

울어도돼 사실 **MNIST**는 없거든.

넥슨코리아
황준식



저는,

넥슨에서 **매크로/해킹**잡는
데이터분석가

#끝없는해공격
#끝없는벙커수리
#살려줘

3줄요약

#머신러닝 in 실무

#엄청 어렵다!

#징징

MNIST?

Modified NIST dataset

ML부적합

NIST: 손글씨 데이터셋

train set by 통계청직원들

test set by 고교학생들

>

크기/위치
전처리

by Yann LeCun 등

>

ML 적합

MNIST

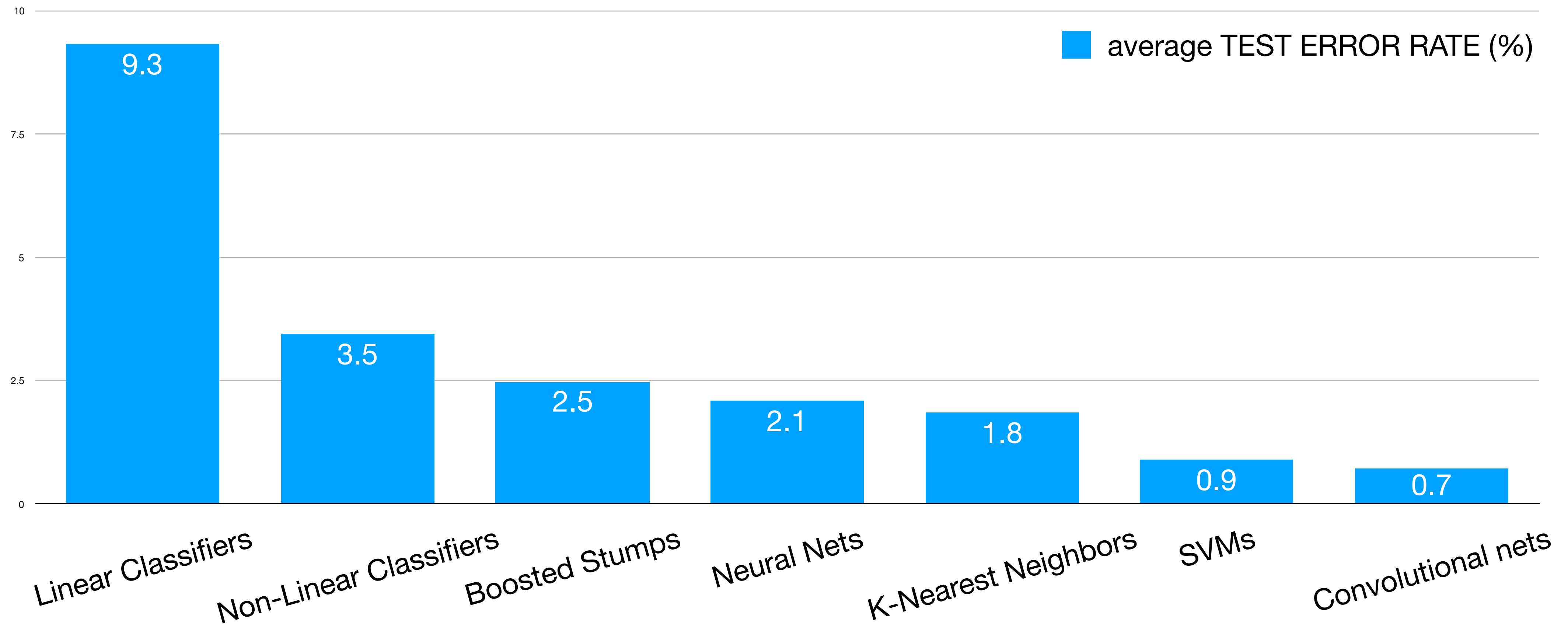
직원샘플

학생샘플

직원샘플

학생샘플

MNIST: born for ML



출처: <http://yann.lecun.com/exdb/mnist/>

MNIST for ML beginners

```
## MNIST 데이터셋 임포트
from tensorflow.examples.tutorials.mnist import input_data

## 데이터셋 경로 및 레이블 형태 지정
mnist = input_data.read_data_sets("MNIST_data/", one_hot=True)

## 텐서플로우 세션 내에서 트레이닝 데이터 배치 불러오기
for _ in range(1000):
    batch_xs, batch_ys = mnist.train.next_batch(100)
```

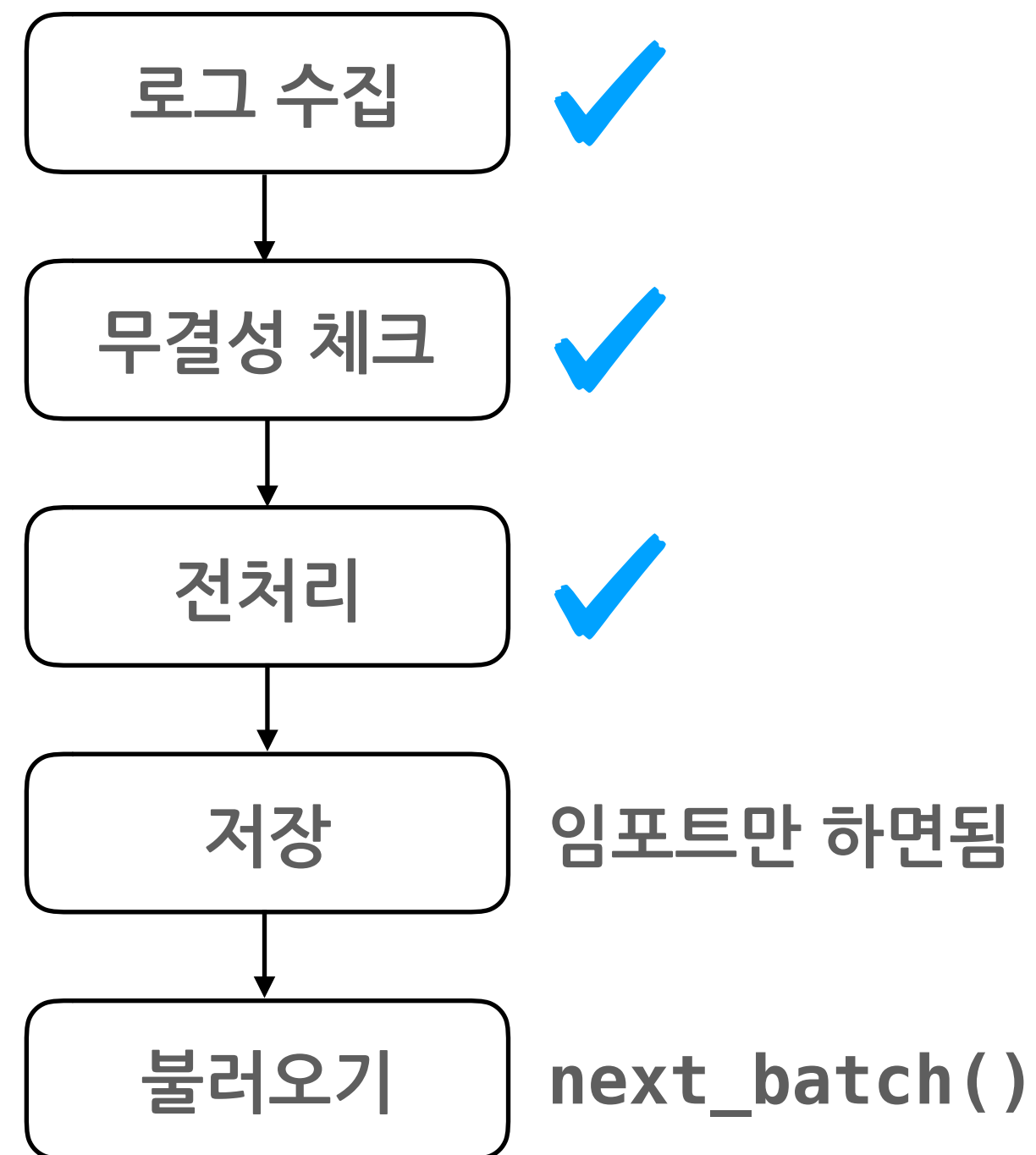
출처: https://www.tensorflow.org/get_started/mnist/beginners

MNIST, 그런거 없다 아들

- 전처리의 문제
- 평가의 문제
- 문제의 문제

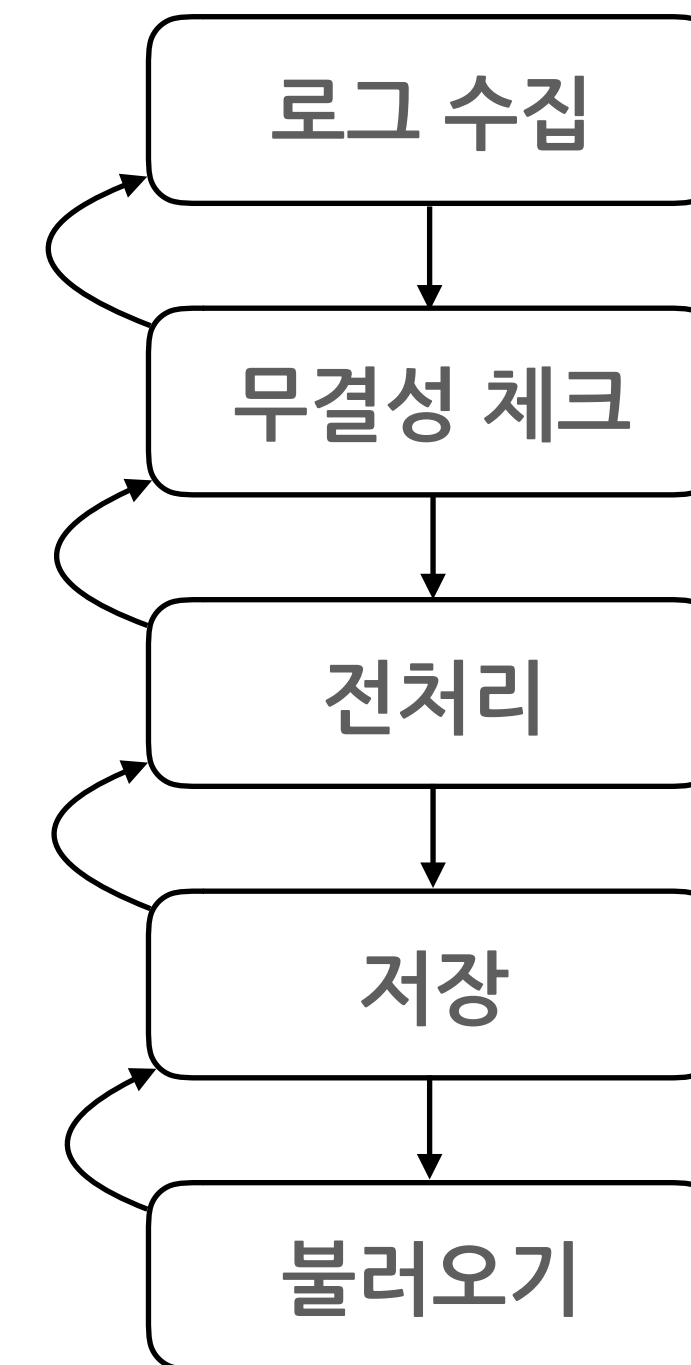
전처리의 문제

MNIST



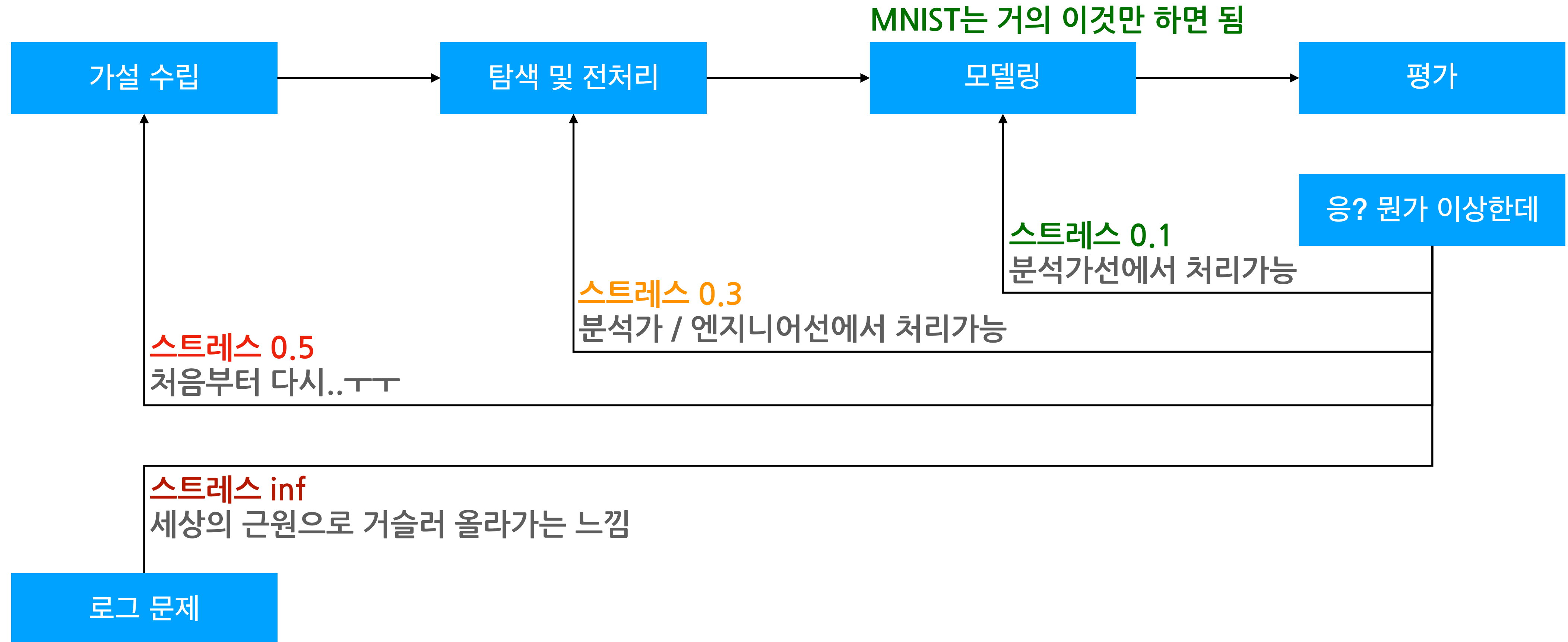
```
from tensorflow.examples.tutorials.mnist import input_data
```

Reality



로그규격변경 / 커뮤니케이션오류 / 시스템오류 등
한번하고 끝나는 개념이 아님

전처리의 문제



전처리의 문제 in GAME

하나라도 로그가 잘못되거나
전처리가 제대로 안되면 **다시**.

캐릭터 기본 정보
캐릭터 능력 스탯
캐릭터 업적

던전 내 액션 기록
액션별 데미지 기록
던전 후 보상 기록
던전 후 경험치 기록

클라이언트 변조기록
서버 접속 기록
신고 유무

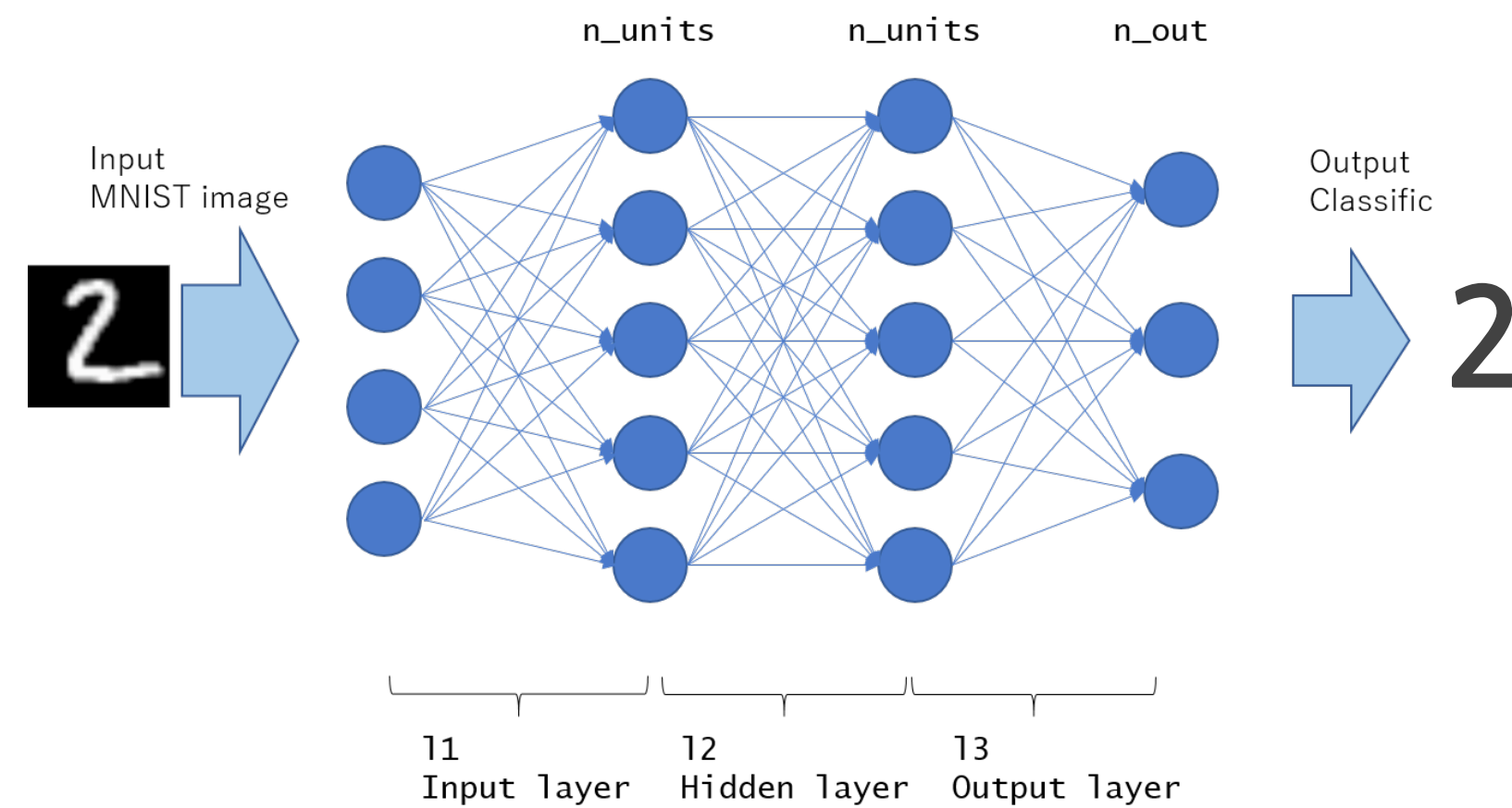
소지품 일람
창고 소지품 일람
경매장 이용내역
우편함 이용내역
골드 내역
아이템 내역

친구 관계 정보
길드 가입 정보
파티 플레이 여부



평가의 문제

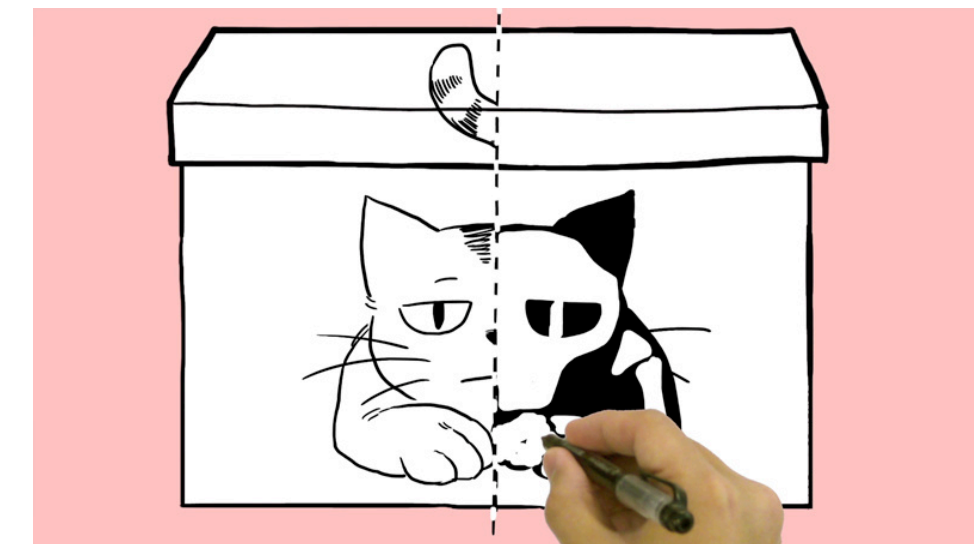
MNIST



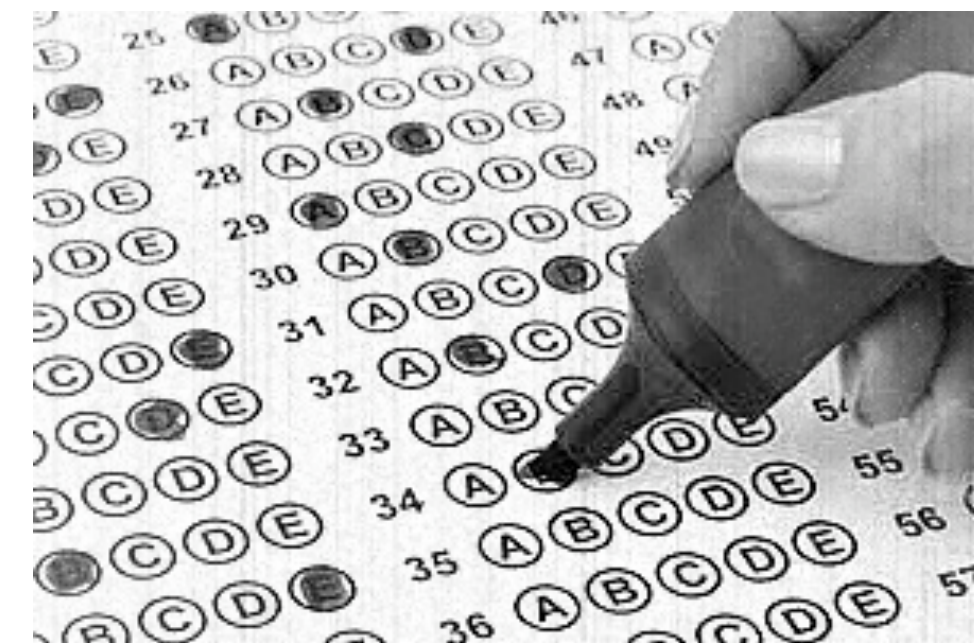
맞으면 맞는거고 틀리면 틀리는
누가봐도 같은 평가가 나옴 **평가에 이견이 없음**

정답지가 있고, 없아해도 노가다하면 됨

Reality



Ban 될때까지 거의 알 수 없음 / **항의가 들어와야 확정됨**



정답지가 없거나 반쪽만 있음 / 만드는데 **비용이 생각보다 큼**

평가의 문제 in GAME

데이터분석팀

이 유저는 저희 머신러닝 알고리즘에 의해 봇으로 검출되었습니다

게임운영팀

더 정확한 정황없이
제재할 수 없어요

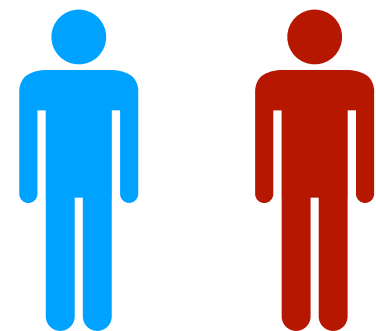
높은 오제재 리스크

제재 사유를 정확히, 직관적으로
이해하기 어려워요

직관적 이해 어려움

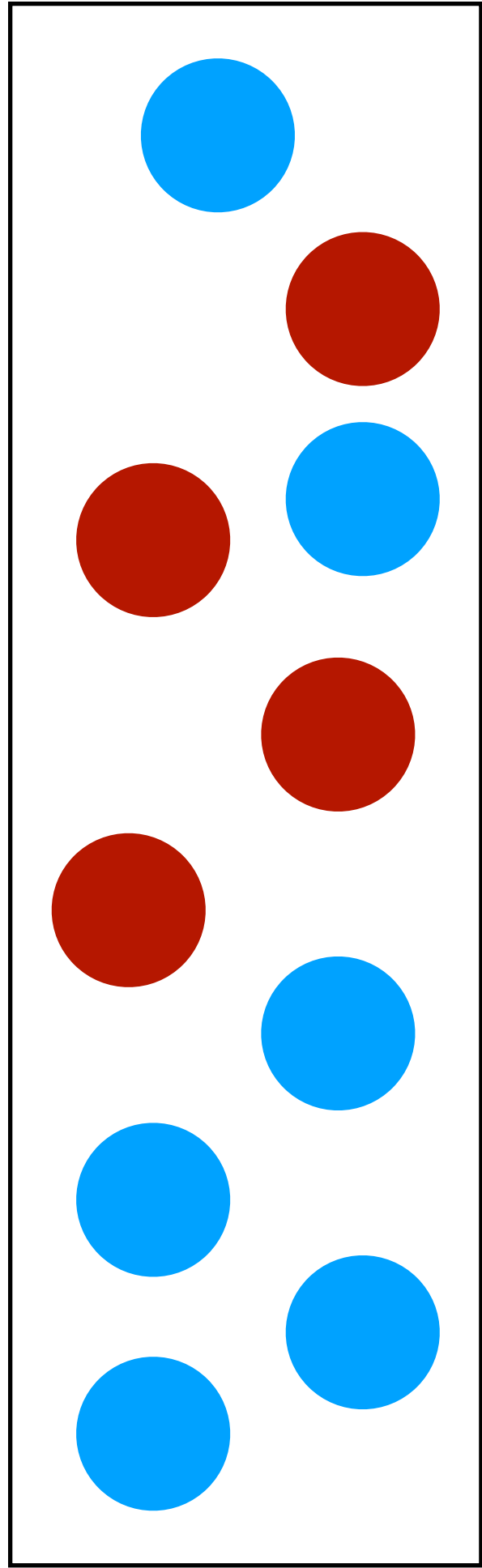
다른 로그를 보니까
정상 유저인 듯 합니다

모델링의 한계
/ 도메인 지식

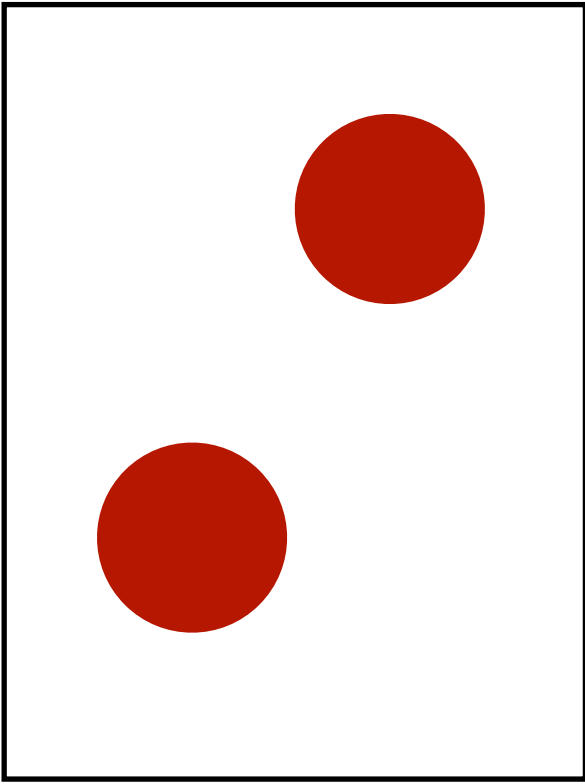


평가의 문제 in GAME

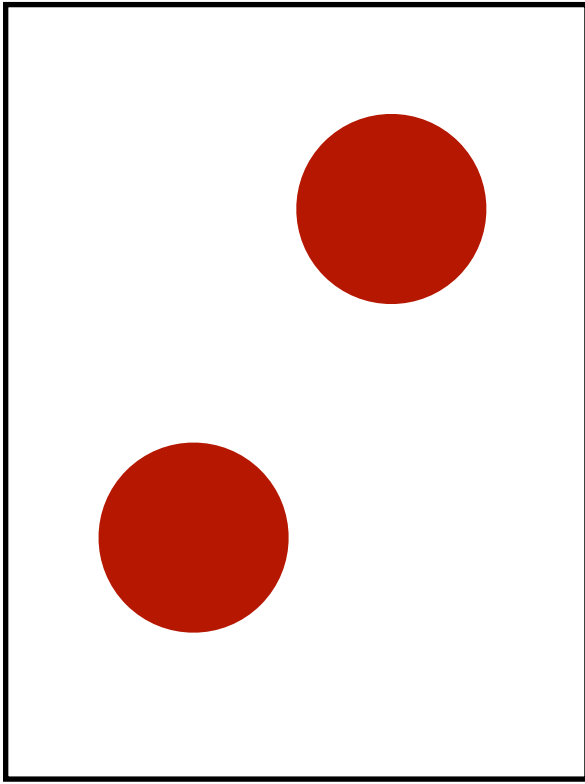
전체 유저풀



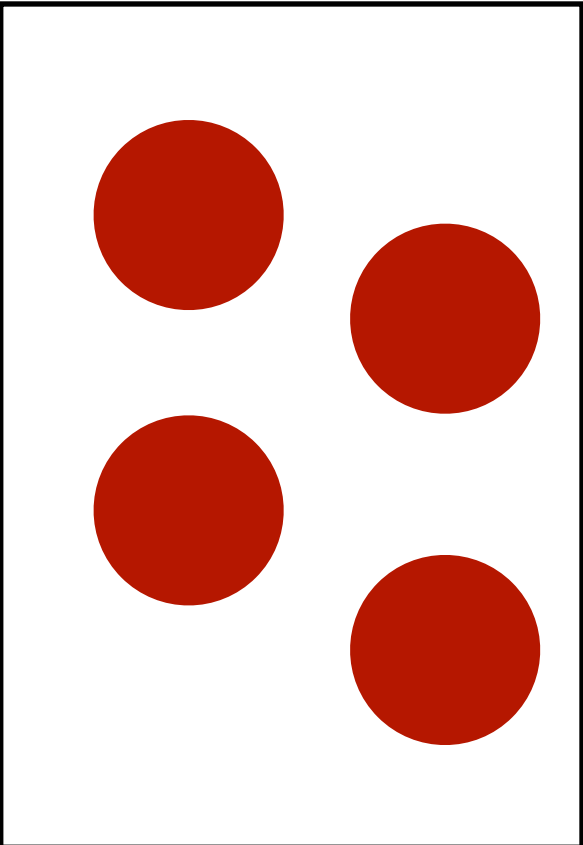
과거 제재기록



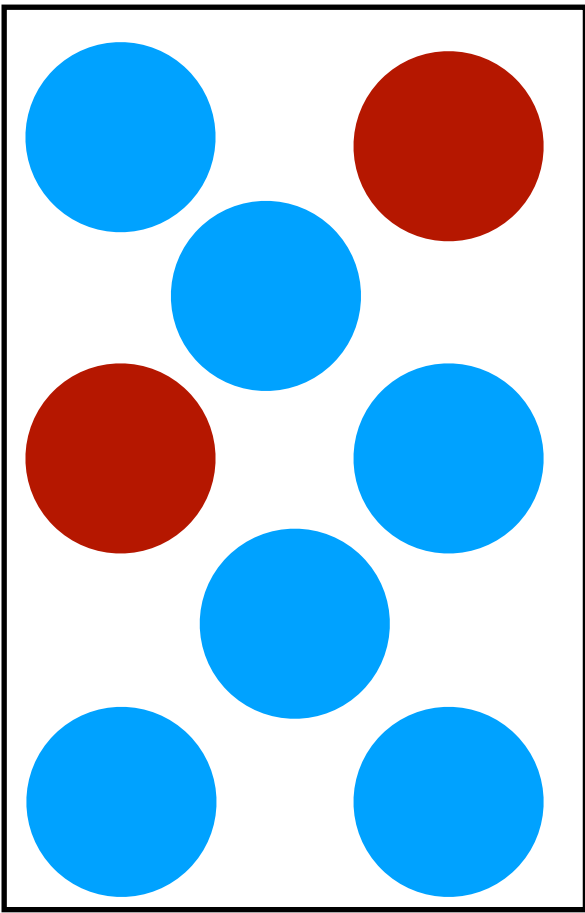
과거 제재기록



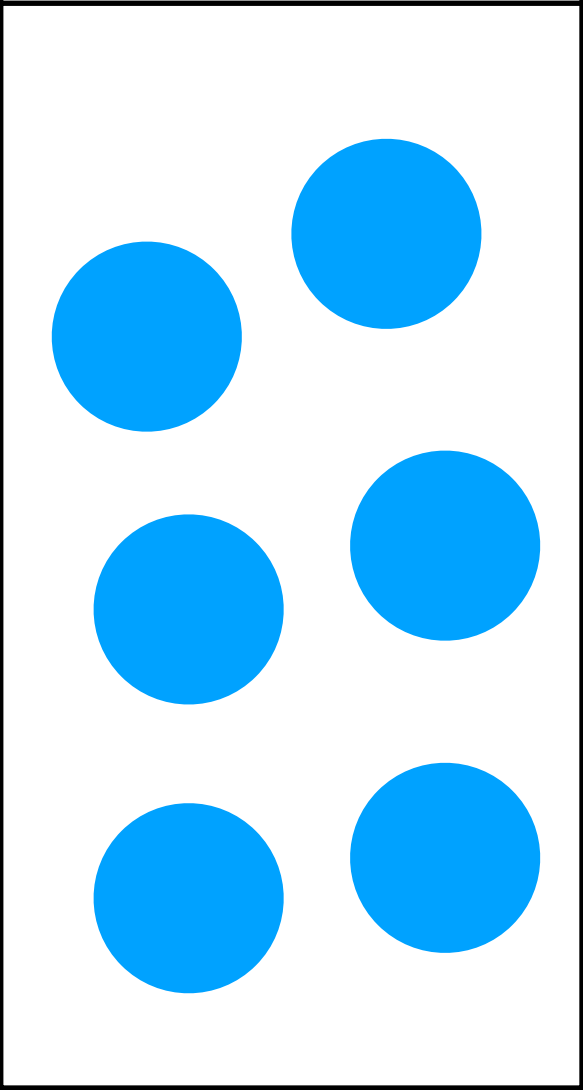
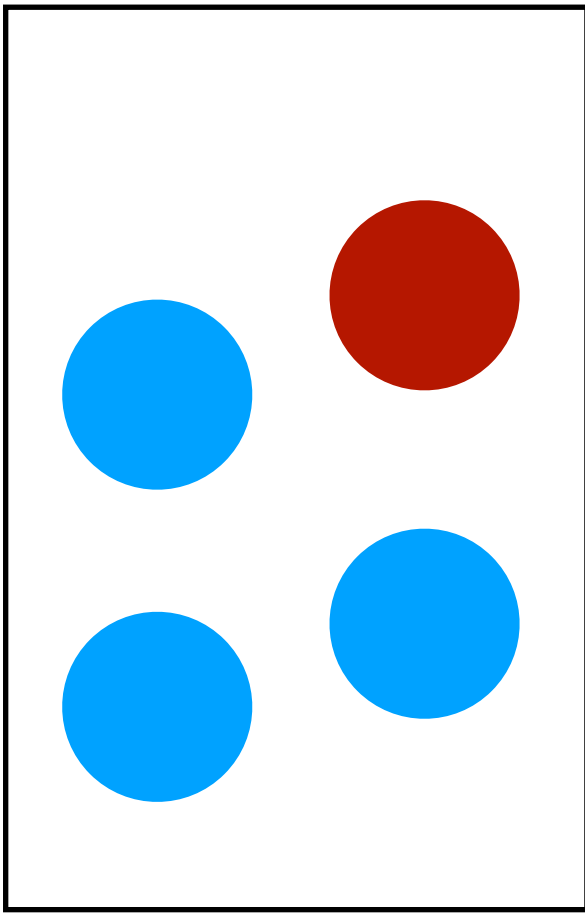
원하는 정답지



나머지

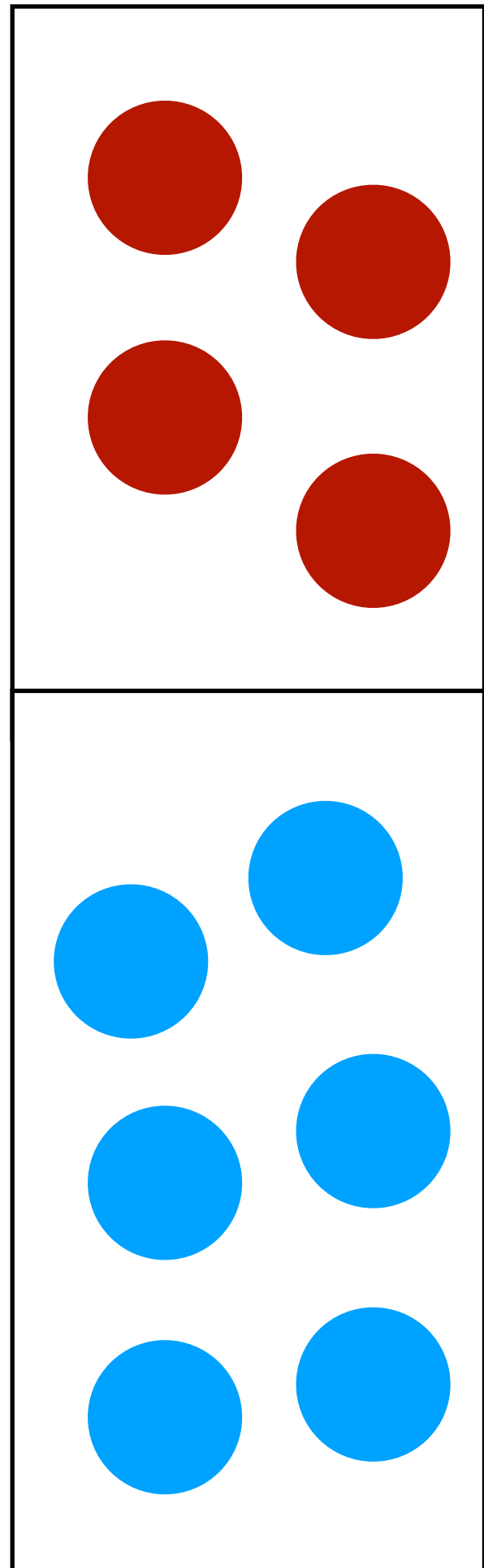


구매유저

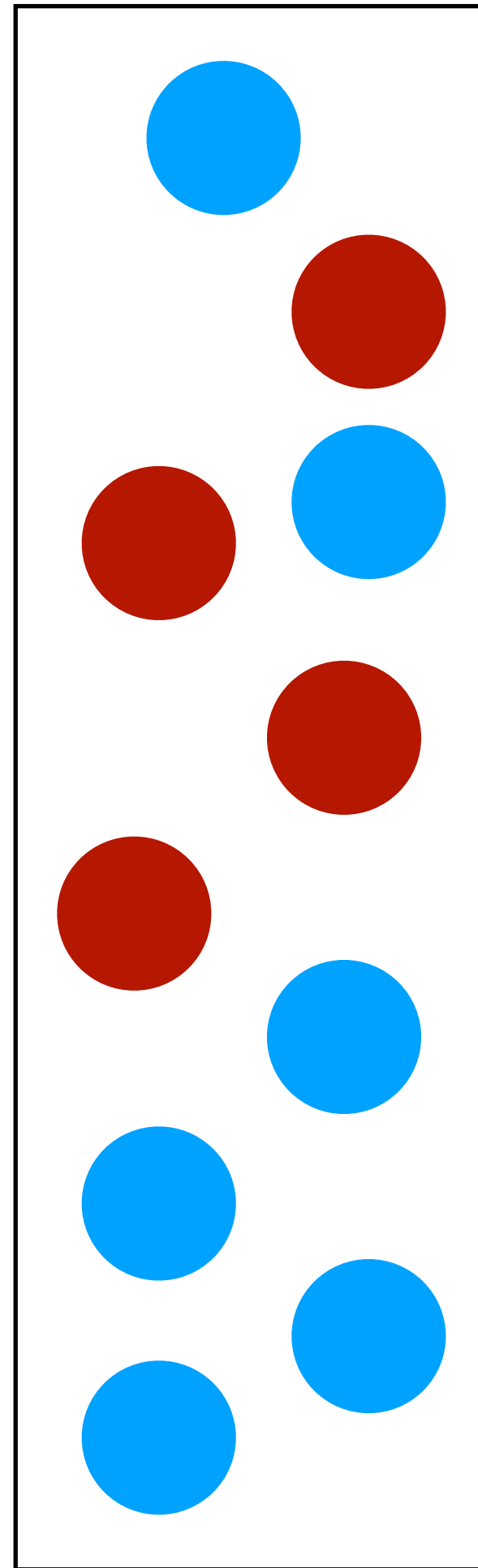


평가의 문제 in GAME

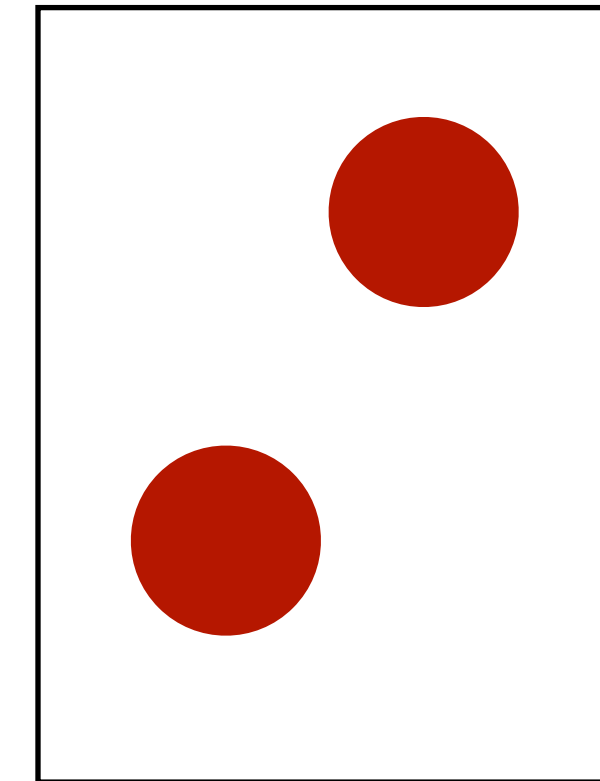
원하는 정답지



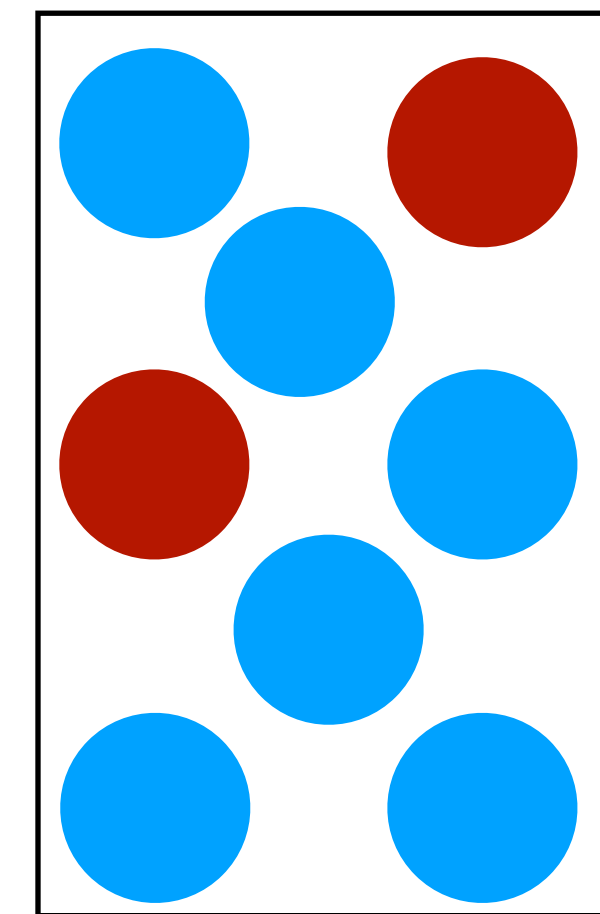
전체 유저풀



과거 제재기록



나머지

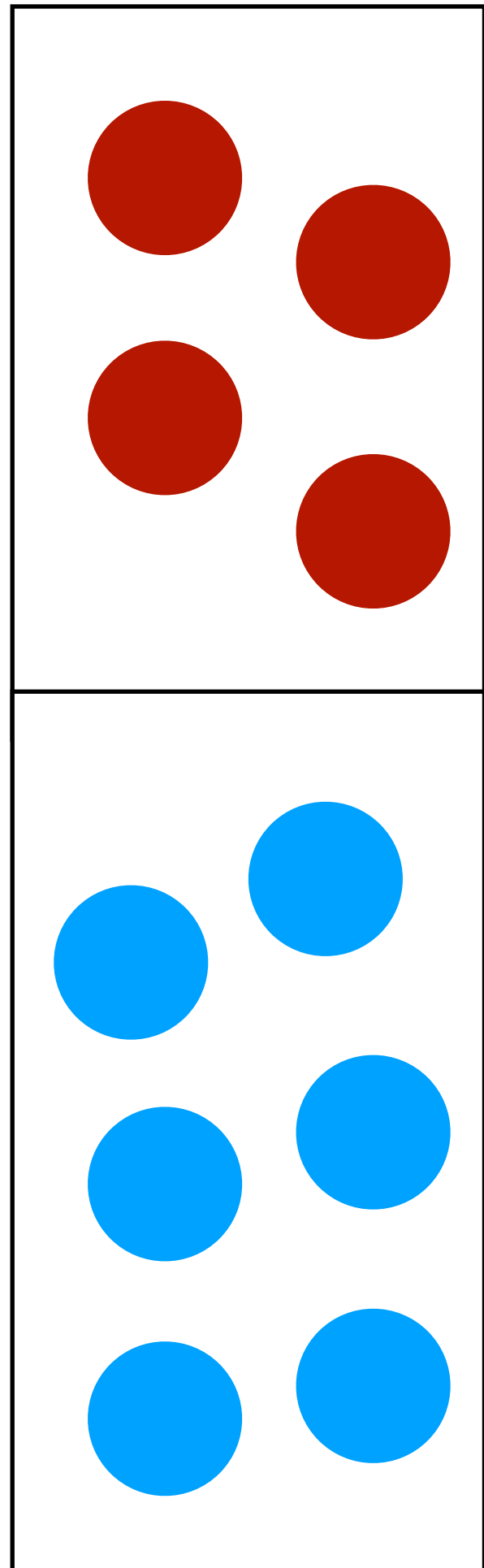


▶ 제재기록
쓰면 되는거
아님? ▶

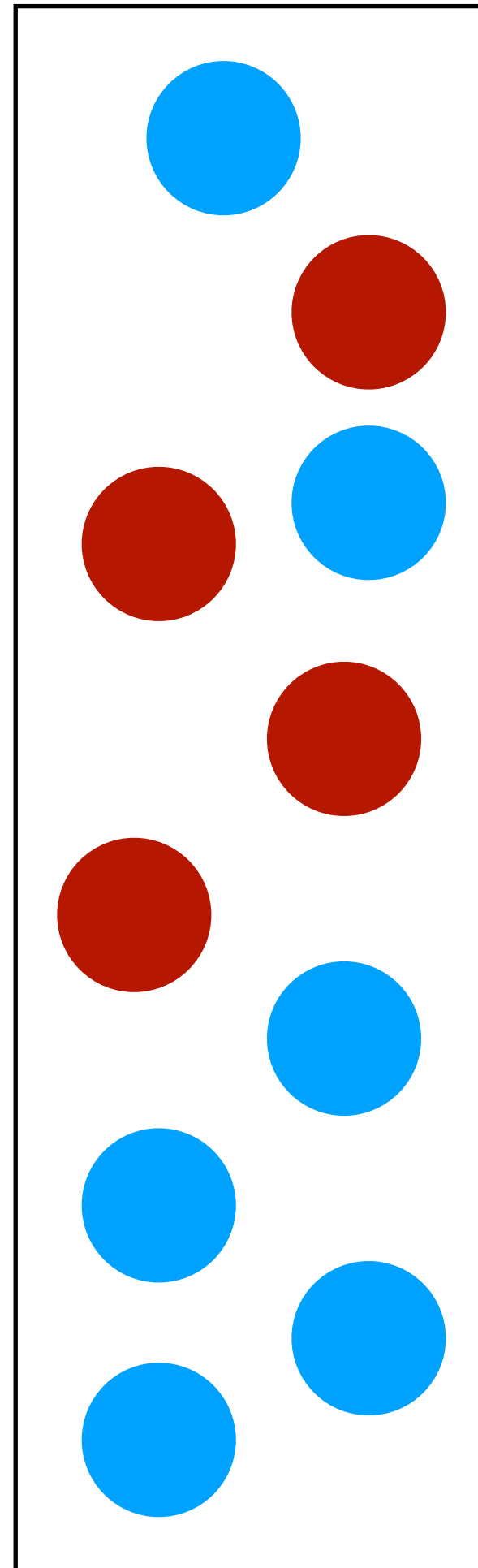
보수적 제재때문에
안걸린 봇 존재함

평가의 문제 in GAME

원하는 정답지

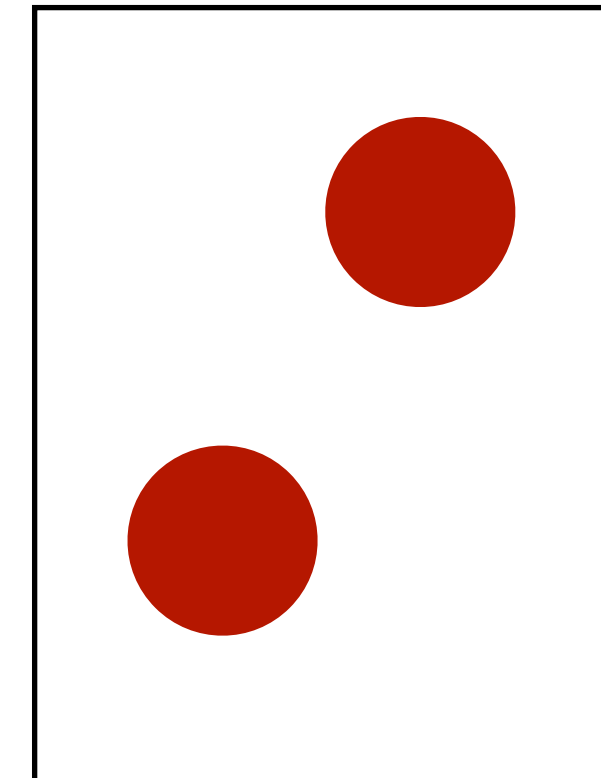


전체 유저풀

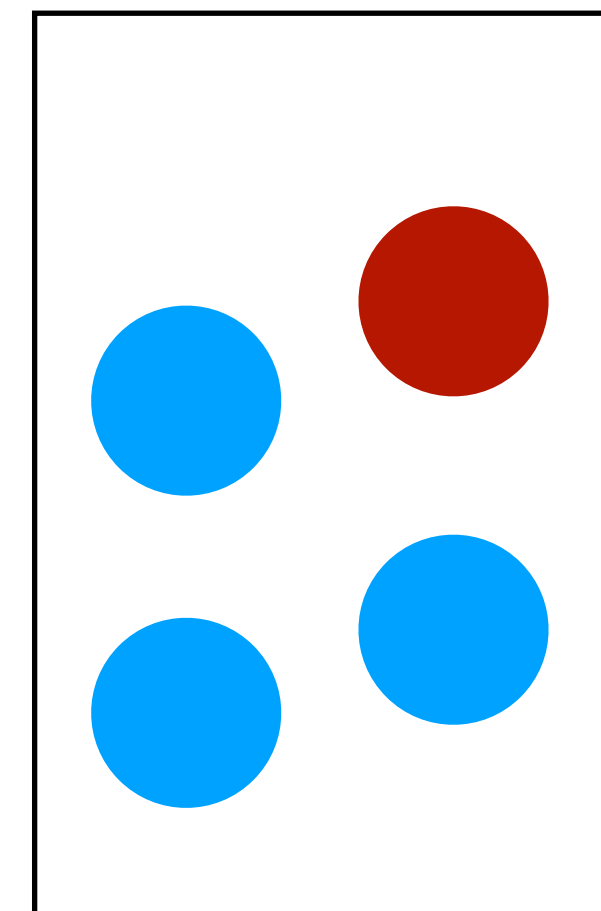


▶ 구매기록
쓰면 되는거
아님? ▶

과거 제재기록



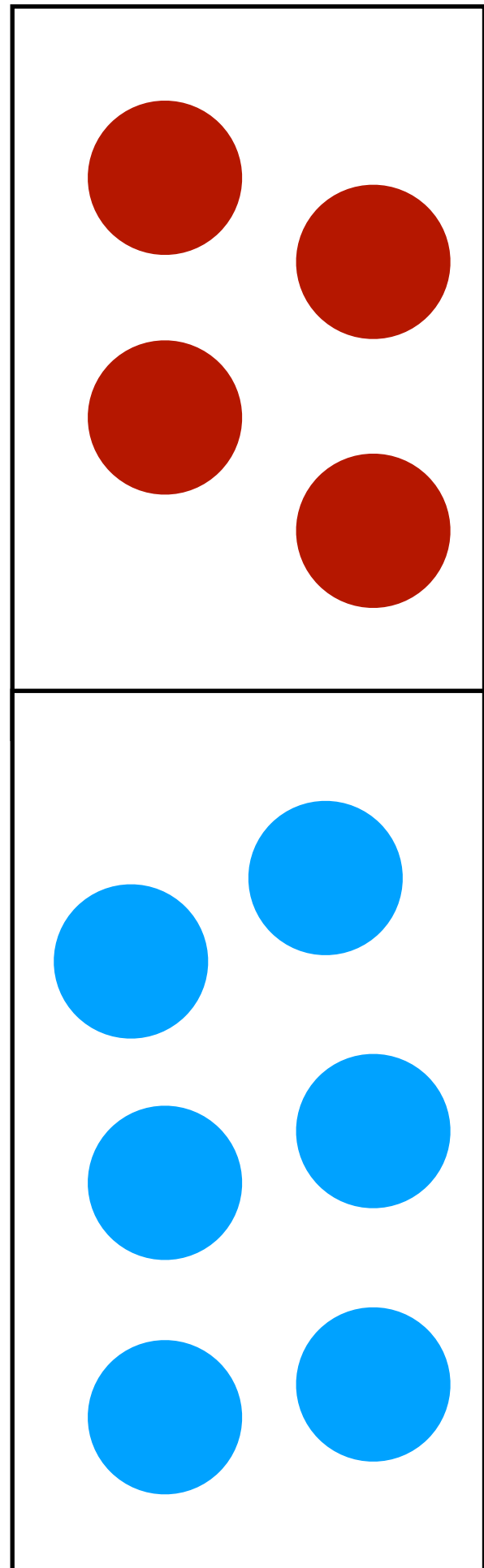
구매유저



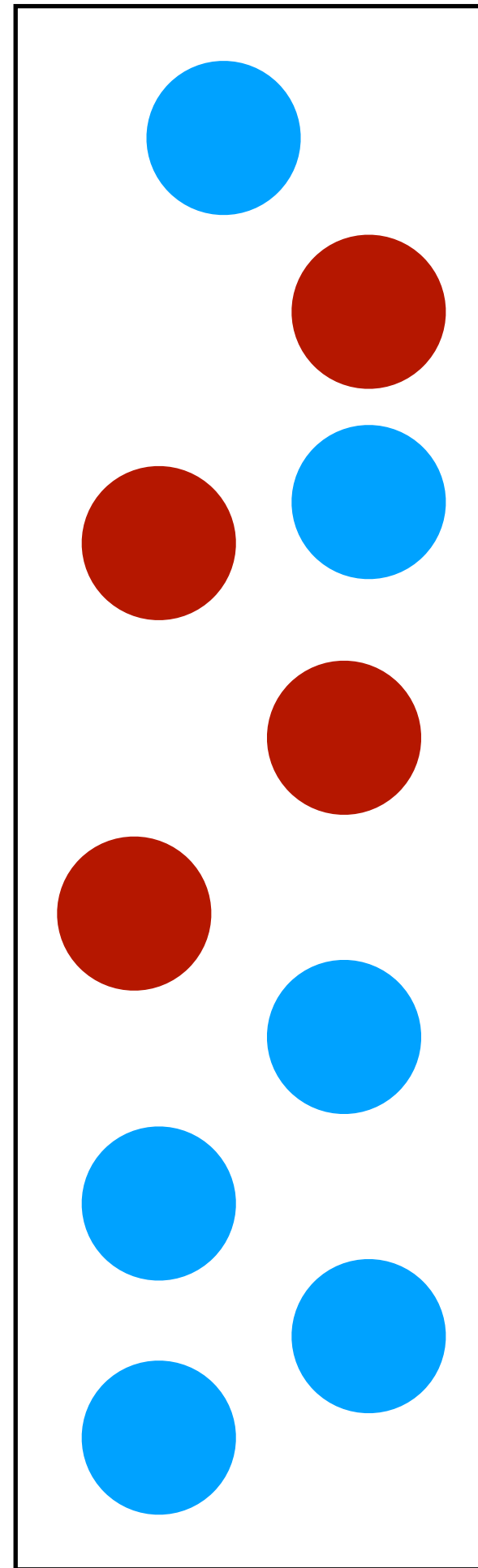
요새는 일부러
약간 과금하기도

평가의 문제 in GAME

원하는 정답지

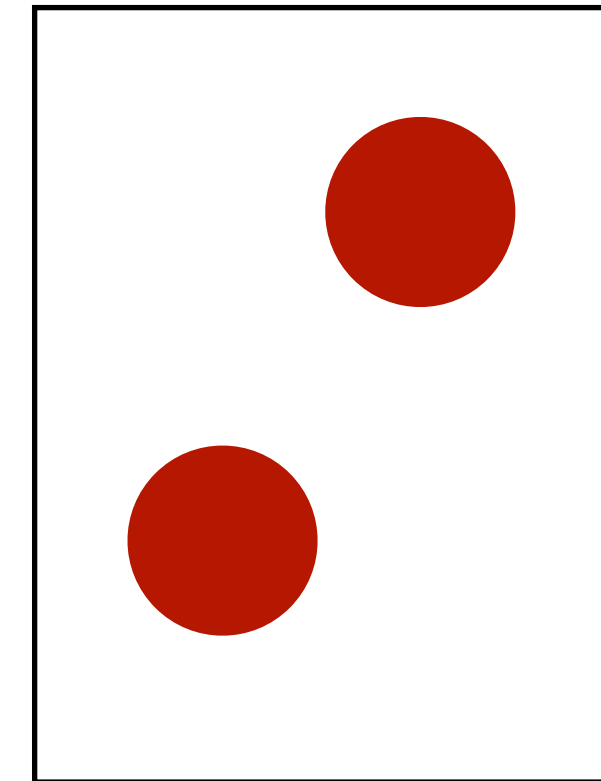


전체 유저풀

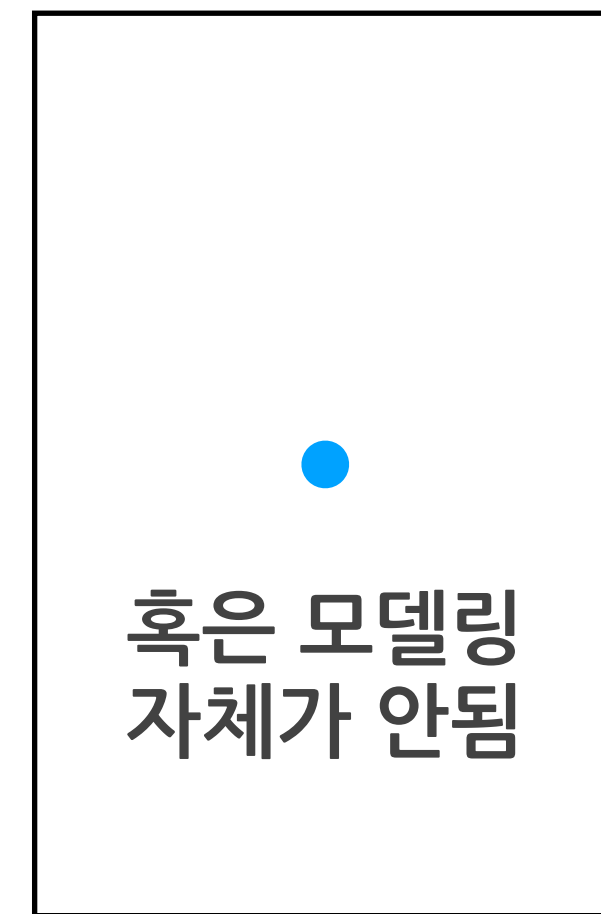


▶ 만렙 등등
조건 걸어서
쓰면 되는거
아님? ▶

과거 제재기록

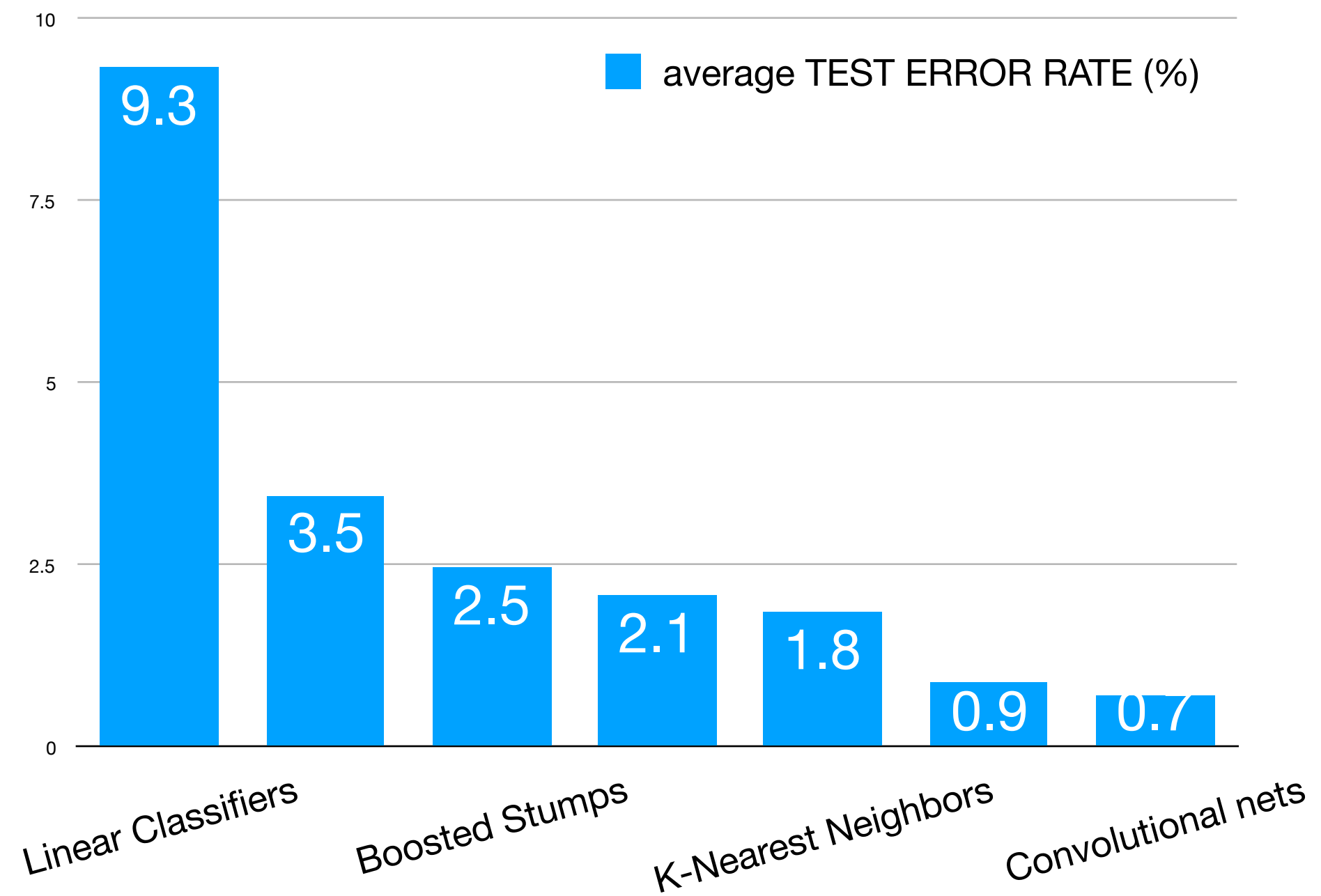


조건유저



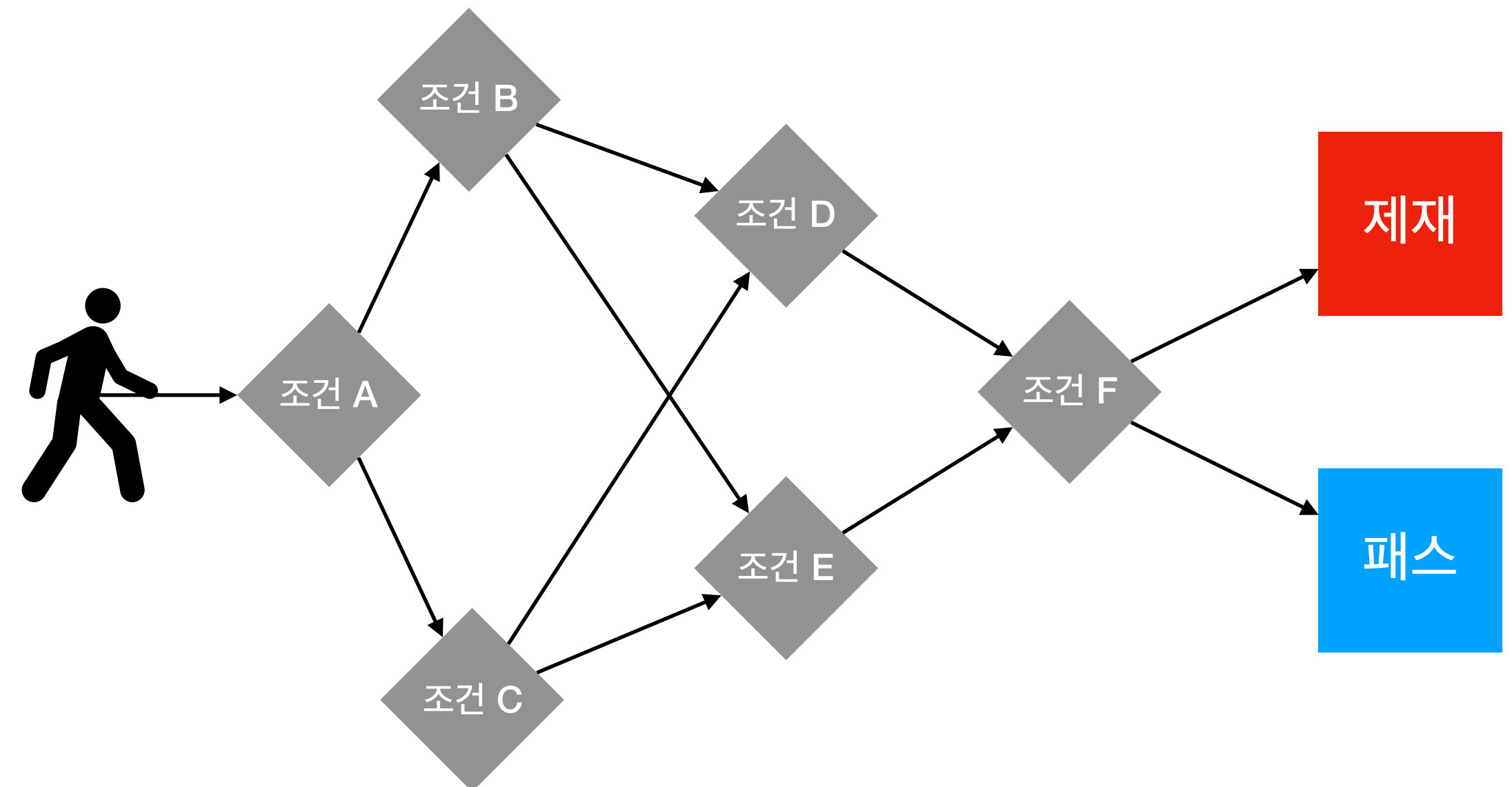
문제의 문제

MNIST



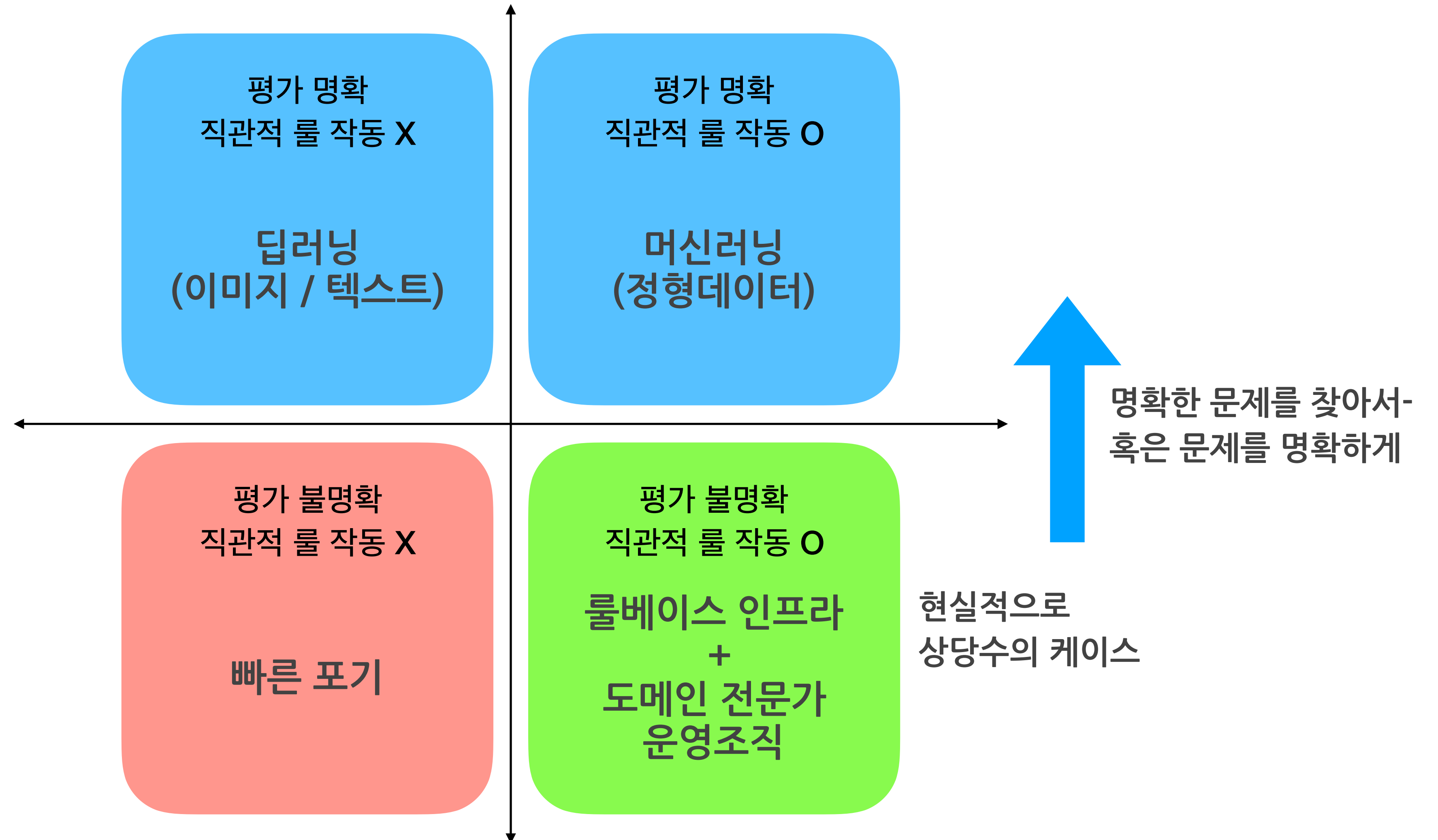
애초에 룰베이스로 풀기가 어려운 문제임

Reality



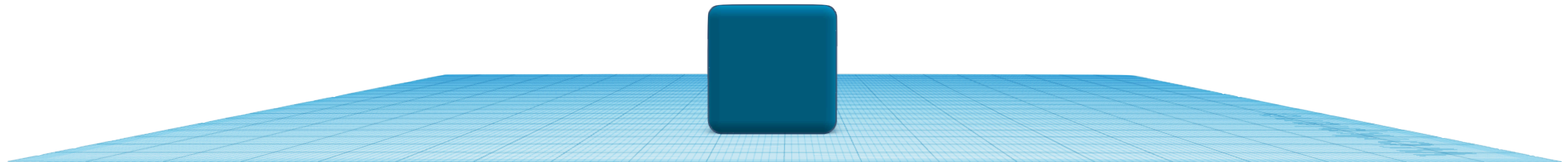
룰베이스 인프라 & 팀만 있으면
잡기도 쉽고 관리도 쉽고 모두가 해피

문제의 문제



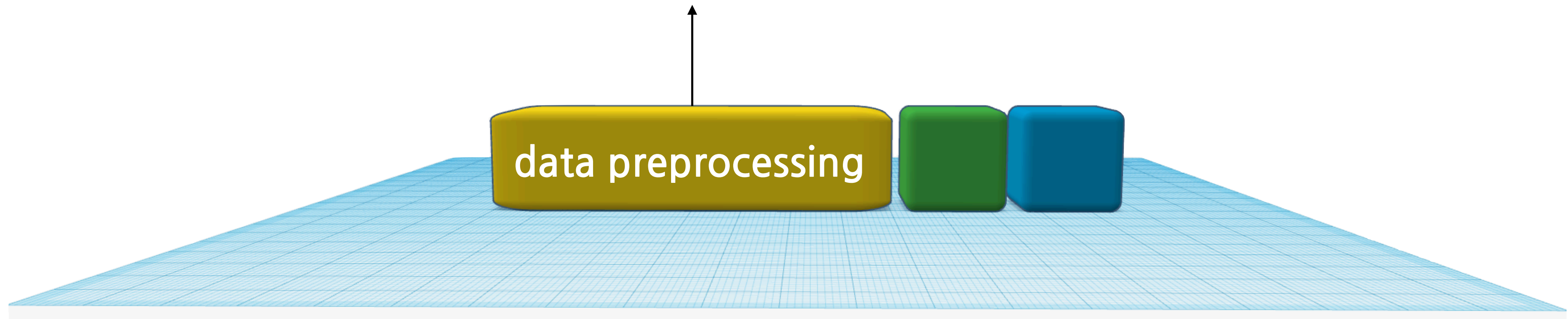
MNIST: 모델링만 해

정확도 97%의 멋진 결과!



MNIST: 전처리는 내가 다해줬음

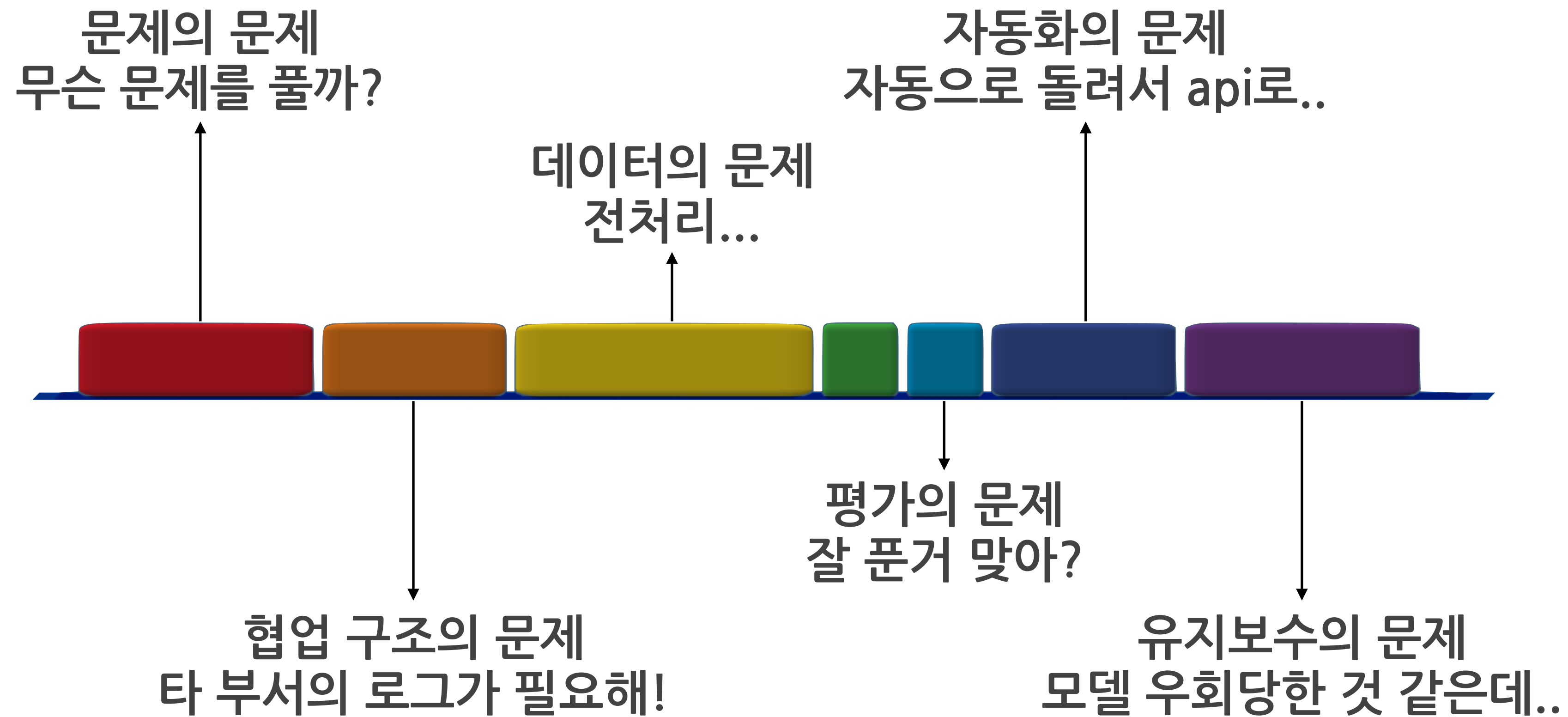
```
from tensorflow.examples.tutorials.mnist import input_data
```



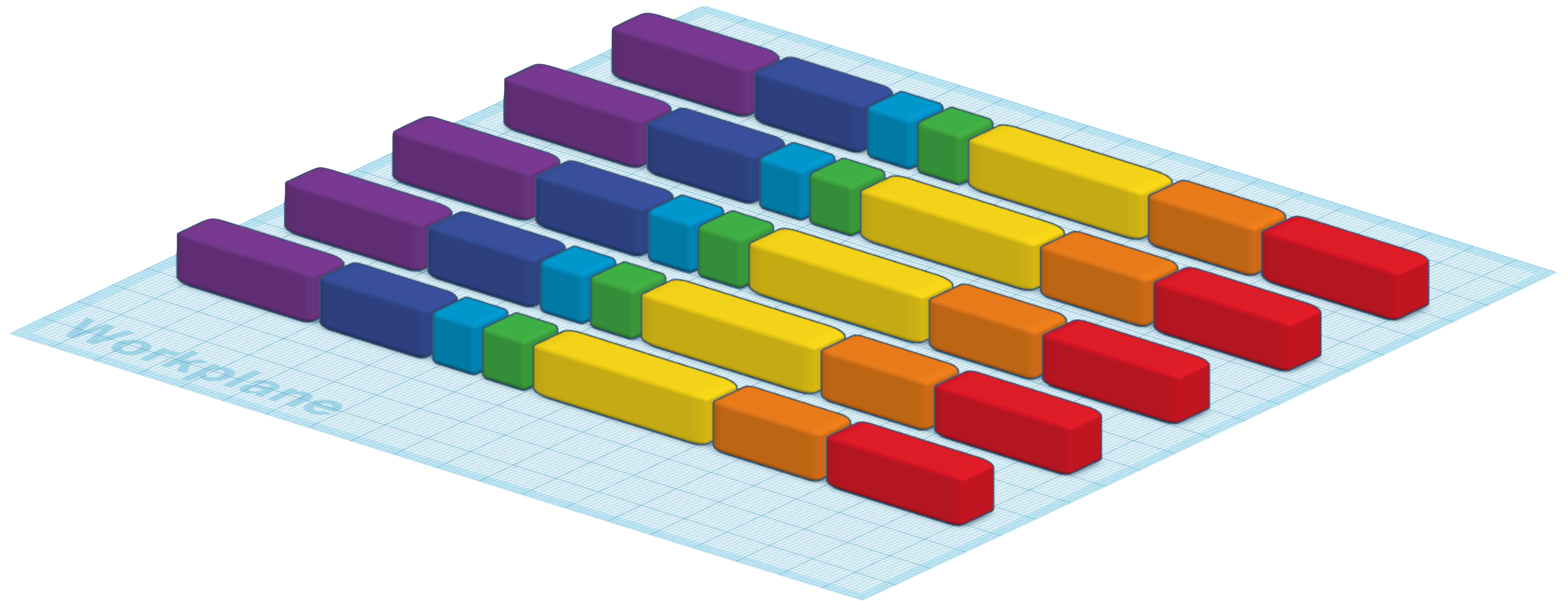
현실에서는 MNIST를 만들어야

- 분석에 필요한 데이터가 있는가?
- 확보한 데이터는 사용가치가 있는가?
- 머신러닝에 적합한 양과 질인가?

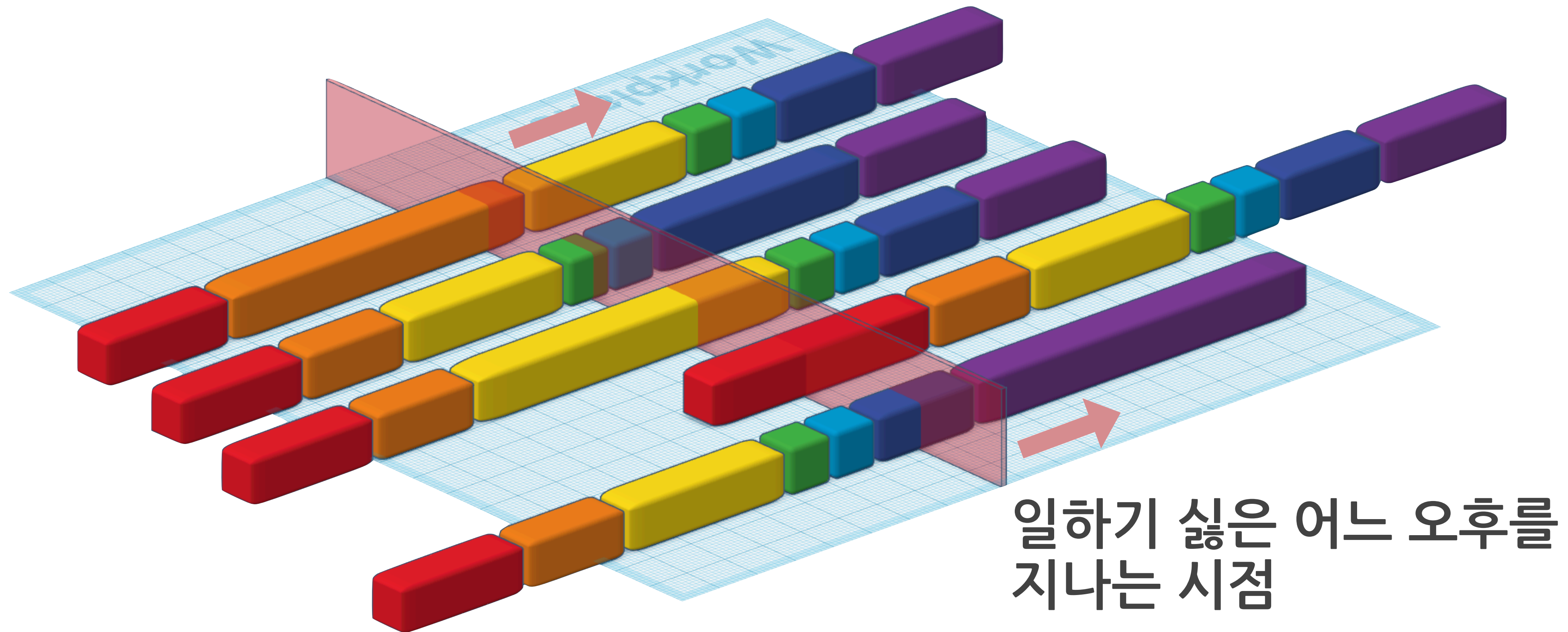
현실: 사실 앞뒤로 더 있음 ㅋ



현실: 그게 병렬로 ㅋㅋ

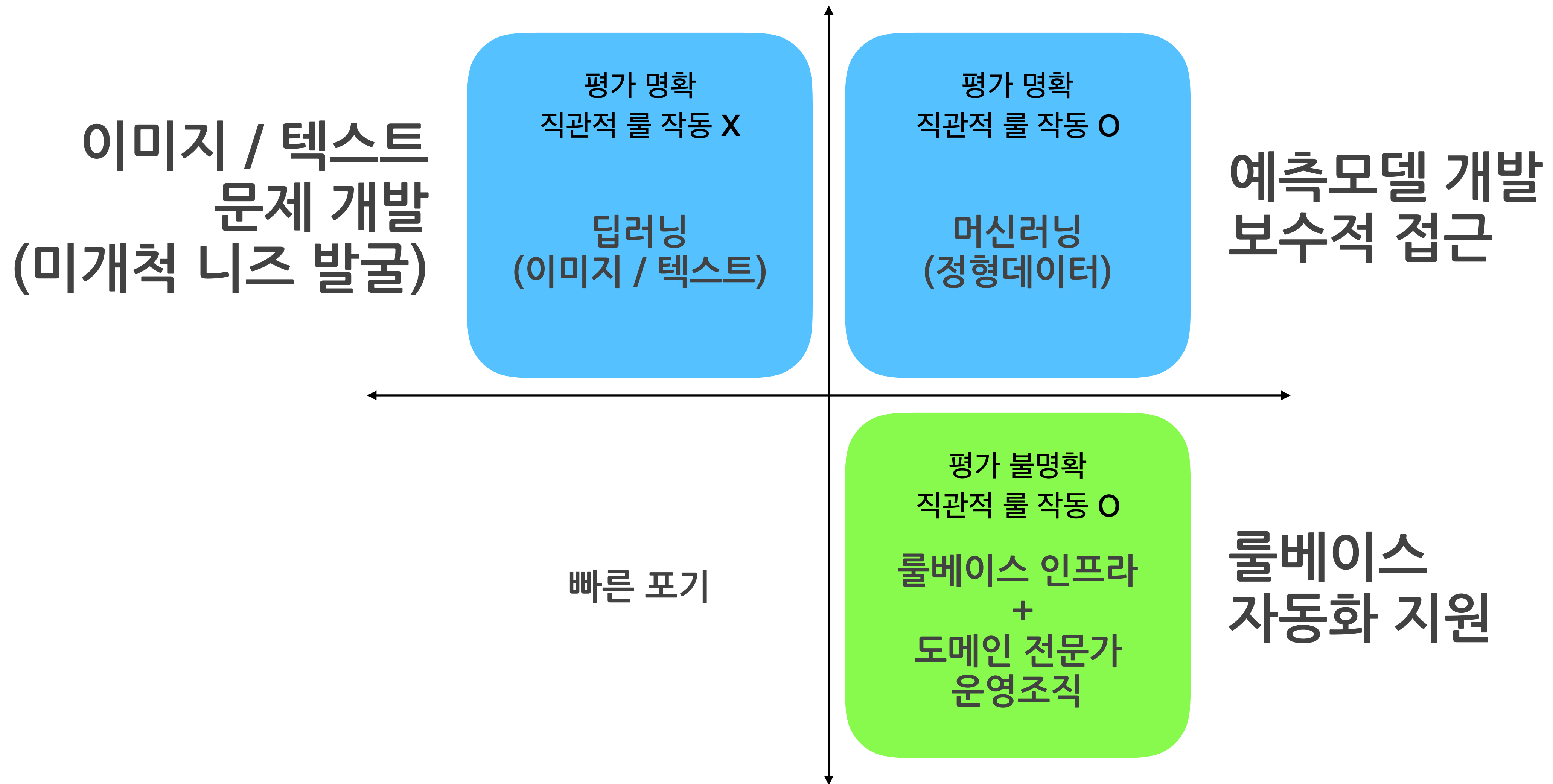


현실: and 다이나믹하게 엇박으록



그래도 머신러닝, 딥러닝

문제의 문제: 빠른포기와 집중



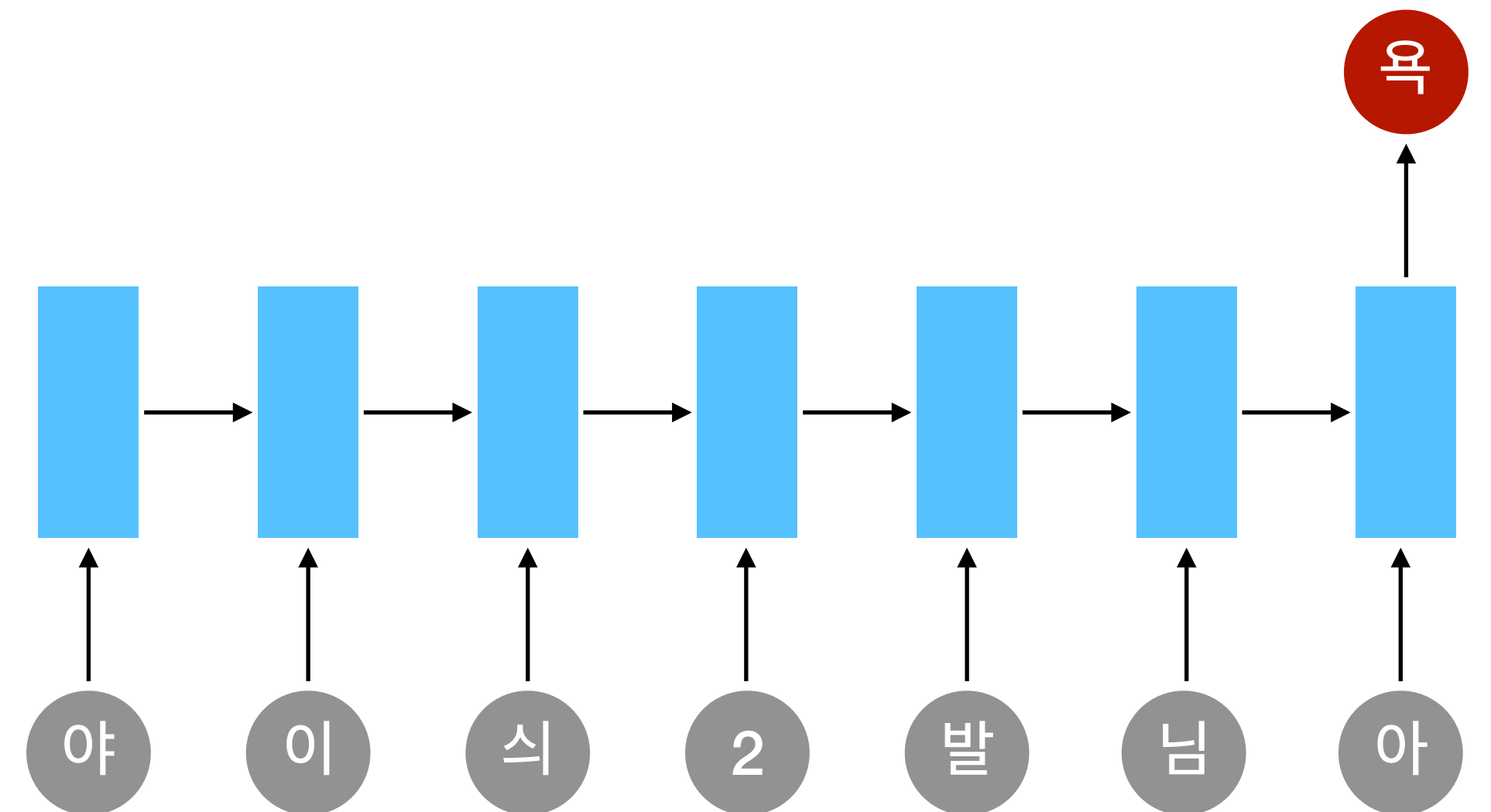
문제의 문제: 딥러닝 가능 문제

FPS 월핵탐지기



Transfer Learning / Image Segmentation

in-Game 욕설탐지기



LSTM / Attention Model

평가의 문제: 설명가능한 것부터

edit slide

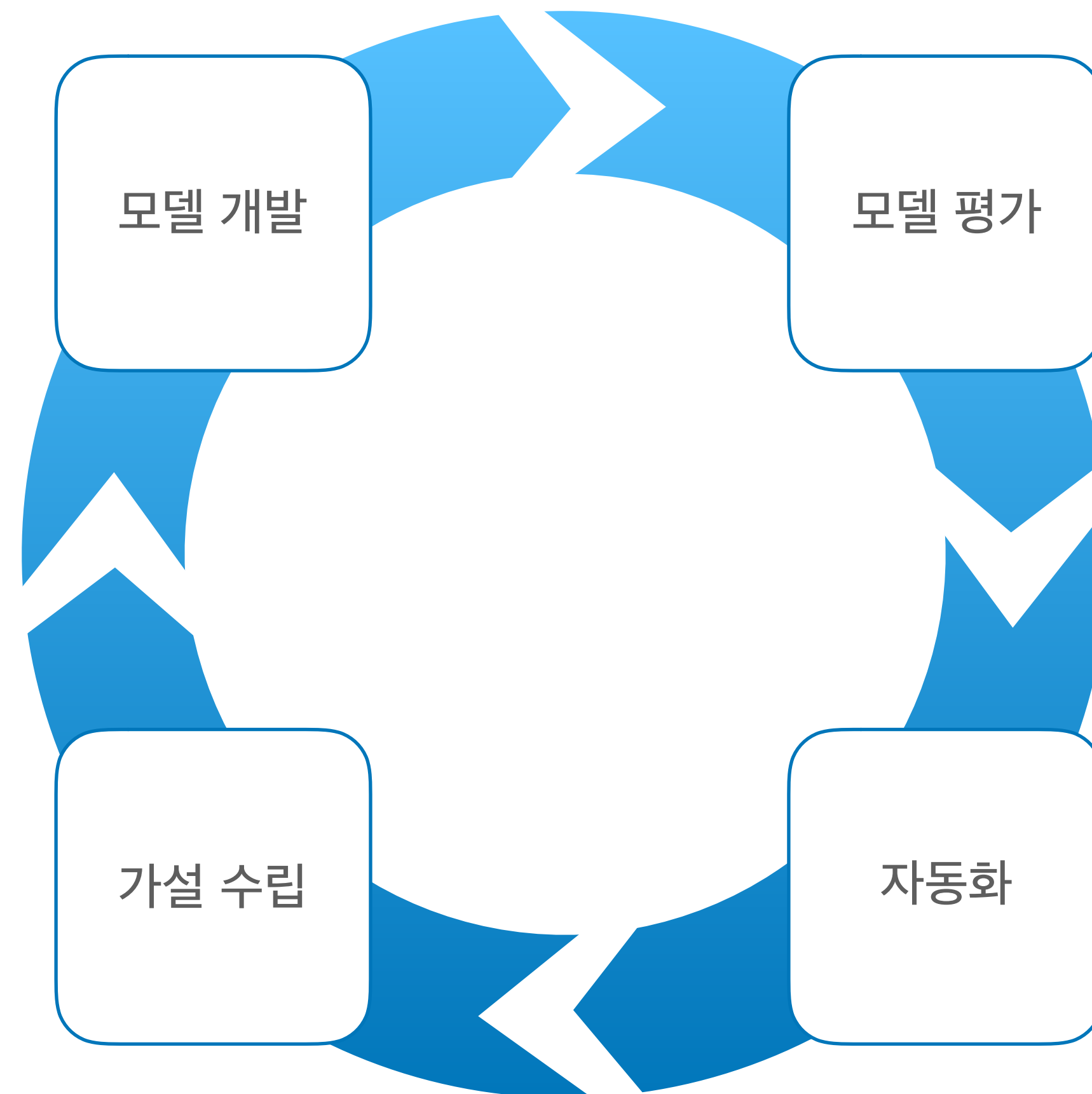
K-Nearest Neighbors
LBD season 1 screenshot

평가의 문제: 패스트 프로토타이핑

모델을 최대한 빠르게 만들어서
평가를 많이 할 수 있도록-



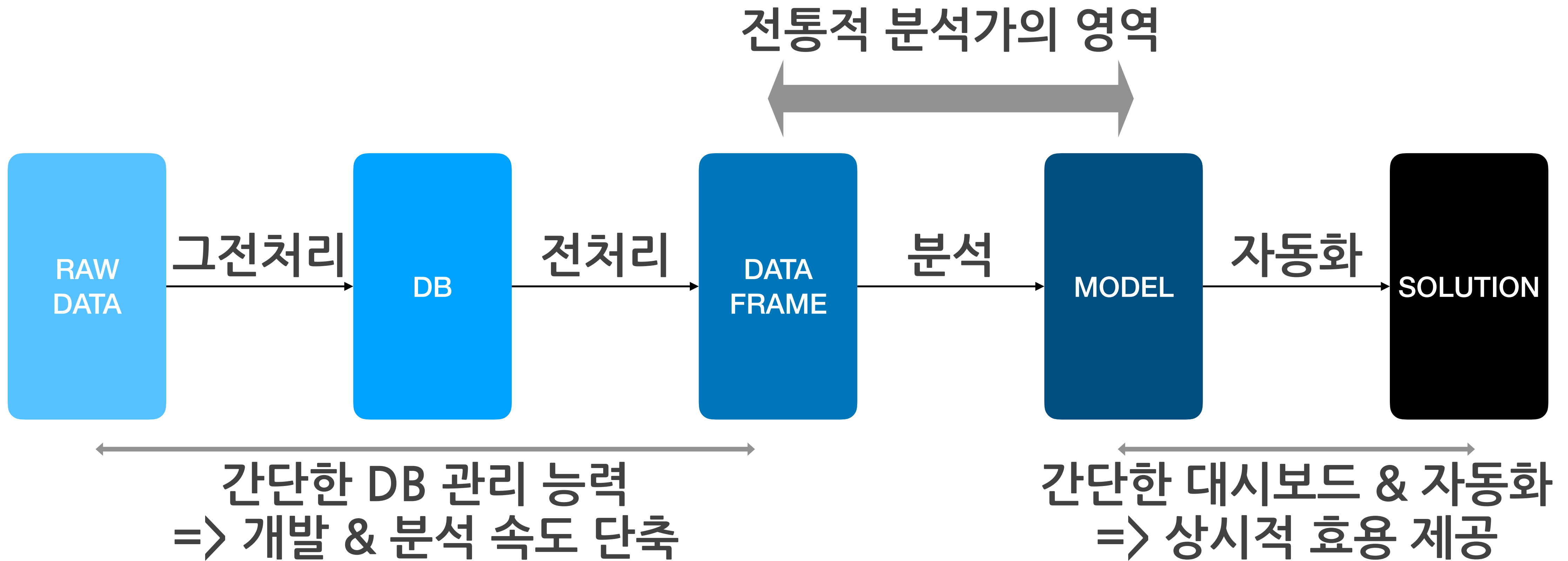
분석가의 엔지니어화
엔지니어의 분석가화



대시보드 / 이메일로
빠른 사용자 피드백

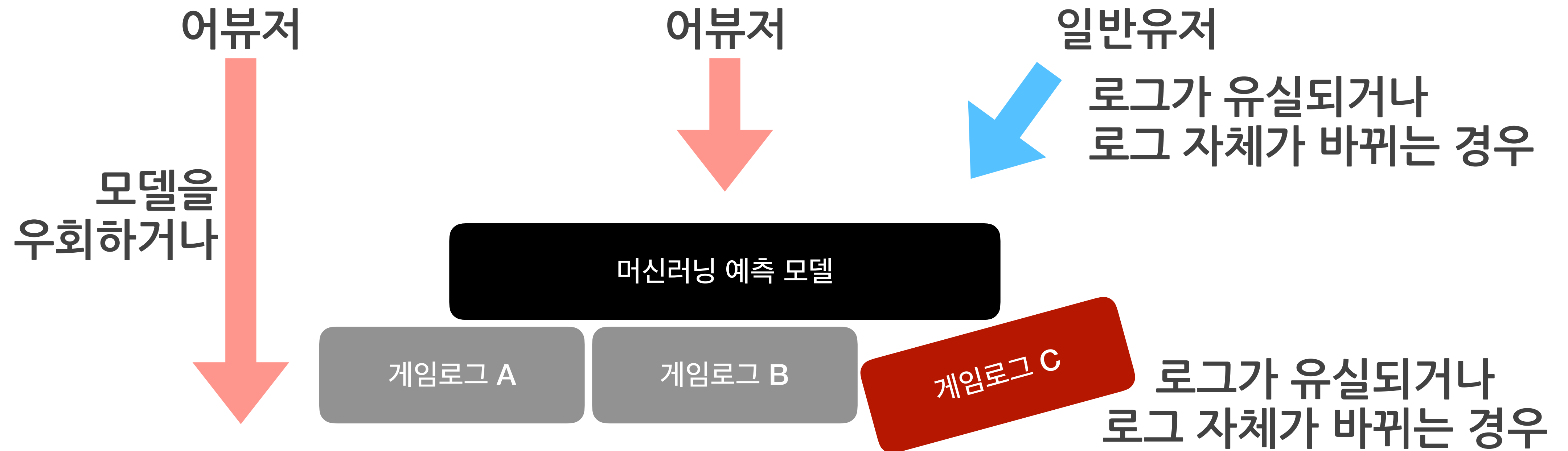


전처리의 문제: 분석외 역량 강화



남은 문제들

남은 문제들: 유지보수의 문제



남은 문제들: 분석을 넘어서

@NEXON

감사합니다