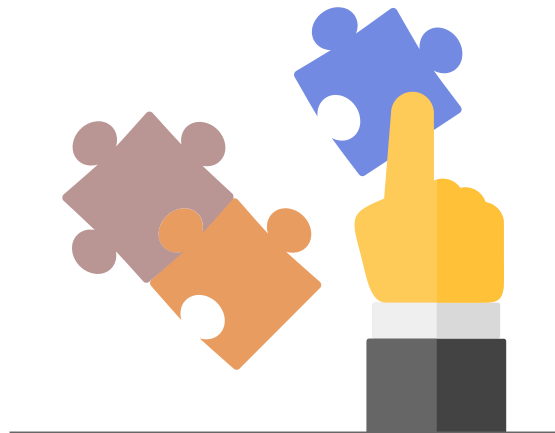


# 백준 문제 추천 및 라이벌 추천 서비스

노명은, 백지원, 임정우, 한상진



01

프로젝트 소개

02

추천시스템

03

데이터

04

모델

05

향후 프로젝트 계획

# 01.프로젝트 소개

## 카카오 채용 개요

지원기간 2021.08.19(목) 11:00 ~ 09.06(월) 17:00

### 전형절차



- 온라인 코딩테스트 : 1차 9/11, 2차 9/25 (2차 코딩테스트는 1차 합격자만 응시함)
- 1차 인터뷰부터 모집회사별로 전형 진행
- 최종 합격발표 11월 중순, 입사 2021년 연중 (모집회사별로 상이함)

## 코딩테스트도 ‘토익’처럼…프로그래머스 인증시험, 개발자 채용 판 바꾼다

디지털데일리 | 발행일 2022-08-23 14:08:21

최민지 |

- 그림, 민간자격인증 ‘프로그래머스 코딩역량인증시험’ 출시
- 다음달 실무역량 과제테스트 선배
- 채용 전형, 졸업 요건에 포함하려는 기업·대학 곳곳에
- 지난해 그림 매출 82억원, 올해 150억원 목표



(디지털데일리, 2022)

01.프로젝트 소개

- 문제 선택에 대한 어려움
- 학습 동기 부족

"AI가 맞춤형 코딩문제 추천"...그렘, '튜터 리포트' 출시

등록 2023.07.19 11:59:59

이메일, 인쇄, 공유, 가, 가

학습자 자료 구조와 문제 해결력 분석



Gold I 20.18%

911,195 / 4,515,400



여기에 단체 이름이 표시됩니다.

1,416 AC RATING 7,232,510 EXP 1,967위 166문제 해결



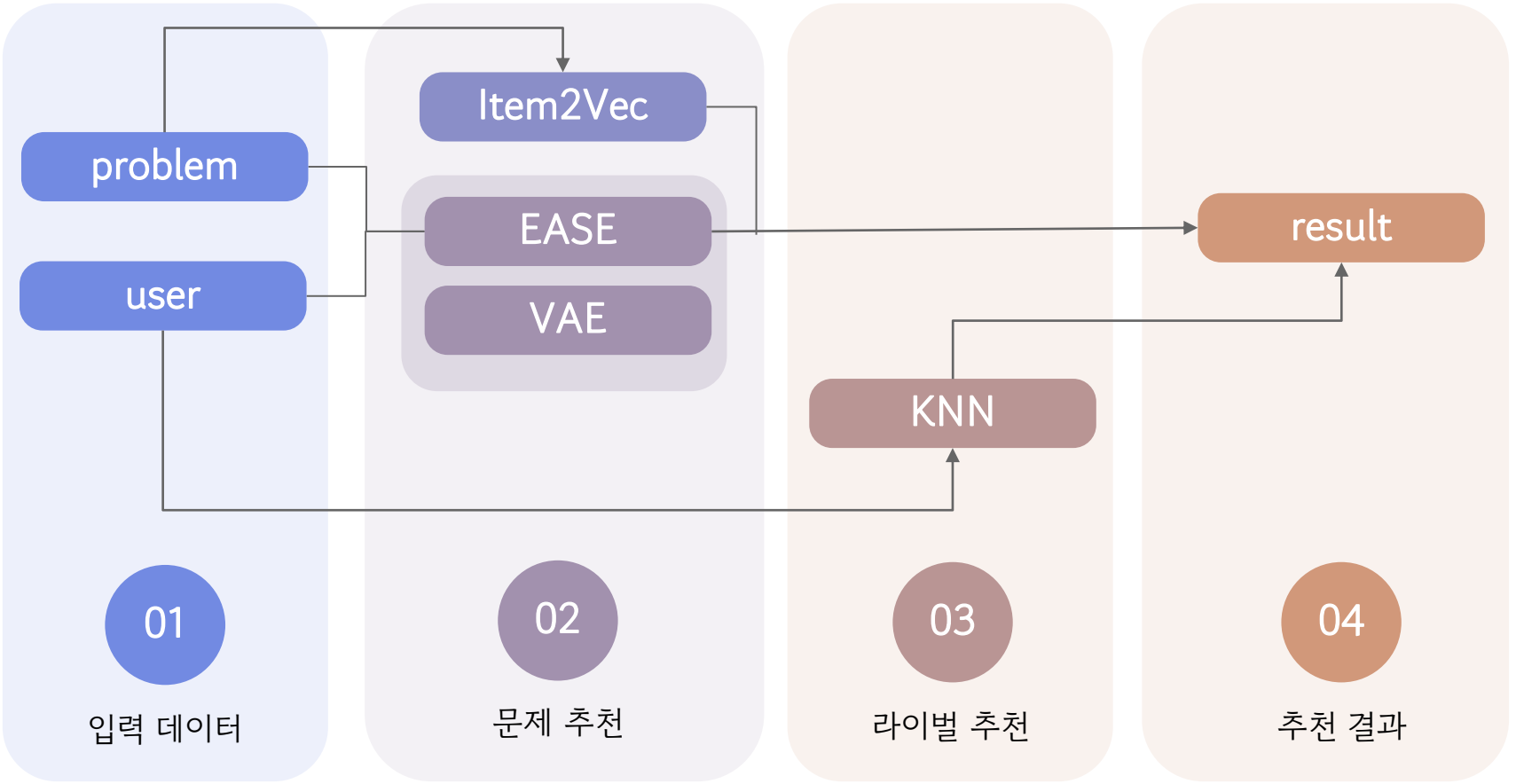
레벨	문제	EXP
• BRONZE	20 12.0%	21,048 0.3%
• SILVER	84 50.6%	820,871 11.3%
• GOLD	49 29.5%	2,467,758 34.1%
• PLATINUM	13 7.8%	3,922,833 54.2%
• DIAMOND	0 0.0%	0 0.0%
• RUBY	0 0.0%	0 0.0%



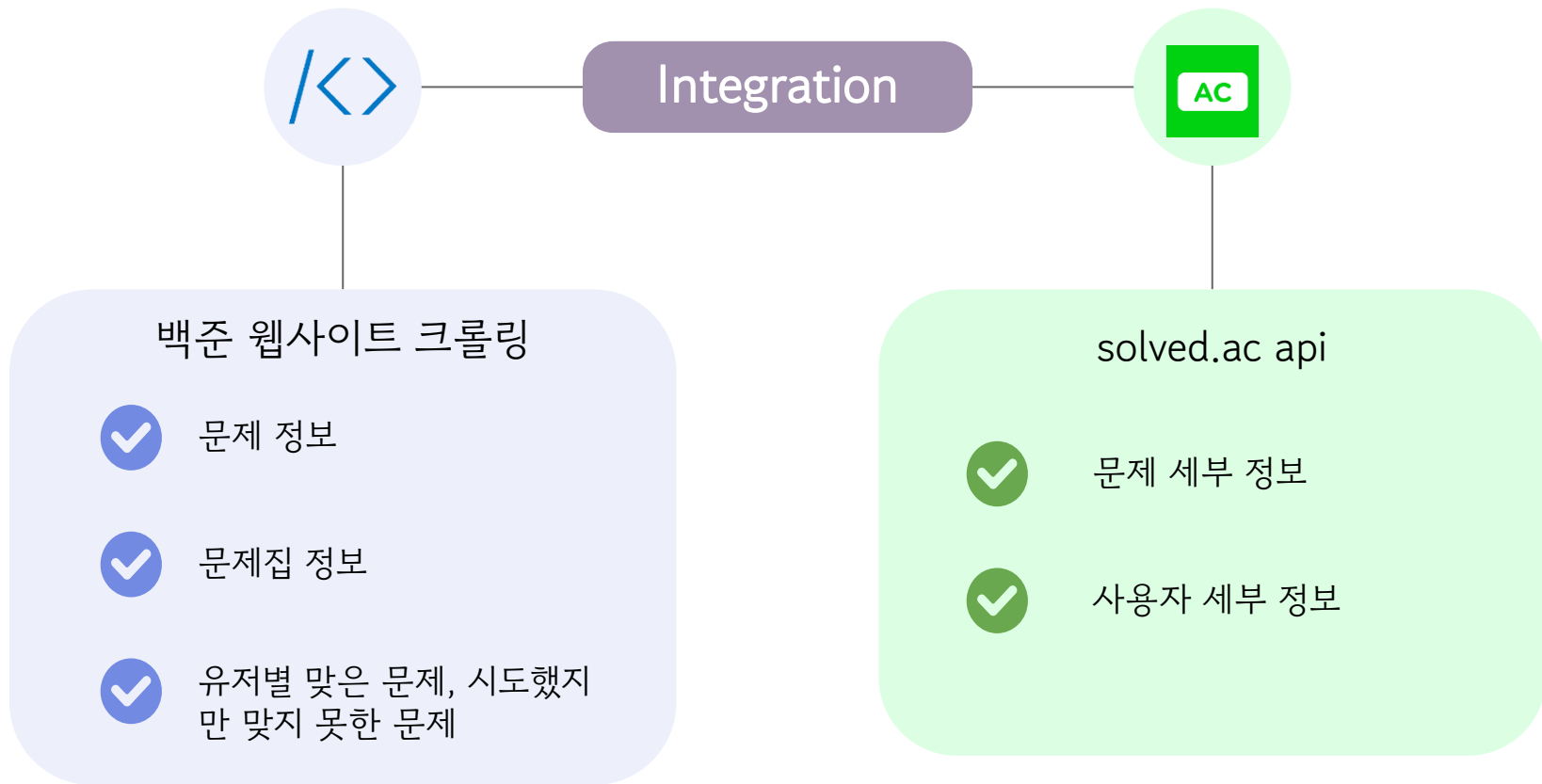
태그	문제	EXP
• 자료 구조	26 15.7%	3,770,190 52.1%
• 세그먼트 트리	18 10.8%	3,705,374 51.2%
• 다이나믹 프로그래밍	68 41.0%	1,877,421 26.0%
• 분할 정복	6 3.6%	692,737 9.6%
• 그래프 이론	21 12.7%	487,560 6.7%
• DFS	1 0.6%	408,280 5.6%
• BFS	1 0.6%	408,280 5.6%
• 오목판인 판	1 0.6%	408,280 5.6%
• 무방향 그래프	1 0.6%	408,280 5.6%
• 계층적 분할	1 0.6%	408,280 5.6%
• 스택	1 0.6%	408,280 5.6%
• 그래프 탐색	19 11.4%	406,716 5.6%
• 문제풀	9 5.4%	336,743 4.7%
• 이진 탐색	8 4.8%	314,237 4.3%
• 나비 무한 탐색	15 9.0%	279,582 3.9%
• 스레	4 2.4%	189,648 2.6%

= 더 보기

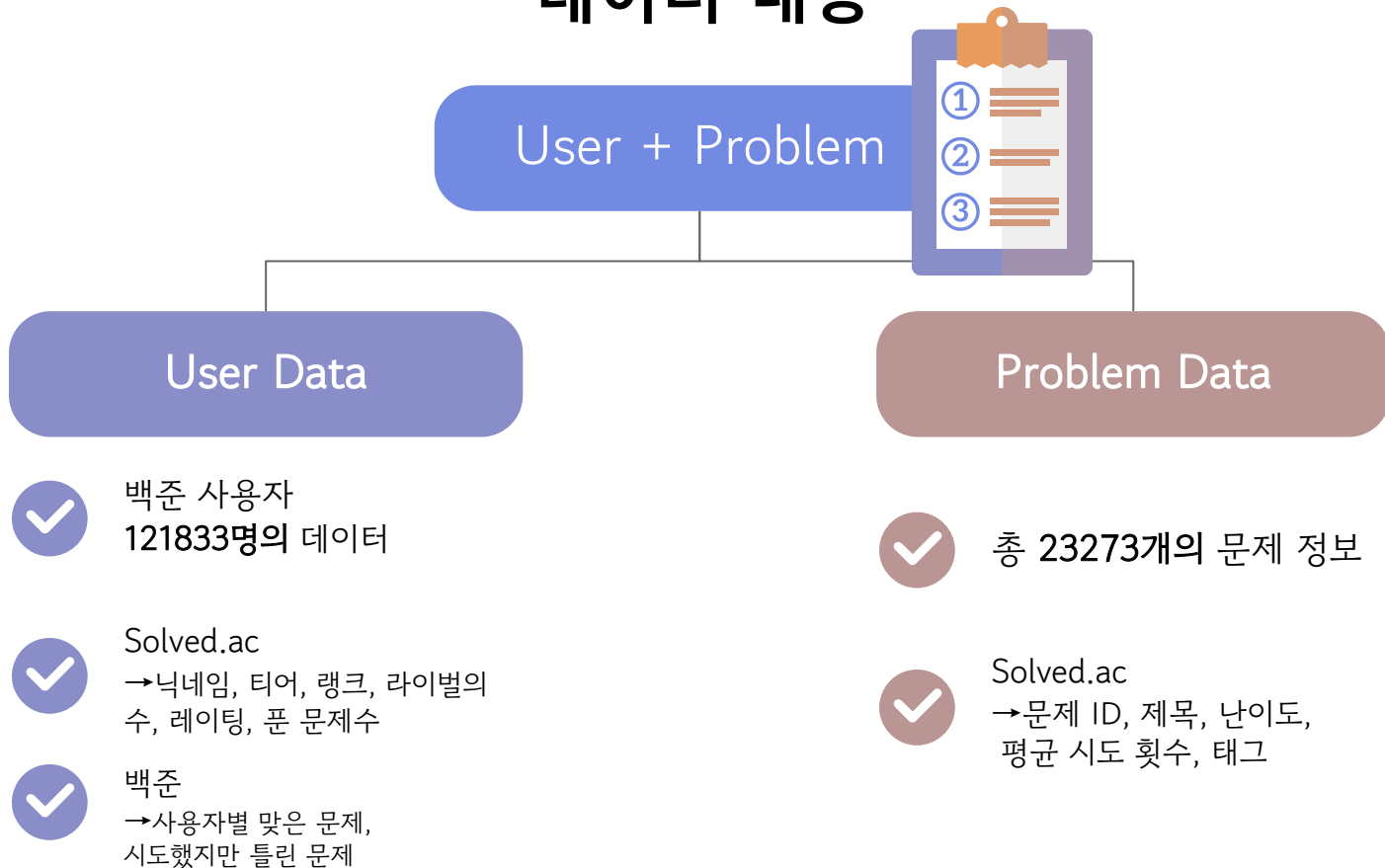
# 추천시스템 개요



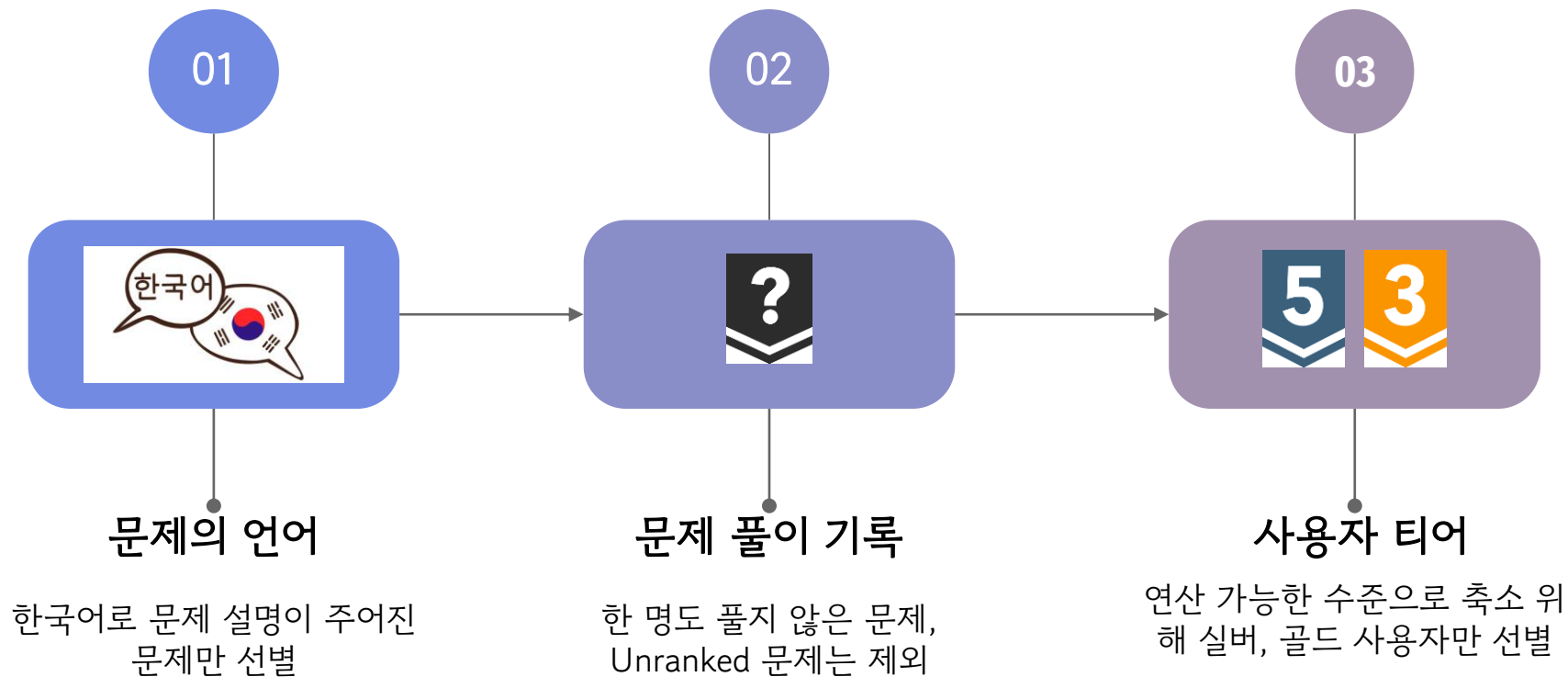
## 데이터 수집 과정



## 데이터 내용



## 데이터 전처리 과정





## Item2Vec

- word2vec 응용
- 문제들의 **난이도, tag** 정보를 바탕으로 유사도 연산
- 유사도 정보 바탕으로 문제 추천 보조 수단으로 이용

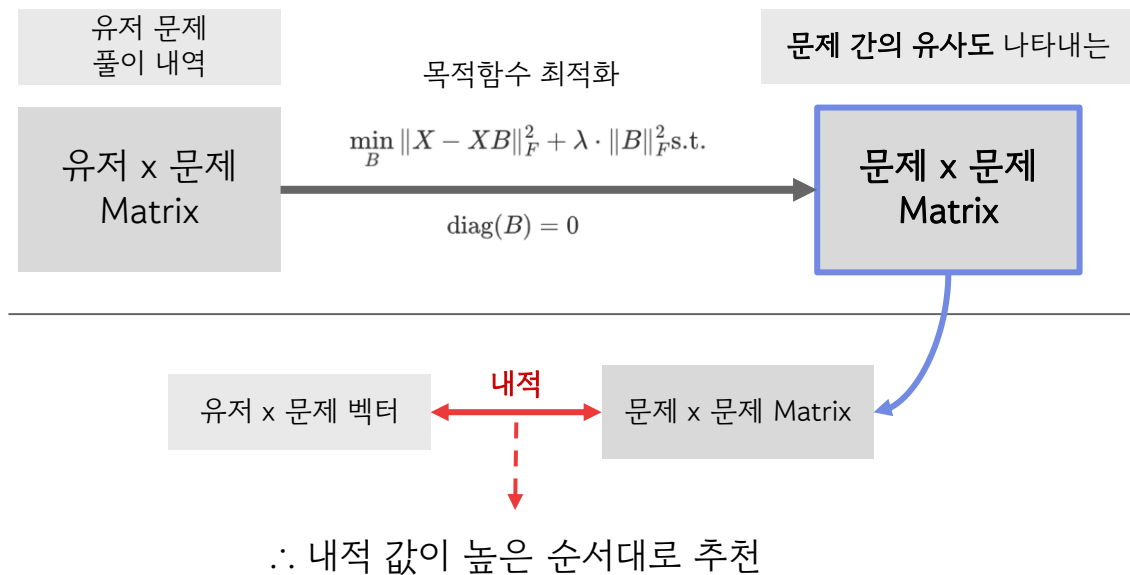


## EASE

## Embarrassingly Shallow AutoEncoders for Sparse Data

## 💡 EASE 모델 핵심

- ✓ AutoEncoder 계열
- ✓ hidden layer가 없다.
- ✓ 목적함수 최적화 연산 수
- ✓ sparse한 데이터에 robust
- ✓ 문제 x 문제 Matrix를 Co-occurrence로 해석 가능  
자주 등장할수록 높은 값  
→ 문제의 인기도 반영



# VAE

## Variational AutoEncoder

- AutoEncoder 계열의 생성 모델
- 유저의 문제 풀이 내역을 바탕으로 유저 x 문제 Matrix 구성
- Sparse한 Matrix를 Dense Matrix로 재구성해 k개의 문제를 추천

Input    1: 풀었음  
          0: 풀지 않음

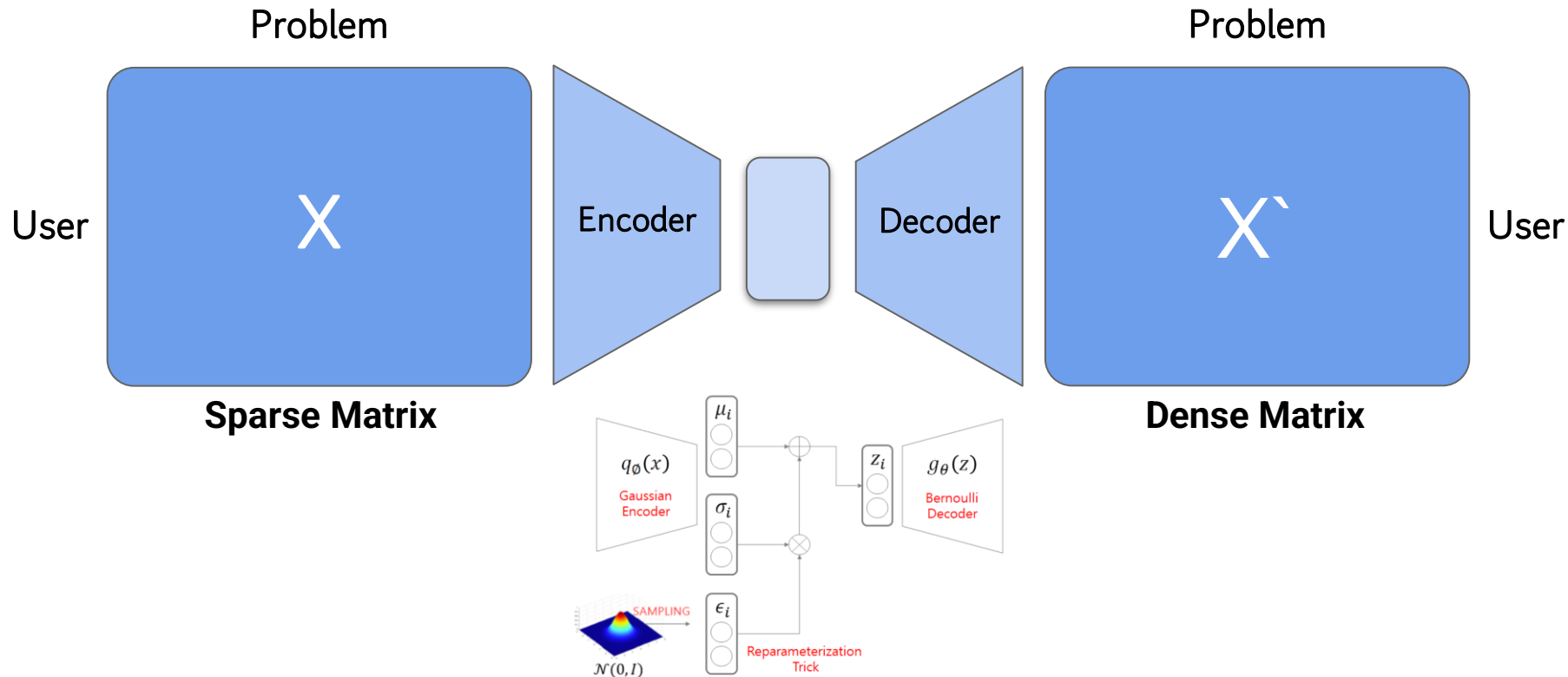
	problem_1	problem_2		problem_m
user_1	1	0	...	0
user_2	0	0	...	1

Output

	problem_1	problem_2		problem_m
user_1	—	0.163	...	0.742
user_2	0.463	0.032	...	—

## VAE

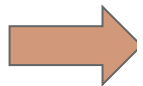
Model 구조



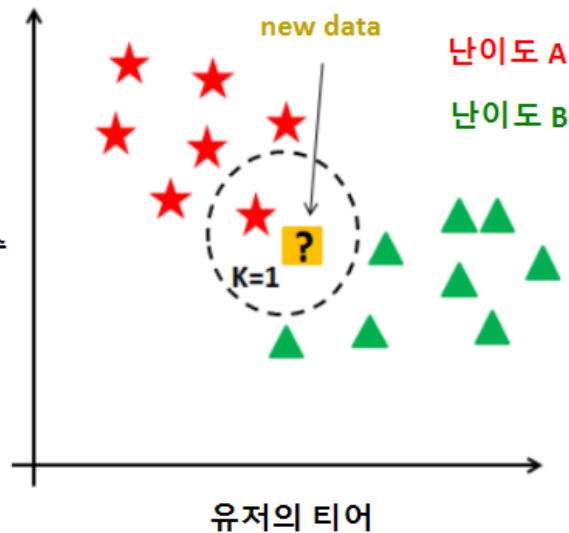
## KNN

- KNN을 이용 > 라이벌 추천
- 유저의 티어, 유저별 맞은 문제와 문제의 난이도 적용
- 가장 유사한 유저를 라이벌로 추천

문제의 난이도	1																			...	30
유저의 티어	6	7	8	9	10	11	12	13	14	15	...	6									
사용자 아이디																					
5990	0	0	0	0	0	0	0	0	0	6	...	0									
5992	0	0	0	0	0	0	0	0	0	7	...	0									
5994	0	0	0	0	0	0	0	0	0	7	...	0									
5996	0	0	0	0	0	0	0	0	0	11	...	0									
5998	0	0	0	0	0	0	0	0	0	6	...	0									



문제의 개수



### #1 추천시스템 고도화

- LightGCN, MF, FM 모델 추가 활용
- 추가 데이터( 예 : 유저별 푼 문제의 총 시도 횟수, 오답 횟수 등) 수집
- 데이터 (예 : 태그별 문제 ) 고도 활용
- MAB, Thompson Sampling 등 활용
- 평가 지표 고안

### #2 웹사이트 제작



- ✓ 유저별 맞춤형 문제 추천
- ✓ 유저별 맞춤형 라이벌 추천
- ✓ 자주 틀린 유형 관련 문제 추천



### #3 프로파일링

- 자주 틀린 문제 유형, 데이터 분석 측면 발전

# 감사합니다.

## [ Reference ]

- 네이버 부스트캠프 AI Tech : <https://github.com/boostcampaitech3/final-project-level3-recsys-05>
- 네이버 부스트캠프 AI Tech : <https://github.com/boostcampaitech3/final-project-level3-recsys-14>