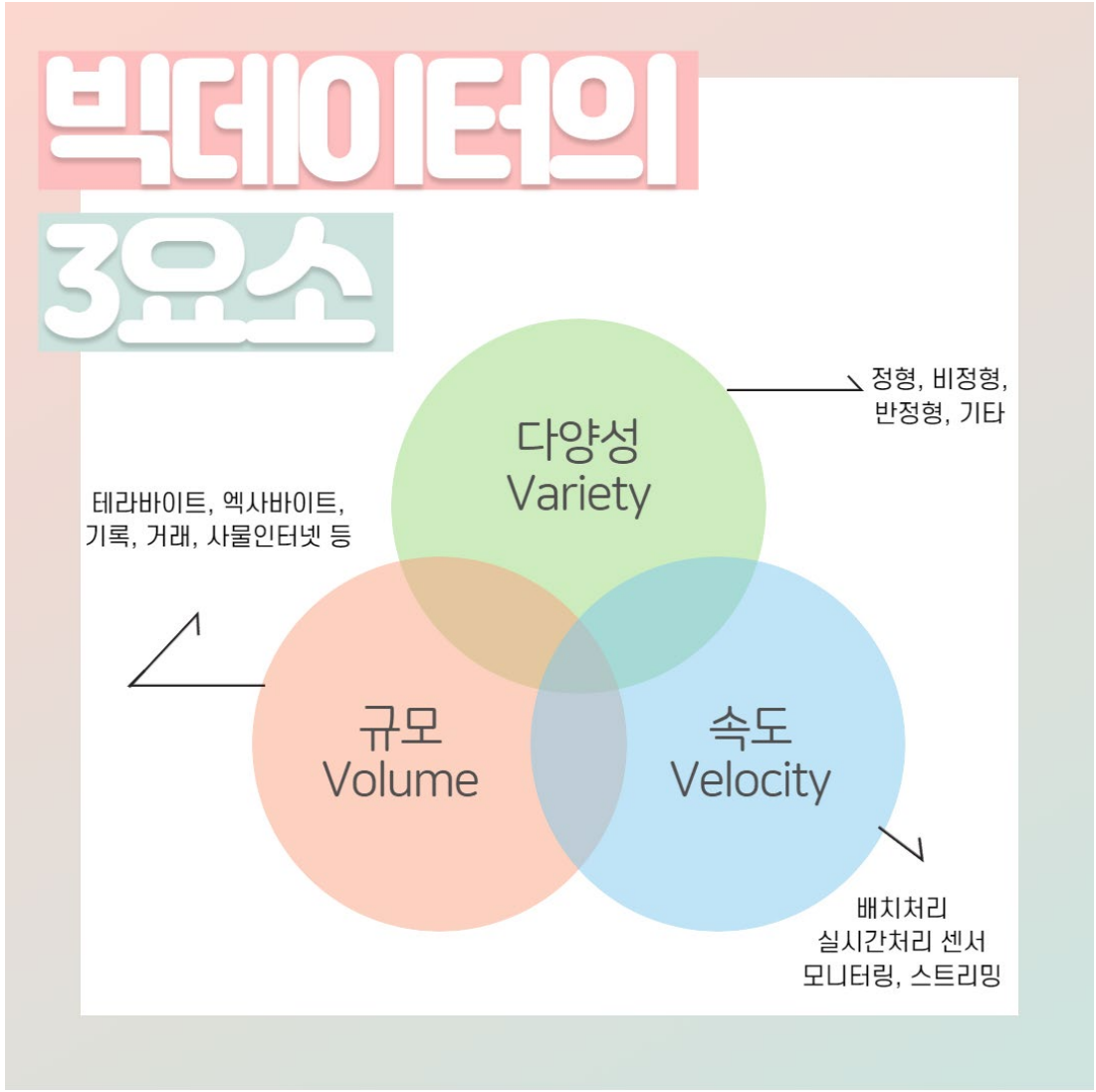
http://www.compchemhighlights.org/2020/05/learning-molecular-representations-for.html

데이터 수가 어느 정도 필요한 이유

- 노이즈를 헤치고 패턴을 찾아야 하기 때문



빅데이터 Big Data



Opinion : 유혁의 데이터 이야기

빅 데이터로 재미 좀 보셨습니까?

중앙일보 | 입력 2019.08.05 00:25 업데이트 2019.08.23 14:32 지면보기

유혁 윌로우 데이터 스트래티지 대표

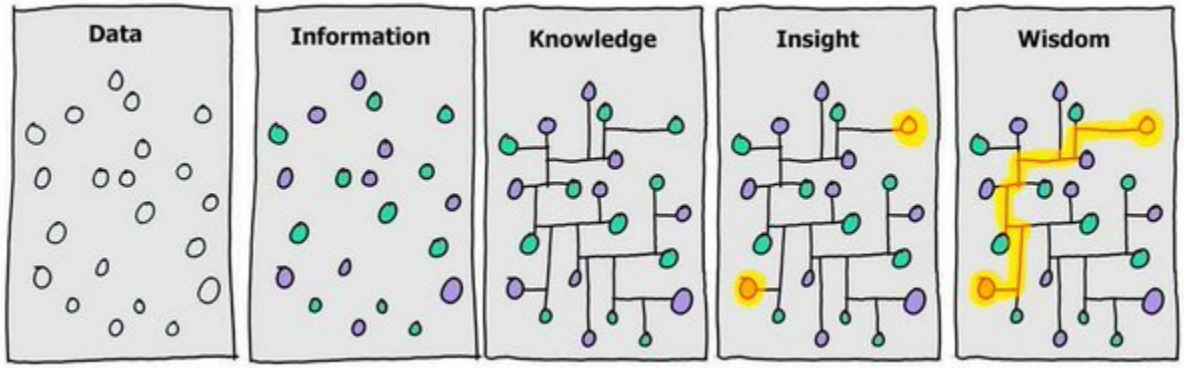
빅 데이터란 말이 유행하기 시작한 지도 꽤 오래되었다. 사실 늘 데이터를 다루며 그것으로 가치를 창출하는 작업을 오랫동안 해 온 사람들에게는 그건 애초부터 적절하지 않은 표현이었다. 데이터를 주 기반으로 하는 구글이나 아마존 같은 거대기업들은 그런 말을 쓰지도 않는다.

우선 그 "빅"이라는 단어에 많은 오해의 소지가 있다. 처음 빅 데이터란 말이 쓰이기 시작했을 때에는 많은 이들이 그것을 3V, 즉 Volume(크기), Velocity(속도), Variety(다양성)로 정의하곤 했다. 그러나 이런 기술적 정의도 다양하고 방대한 양의 데이터가 빨리 돌아다니기만 하면 저절로 가치가 창출된다는 오해를 낳아서 바람직하지 않다. 그 광고적 표현을 만들어낸 소프트웨어 회사들의 의도도 "이제는 아주 방대한 데이터도 처리할 수 있다"이지 "데이터는 커야만 한다"가 아니었다.

https://www.joongang.co.kr/article/23543744#home

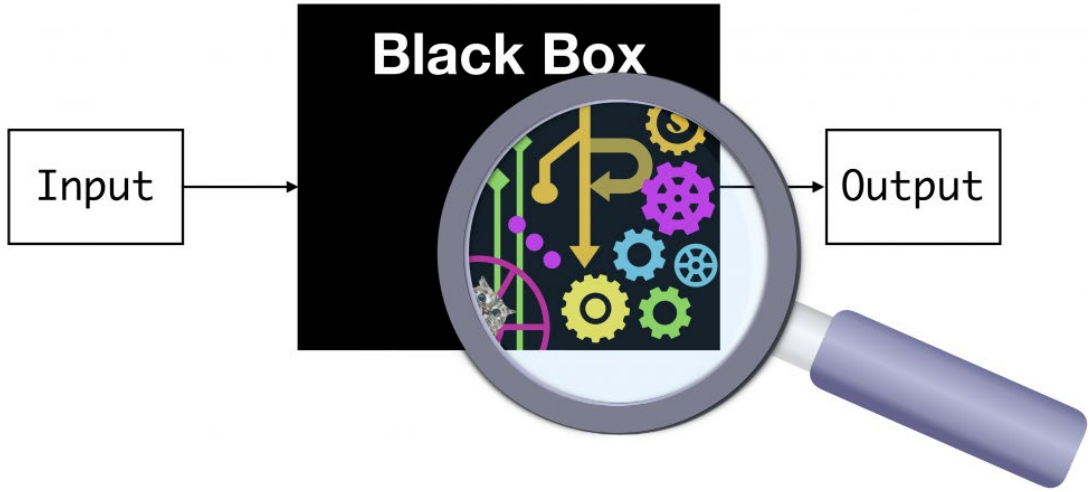
데이터 분석 vs 머신 러닝

- **데이터 분석 :**



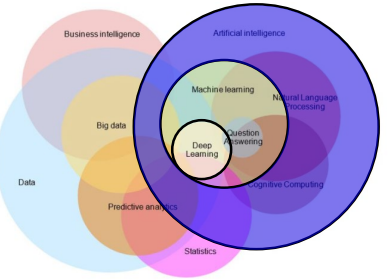
이미지 = www.kaushik.net

- **머신 러닝 :**
  (딥 러닝)

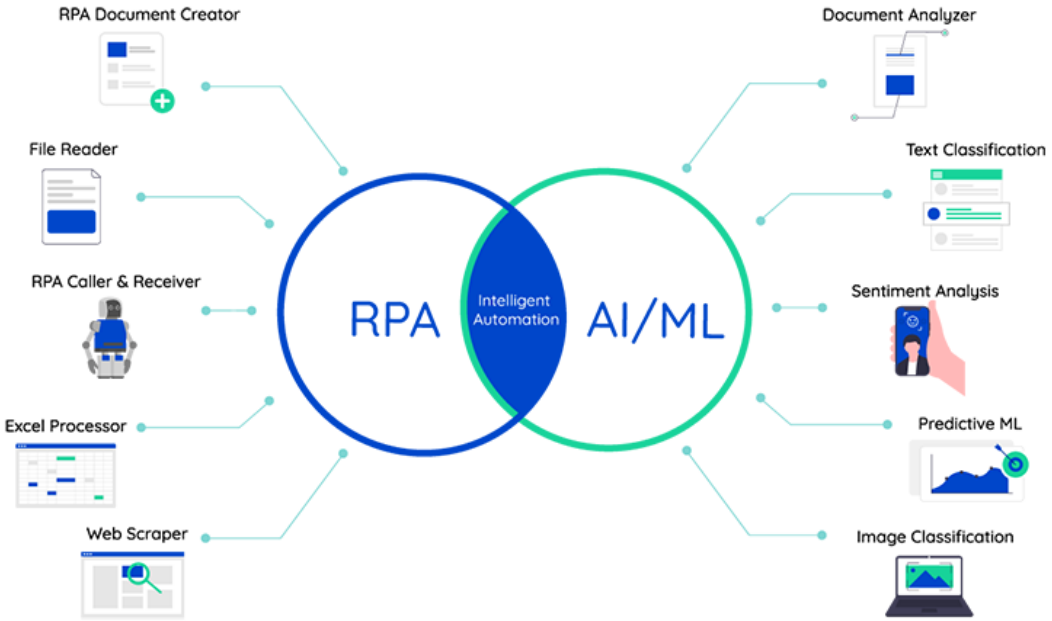


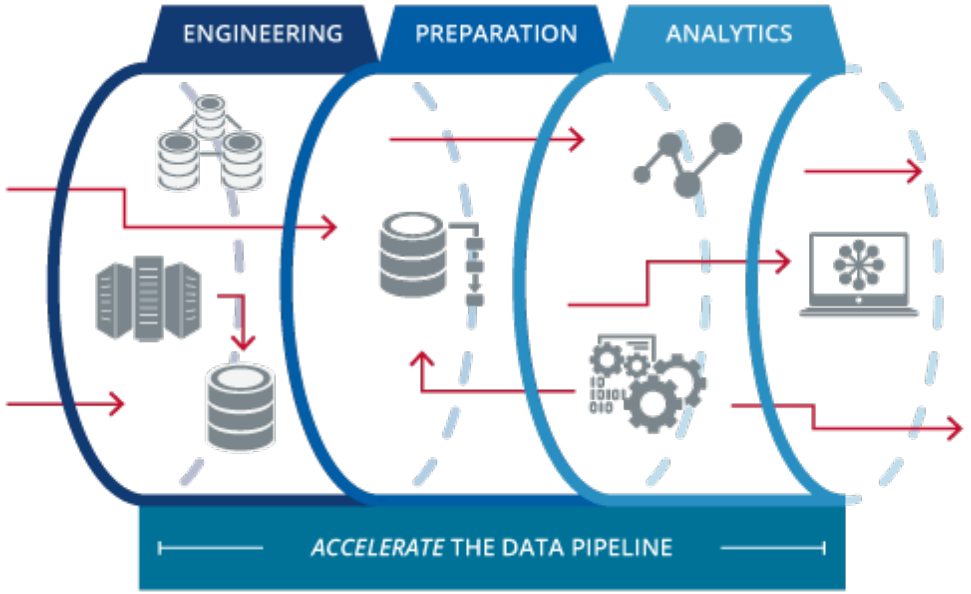
이미지 = CMU ML blog

업무 효율화 RPA vs 머신 러닝 ML



- **RPA :** 원활한 ML을 위한 선결 조건
  - "전방에 탄약이 모자랍니다!" vs "기다려. 장인들이 한 땀 한 땀 만들고 있으니까. 중요한 거라 아무나 만들면 안 돼."
- **Data Pipeline :** 조직에 데이터가 흐르는 경로
  - 필요할 때마다 강에서 양동이로 물 떠오기 (성실함) vs 수도꼭지 틀기 (시스템)

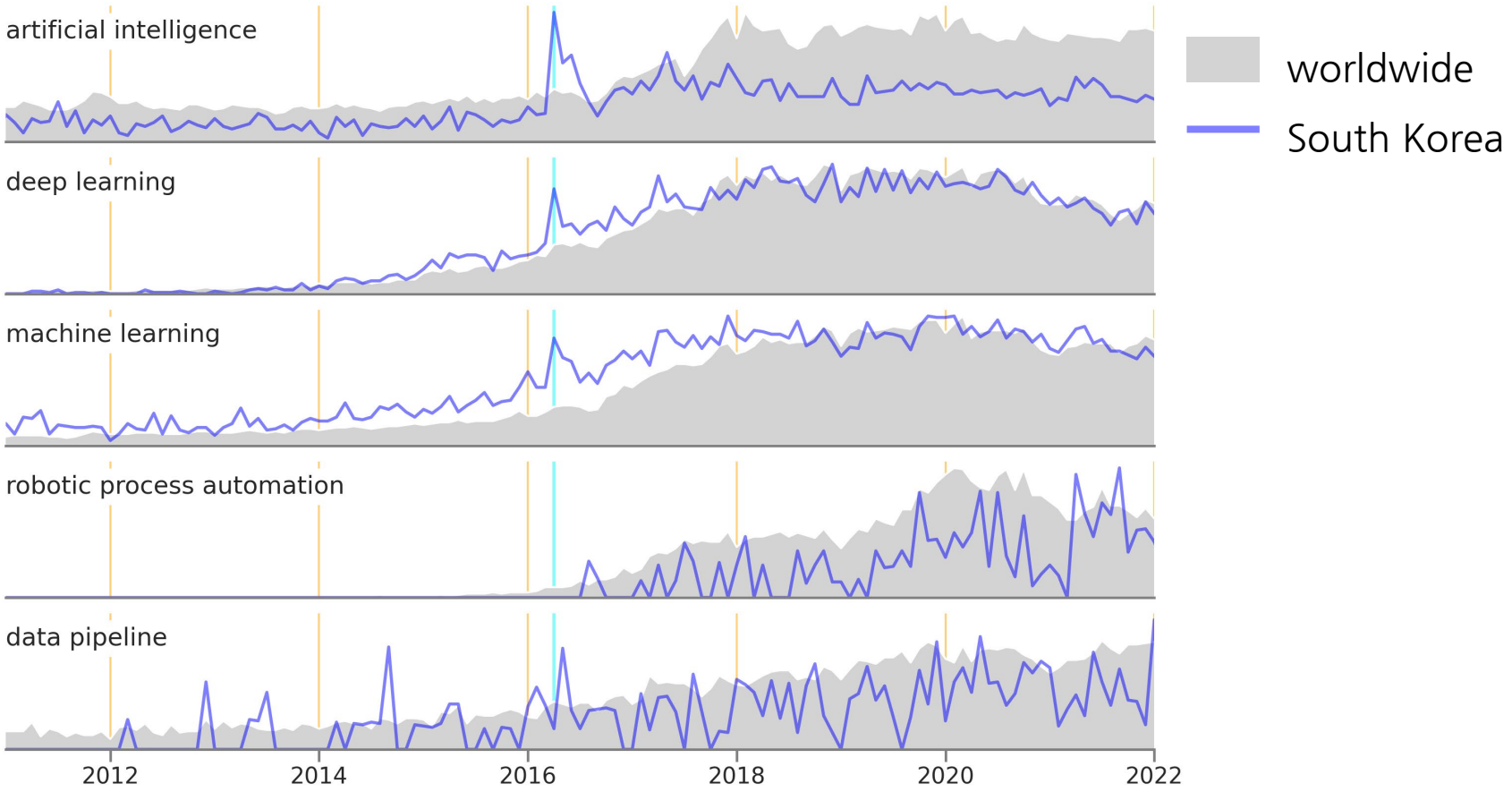




https://codimor.com/ https://tigosoftware.com/what-data-pipeline

Google Trends : World vs South Korea

- 알파고 이후 딥 러닝, 머신 러닝에 대한 관심은 세계 평균보다 빠름
- 그러나 기초 체력에 해당하는 RPA, 데이터 파이프라인은 매우 미진함.



타자기

- 기술은 있지만 시스템이 없는 경우

주의 사항

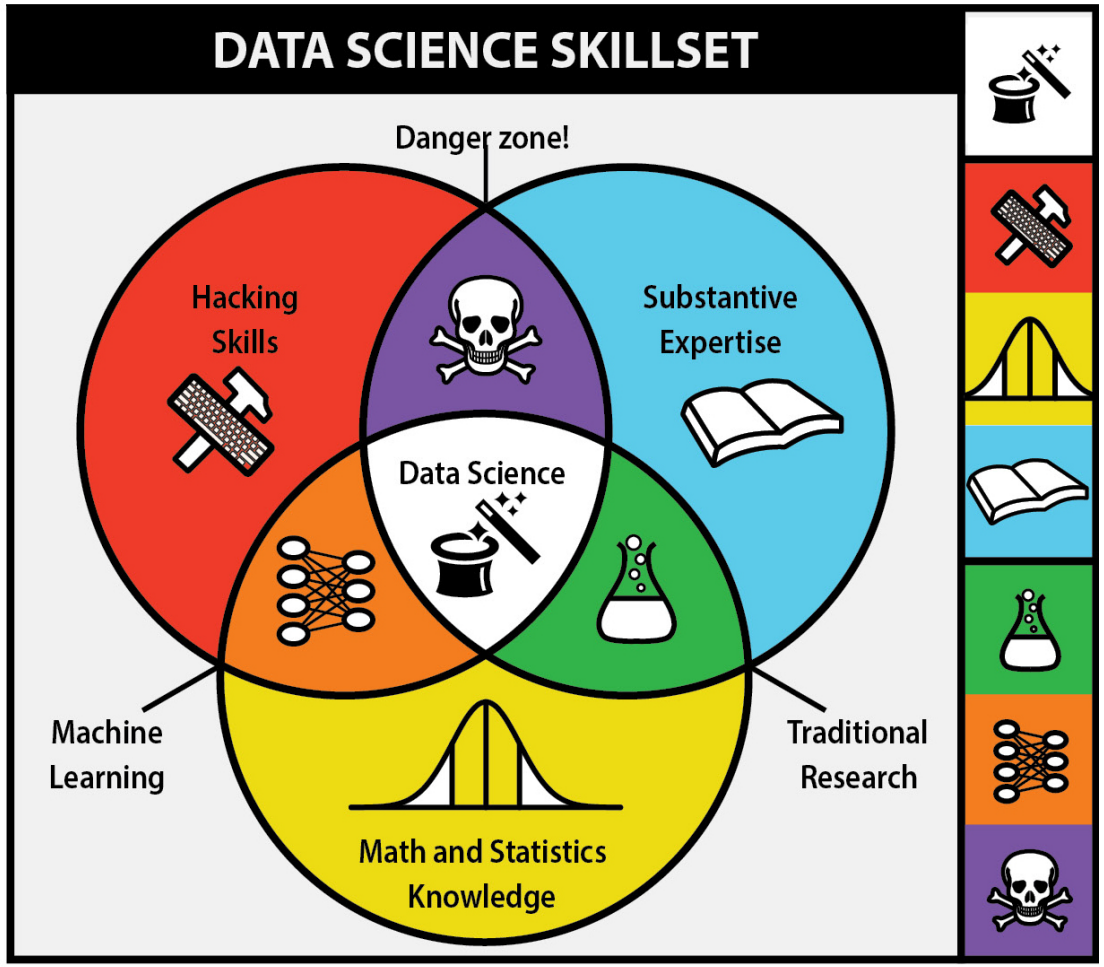
- 아무 모델이나
- 최신 모델이라

- **이 문제**와 **이** 어야 합니다.

- **RPA, Data P** 되어야 합니다.

주의 사항

- 도메인 지식만큼 수학적 지식, 구현 능력도 중요합니다.



데이터 과학자

컴퓨터로 구현하기

수학에서 근거 찾기

분야별 전문 지식

전통 연구

연구 과학자, 머신 러닝 공학자

함정 : 문제를 엉뚱하게 풀어요!