



# 기계학습과 확률

장하영

분당차병원 정보의학 컨퍼런스

2021-04-21

# ARTIFICIAL INTELLIGENCE

Early artificial intelligence  
stirs excitement.



## MACHINE LEARNING

Machine learning begins  
to flourish.



## DEEP LEARNING

Deep learning breakthroughs  
drive AI boom.



1950's

1960's

1970's

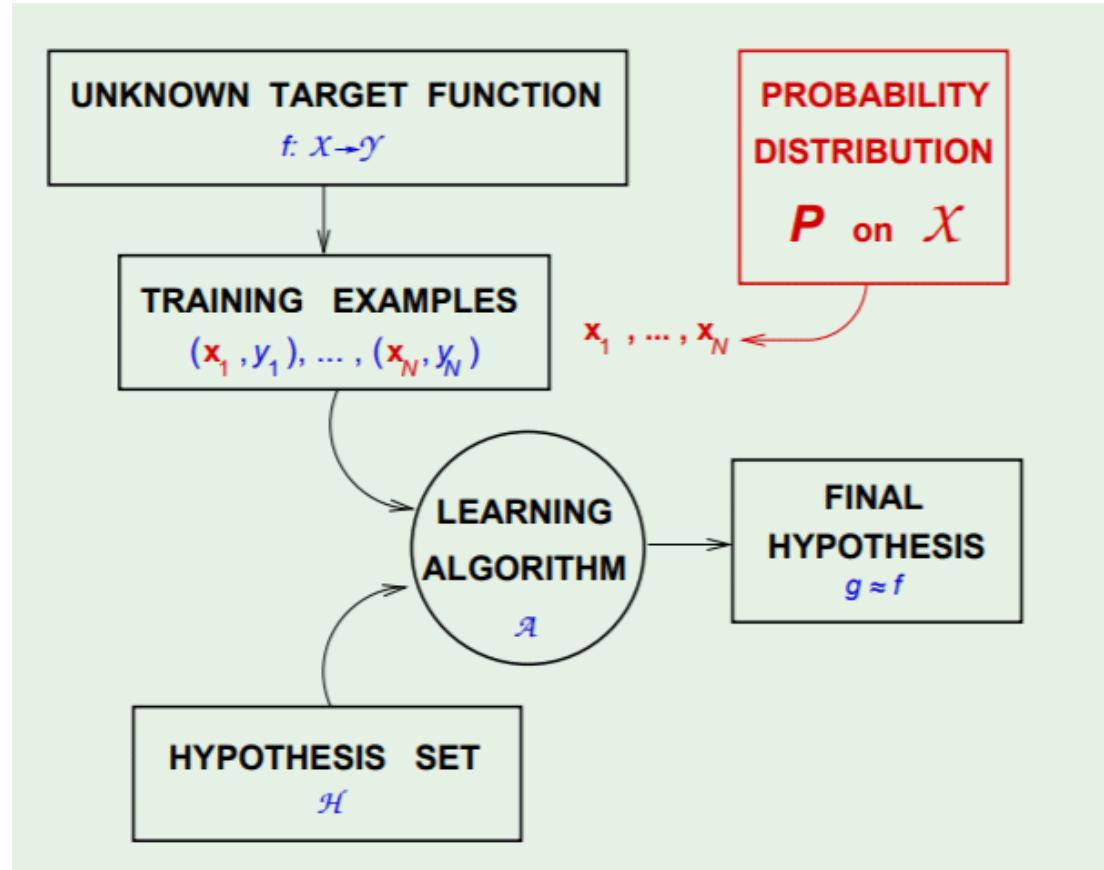
1980's

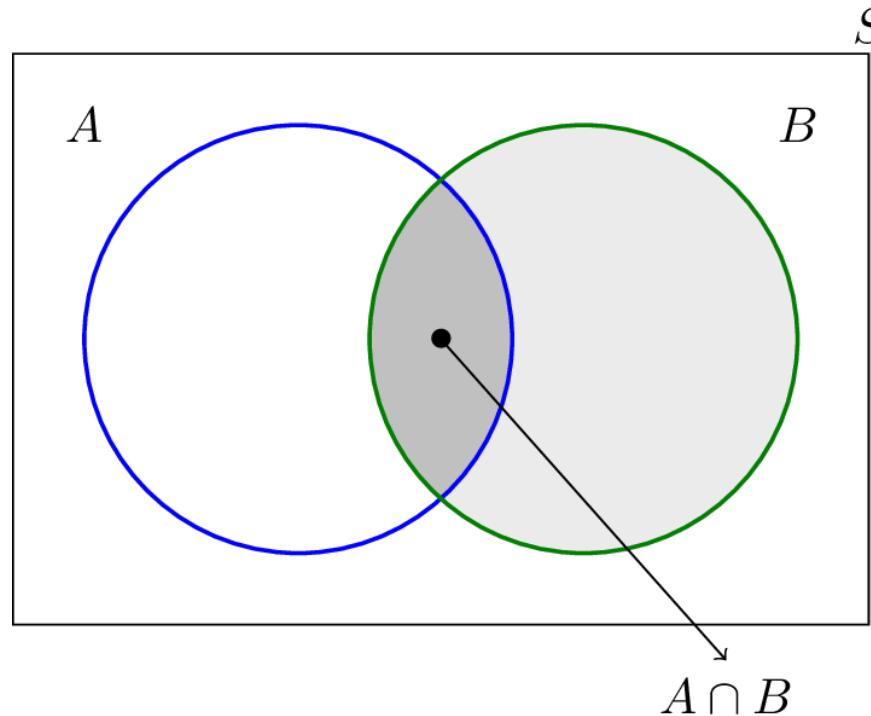
1990's

2000's

2010's

Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.



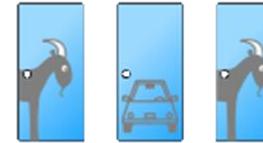
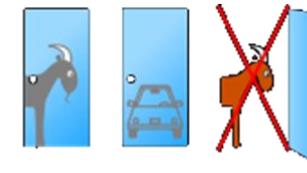


$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

# 몬티홀 문제



# 몬티홀 문제

Car hidden behind Door 1	Car hidden behind Door 2	Car hidden behind Door 3
Player initially picks Door 1		
		
Host opens either Door 2 or 3	Host must open Door 3	Host must open Door 2
		
Switching loses with probability 1/6	Switching loses with probability 1/6	Switching wins with probability 1/3
Switching loses with probability 1/3	Switching wins with probability 2/3	

- Frequentist
  - Ratio of frequencies as  $n \rightarrow \infty$
  - Repeated "identical" trials
  - Not applicable to single event or physical constant
- Bayesian
  - Degree of belief
  - Can be applied to single event or physical constant (even though these have unique truth)
  - Varies from person to person

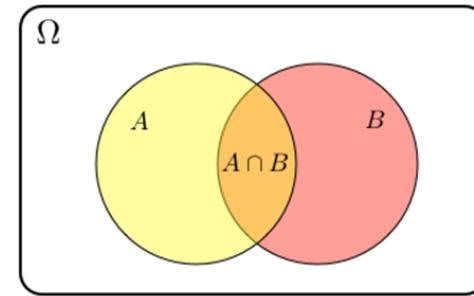
- 타율이 3할인 야구선수가 있다. 앞선 세번의 타석에서는 안타를 기록하지 못하고 네 번째 타석에 들어섰을 때 안타를 칠 확률은?
  - 0.3
  - 0.3보다 낮다
  - 0.3보다 높다

# Bayes' Theorem

- Bayes' Theorem is simply a consequence of the definition of conditional probabilities:

$$p(A | B) = \frac{p(A, B)}{p(B)} \rightarrow p(A, B) = p(A | B)p(B)$$

$$p(B | A) = \frac{p(A, B)}{p(A)} \rightarrow p(A, B) = p(B | A)p(A)$$



Thus  $p(A | B)p(B) = p(B | A)p(A)$

$$\rightarrow p(A | B) = \frac{p(B | A)p(A)}{p(B)}$$

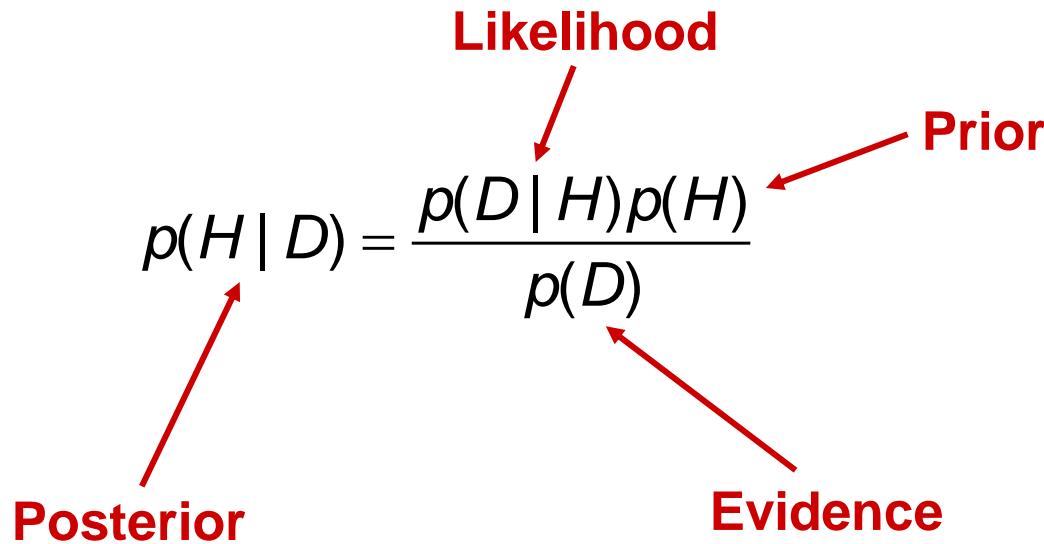
Bayes' Equation

# Bayes' Theorem

- Bayes' theorem is most commonly used to estimate the state of a hidden, causal variable H based on the measured state of an observable variable D:

$$p(H | D) = \frac{p(D | H)p(H)}{p(D)}$$

**Likelihood**  
**Prior**  
**Posterior**      **Evidence**

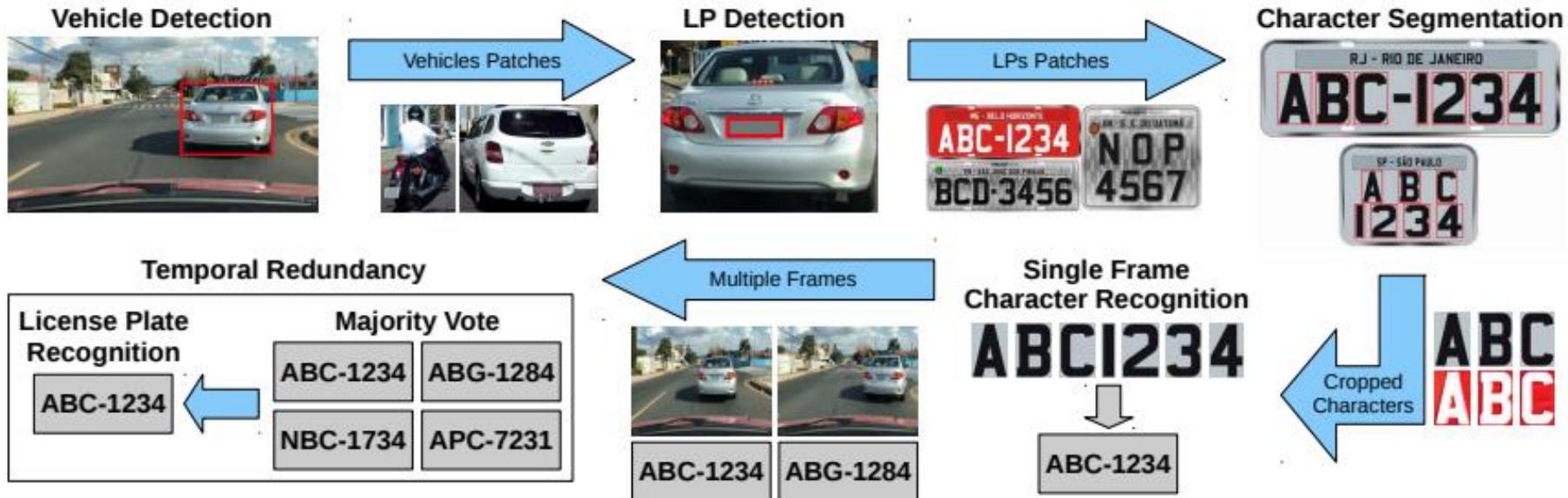


The diagram illustrates the components of Bayes' Theorem. At the top, the term "Likelihood" is written in red, with a red arrow pointing down to the term "p(D | H)" in the numerator of the equation. To the right of the equation, the term "Prior" is written in red, with a red arrow pointing left to the term "p(H)" in the numerator. Below the equation, the term "Posterior" is written in red, with a red arrow pointing up to the term "p(H | D)" on the left. To the right of the equation, the term "Evidence" is written in red, with a red arrow pointing down to the term "p(D)" in the denominator.

- Your cancer test result is positive
- Sensitivity of the test is 95%
- Attack rate of the cancer is 0.1%

$$\begin{aligned} p(\text{cancer}|\text{positive}) &= \frac{p(\text{positive}|\text{cancer})p(\text{cancer})}{p(\text{positive})} \\ &= \frac{p(\text{positive}|\text{cancer})p(\text{cancer})}{p(\text{positive}|\text{cancer})p(\text{cancer}) + p(\text{positive}|\neg\text{cancer})p(\neg\text{cancer})} \\ &= \frac{0.95 \times 0.001}{0.95 \times 0.001 + 0.05 \times 0.999} = \frac{0.00095}{0.0509} \approx 0.01866 \end{aligned}$$

# 딥러닝을 이용한 번호판 인식 과정



# 머신러닝 프로덕트 개발 단계와 비용

비전문가의 예상

Expectation



Google의  
전문가 산정 비율

Reality



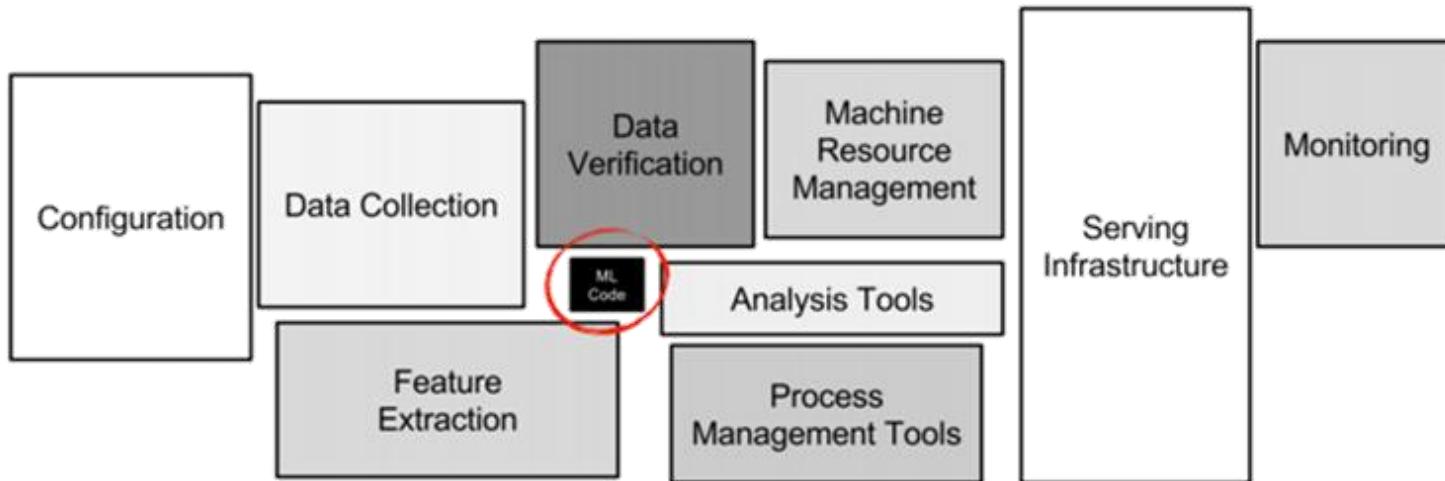
0.25

0.5

0.75

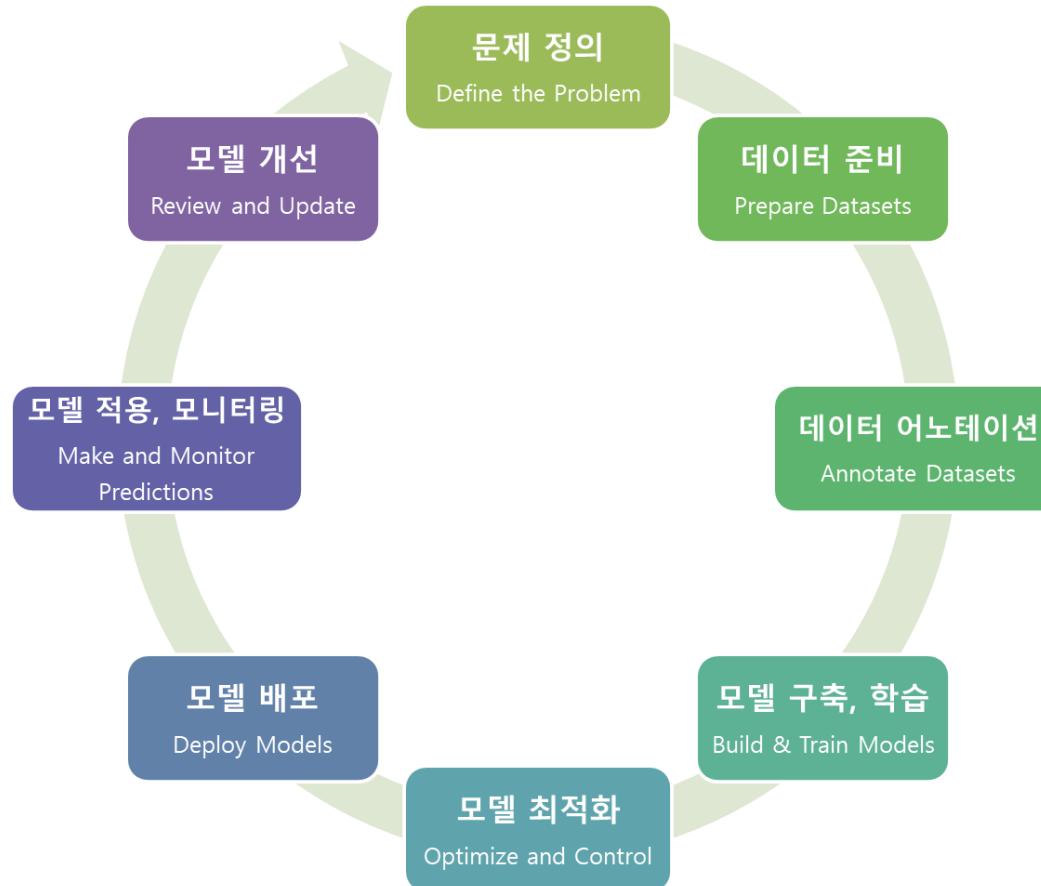
.1

- Defining KPI's  
KPI 정의
- Collecting data  
Data 수집
- Building infrastructure  
인프라 구축
- Optimizing ML algorithm  
머신러닝 알고리즘 최적화
- Integration  
제품 통합 및 출시



## Machine Learning: The High-Interest Credit Card of Technical Debt

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov,  
Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young



## 레이블

## 구축

## 학습 및 튜닝

## 배포 및 관리

**Amazon SageMaker Ground Truth**  
Build and manage training data sets

**Amazon SageMaker Studio**  
Integrated development environment (IDE) for machine learning

**Amazon SageMaker Autopilot**  
Automatically build and train models

**Amazon SageMaker Model Monitor**  
Automatically detect concept drift

**Amazon SageMaker Notebooks**  
One-click notebooks with elastic compute

**Amazon SageMaker Experiments**  
Capture, organize, and search every step

**Amazon SageMaker Neo**  
Train once, deploy anywhere

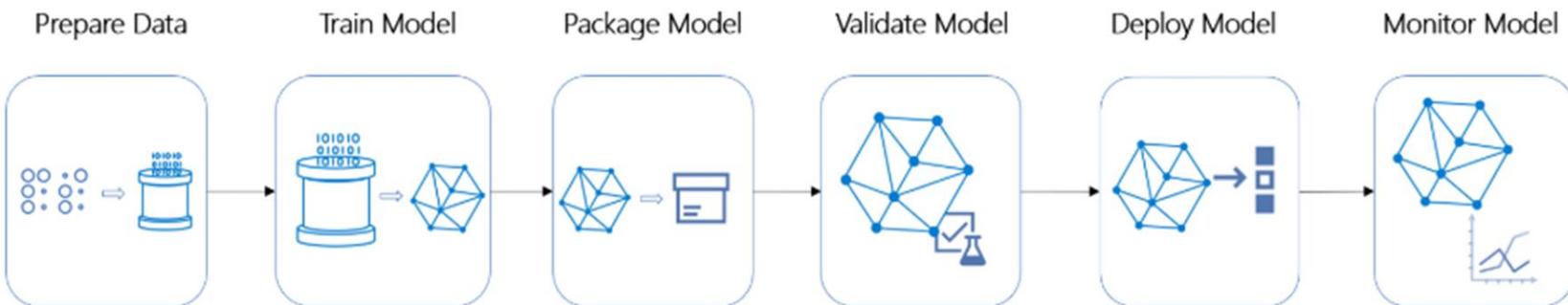
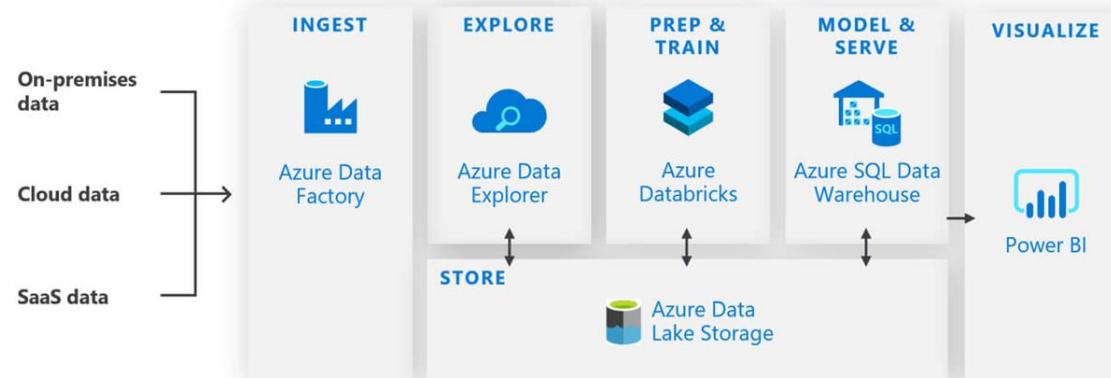
**AWS Marketplace**  
Pre-built algorithms and models

**Amazon SageMaker Debugger**  
Debug and profile training runs

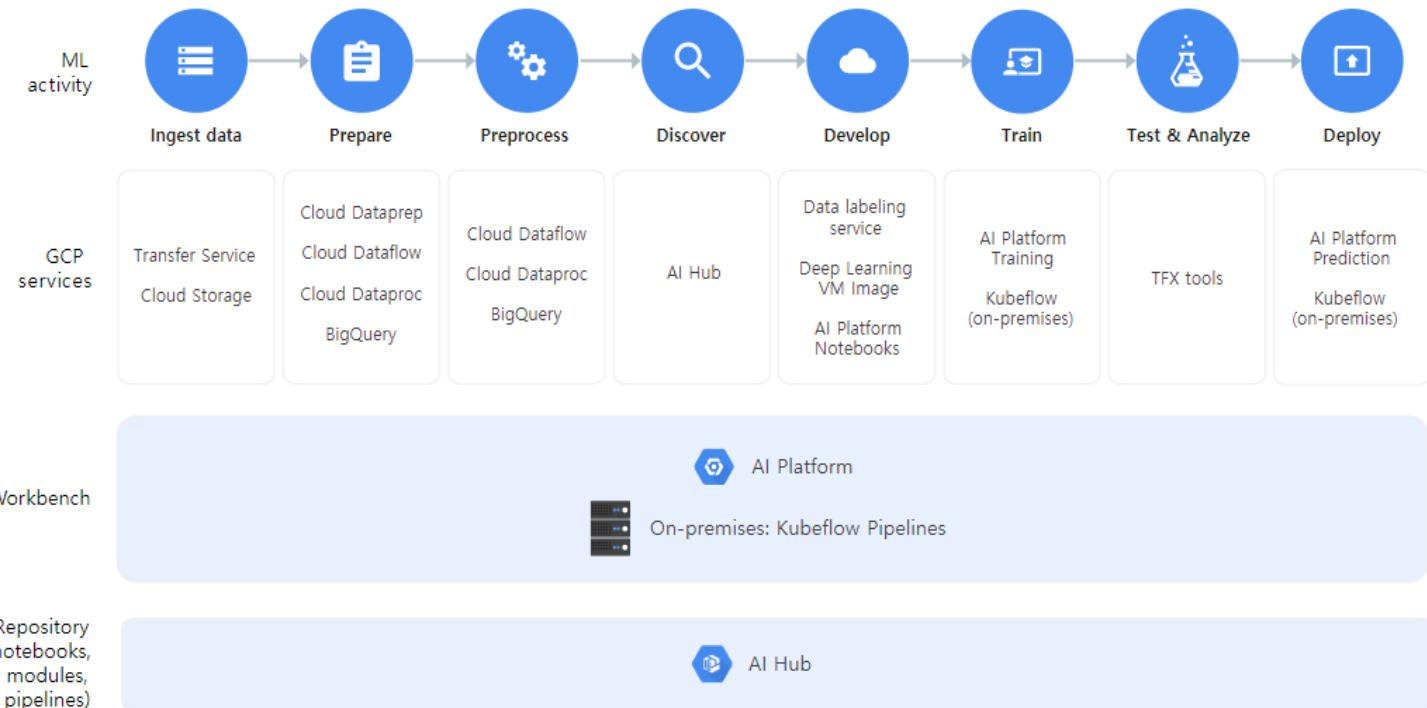
**Amazon Augmented AI**  
Add human review of model predictions

**Automatic Model Tuning**  
One-click hyperparameter optimization

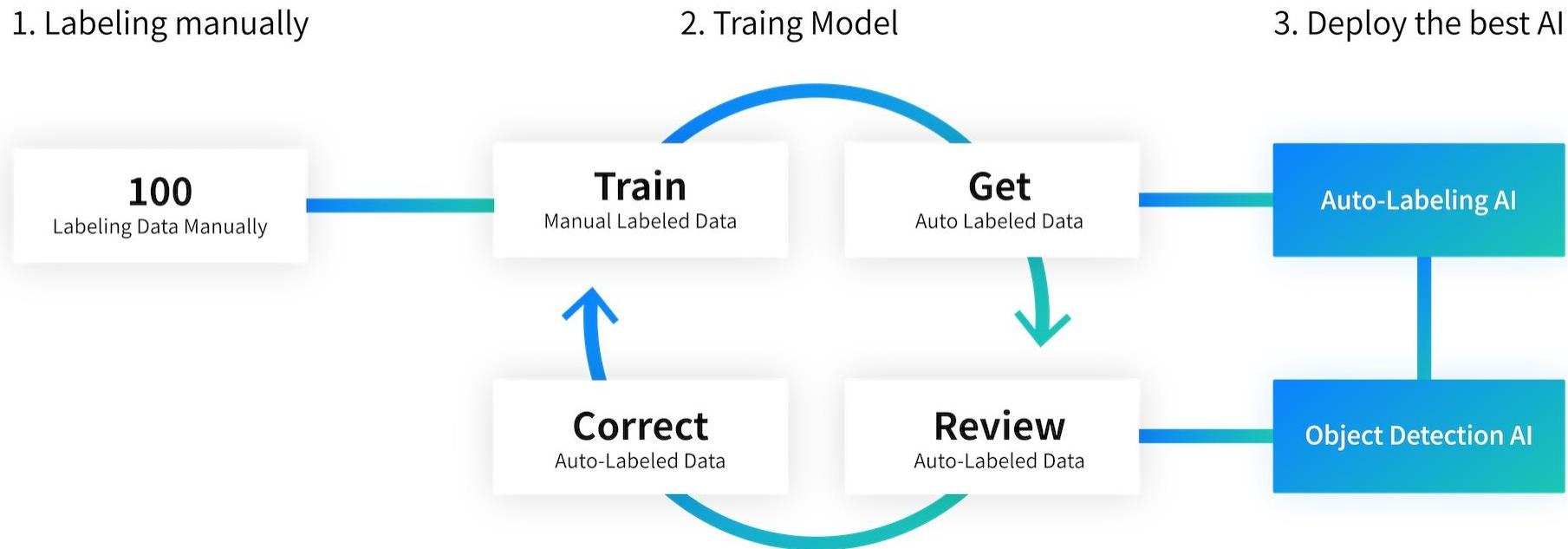
# Azure ML Studio

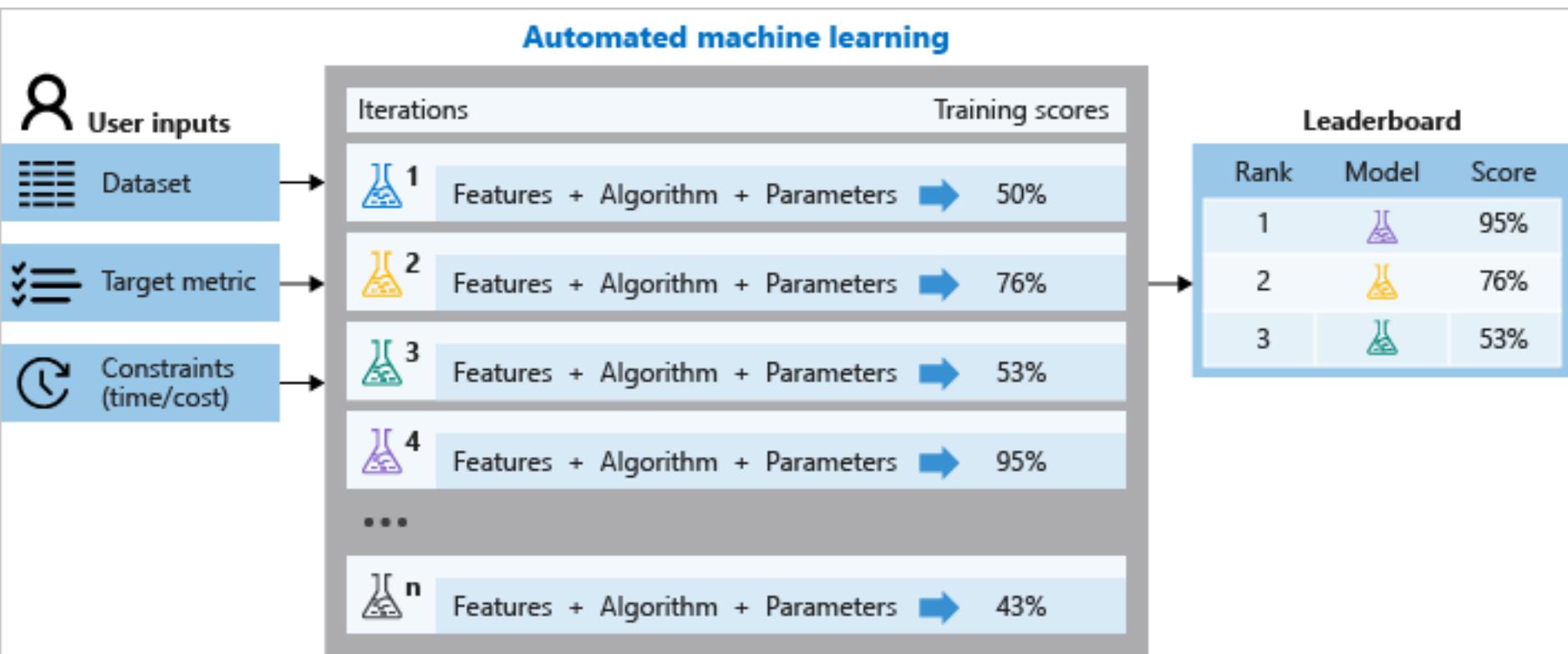


## 머신러닝 개발: 엔드 투 엔드 주기

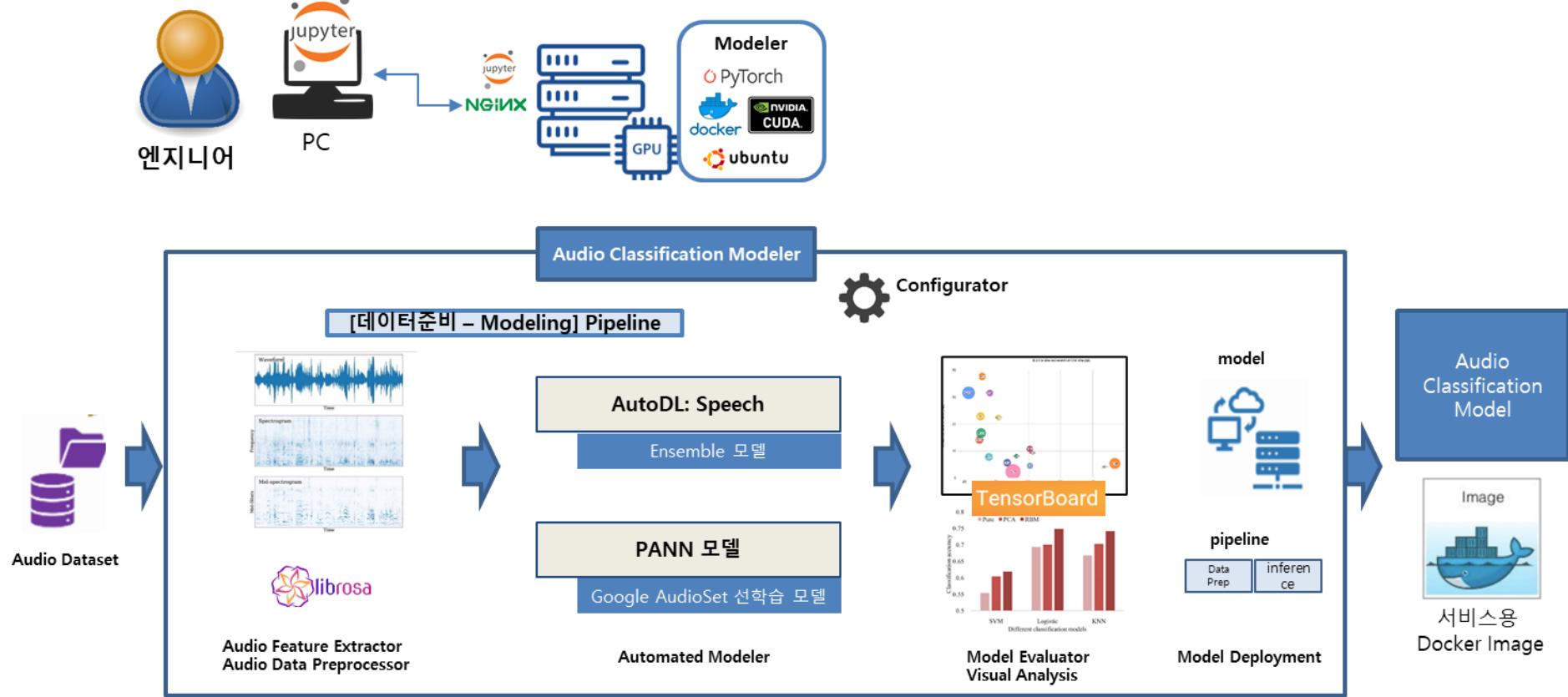


# Auto Labeling

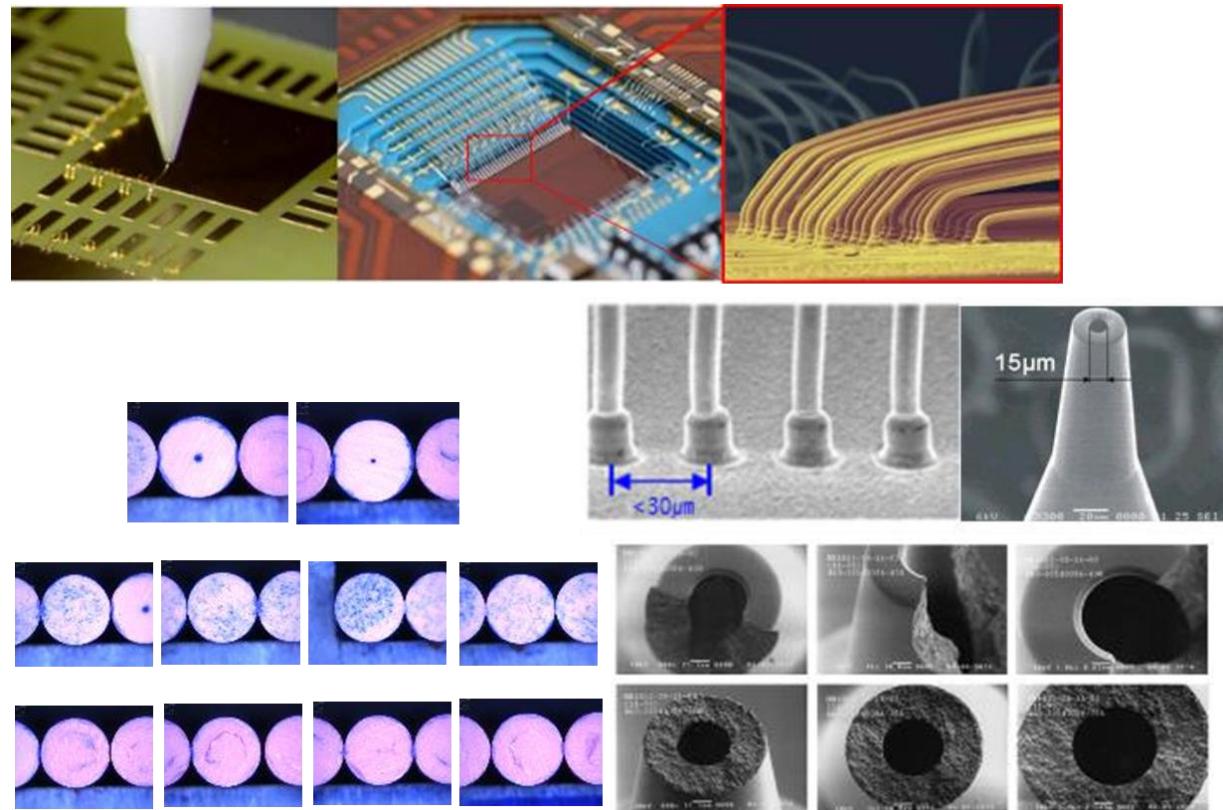
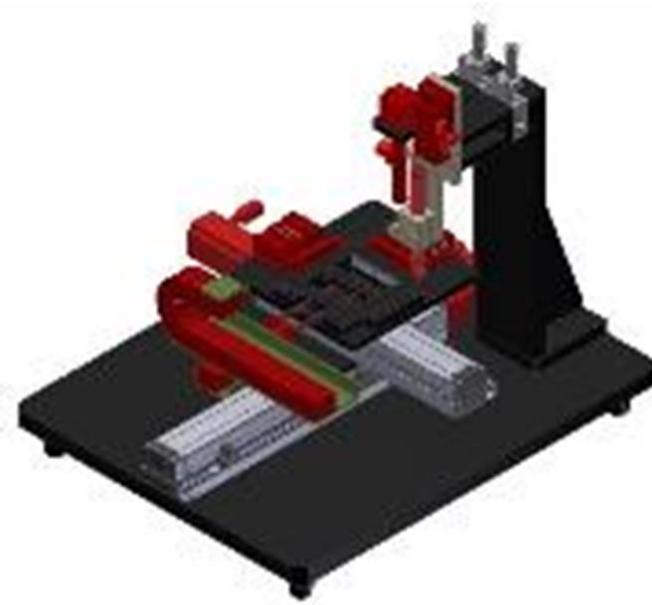




# Use Case



# Use Case





Bd 138, Seoul National Univ.  
Gwanak-Gu, Seoul 08826,  
South Korea



+82-2-872-5127



[surromind.ai](http://surromind.ai)  
[contact@surromind.ai](mailto:contact@surromind.ai)

# Thank You

 SURROMIND