

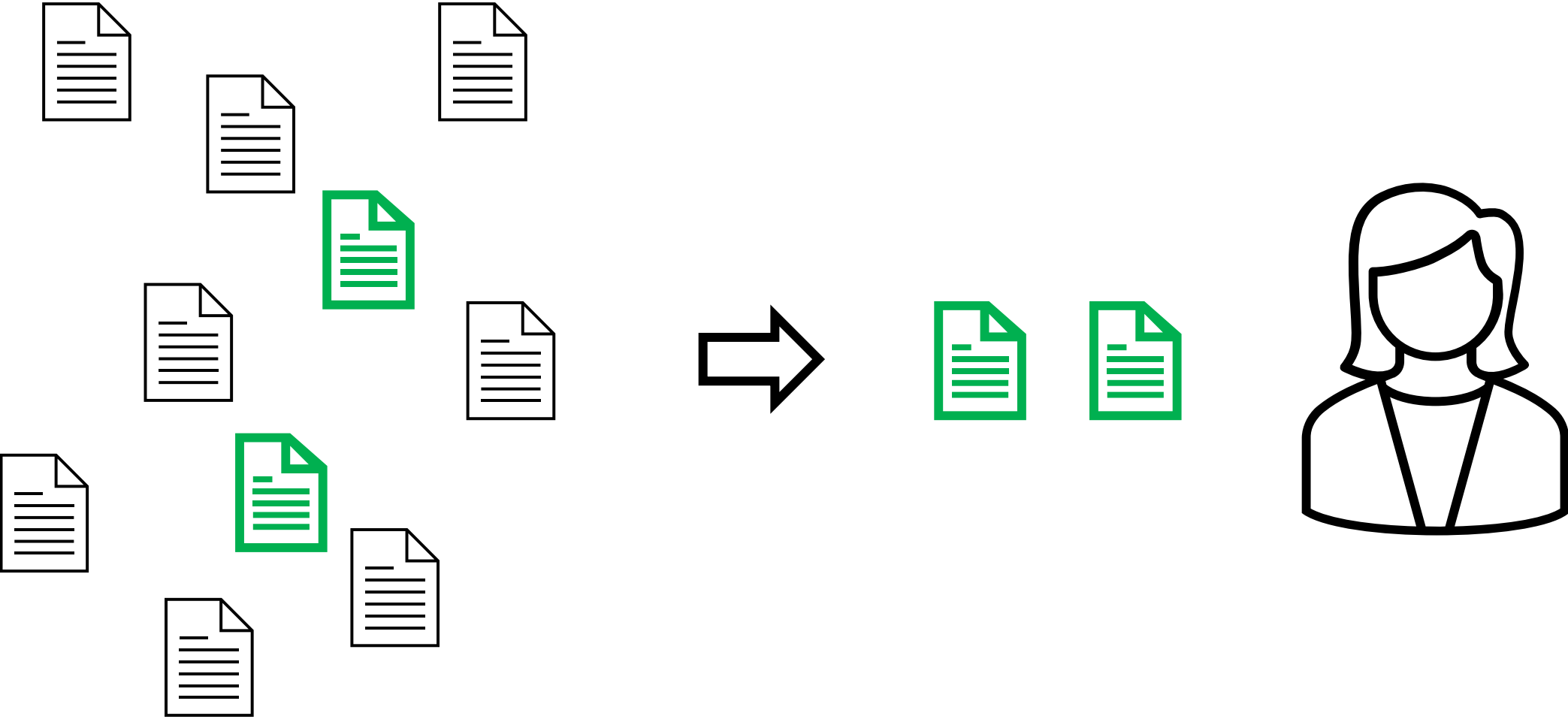


문화 A0007

데이터사이언스입문

김 태 완

kimtwan21@dongduk.ac.kr



추천시스템

- 활용 예시 1 : 카카오 페이지
 - 이 작품 다음엔 뭘 보면 좋을까?
 - Word2vec
 - TF-IDF

10:47

작품홈

작품소개

작품 설명

영원불멸의 삶을 위해 사후세계에서 지구로 도망쳐 온갖 악행을 저지르고 다니는 악한 영혼들.
그리고 그들을 쫓는 악귀 사냥꾼 '카운터'!

평범하던 고등학생 소문이 범상치 않은 능력을 보이며 카운터에 합류하게 되고,
다양한 능력을 쓰며 악귀를 쫓는 팀원들과 함께
국숫집 직원으로 위장하여 악한 영혼들을 추격하기 시작한다!

분류

기다무웹툰 | 드라마

발행자

다음웹툰컴퍼니

연령등급

15세 이용가

전자책 정가

400원/회차 당

작가

글

장이

그림

장이

이 작가의 작품

15 경이로운...
장이

추천작품

승리호
홍작가

당골의 신부
오주,나은성

카카페 매거진
카카오페이지

15 페이크뷰티
Liz

라이아
Shinc

★★★★★ 9.9 (406명)

별점주기 >

댓글 (183)

더보기 >

BEST

하르온♡루시아언니_사랑해요♡

하.. 가장 썸보이는 놈이 저러고 있는데...
울 팀이었음 좋겠다ㅏ...

BEST

Song

드라마 덕에 알게 돼서 접했다가 하루 만에 전부 다
보고, 몇 번째 정주행인지.. ㅎㅎ 여러분 이거 시즌2도...

S/A/L/E 53,000→6,930원

STL #힙커버 #체형커버 #롱슬리브

완결된 작품입니다.

추천

승리호

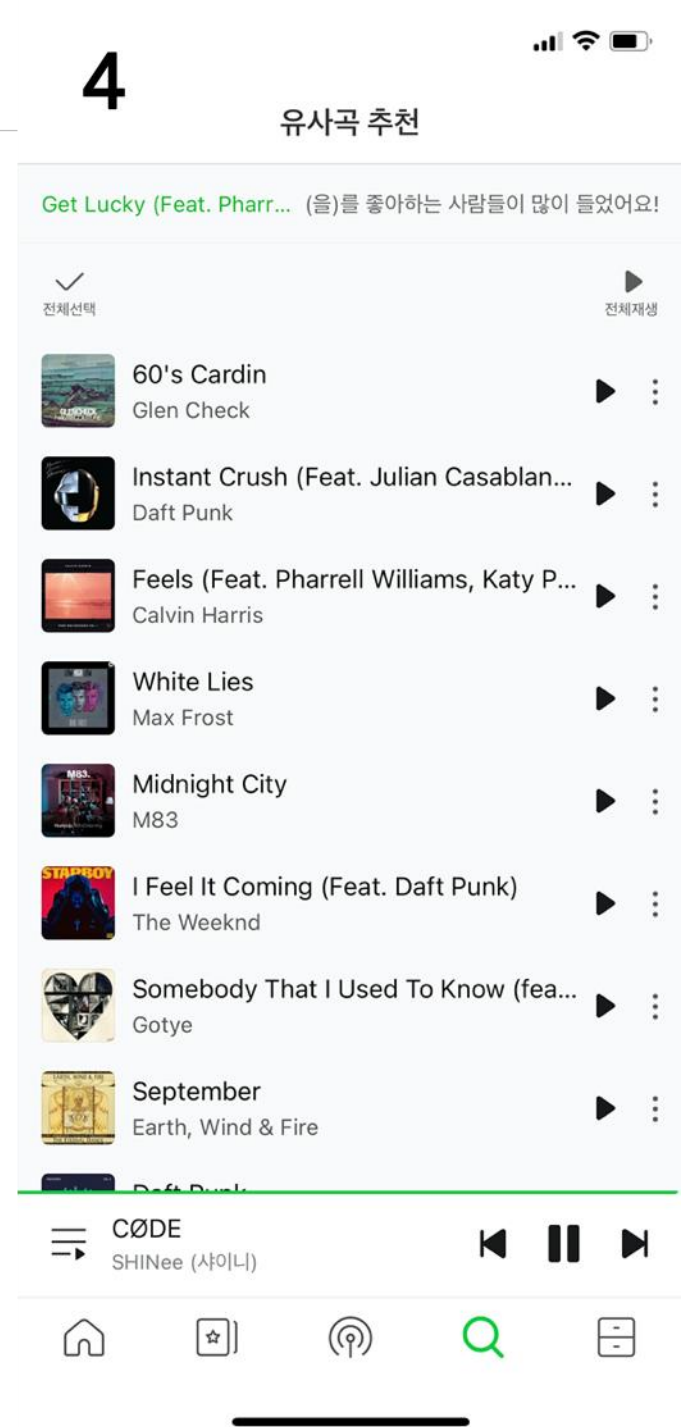
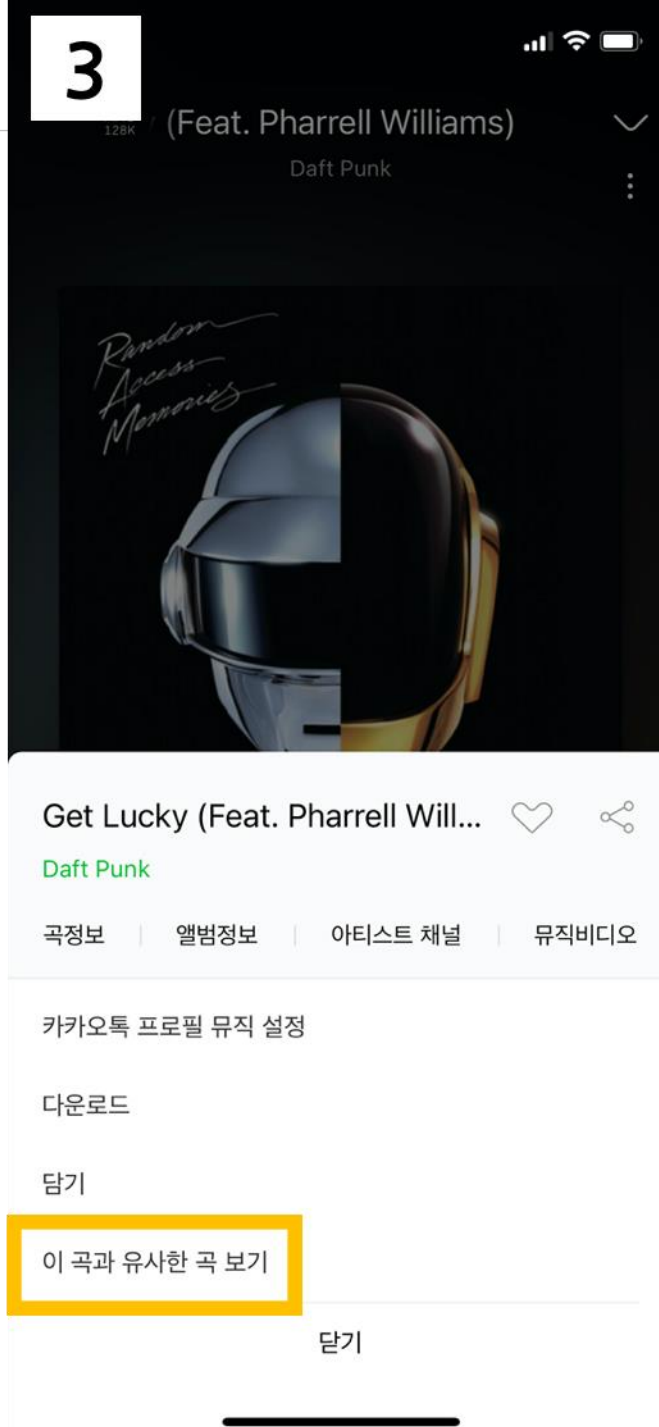
웹툰 | 드라마 | 홍작가

추천 작품 보러가기

작품홈 가기 >

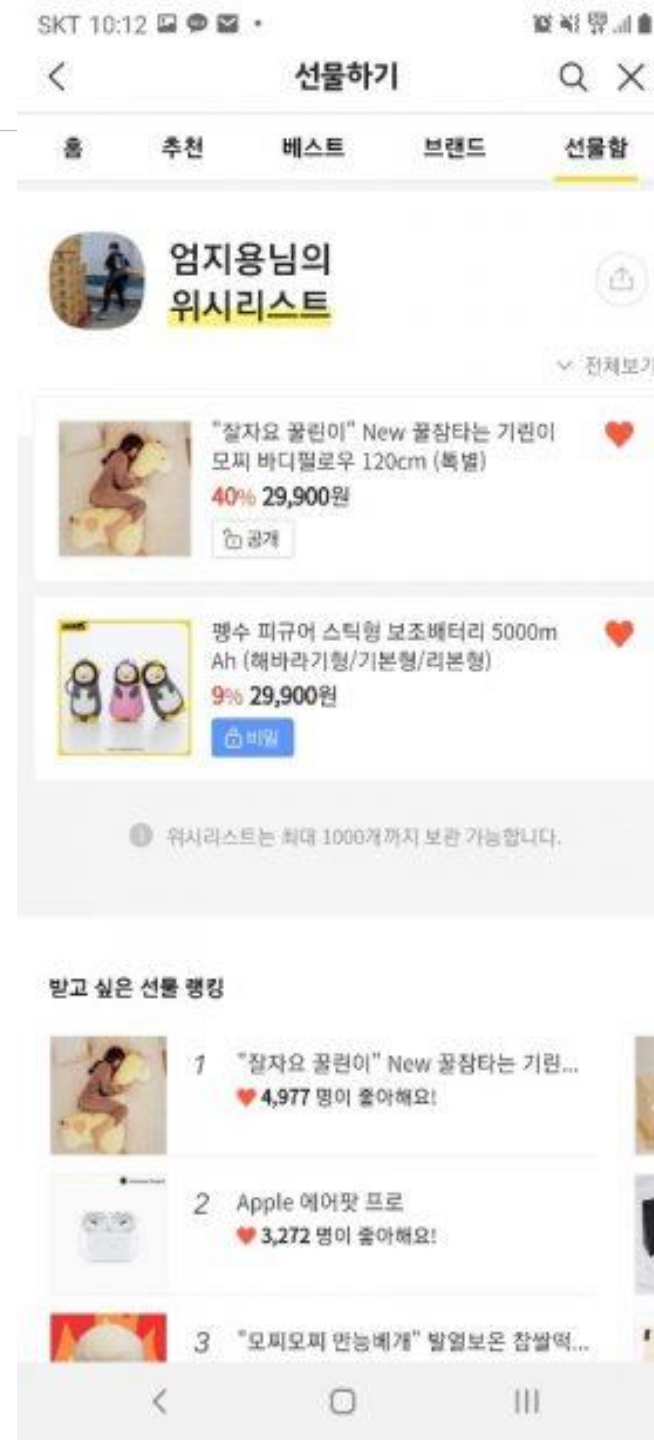
추천시스템

- 활용 예시 2 : 멜론
 - 지금 이 음악 스타일이 마음에 든다면...
 - 협업 필터링 (COLLABORATIVE FILTERING)



추천시스템

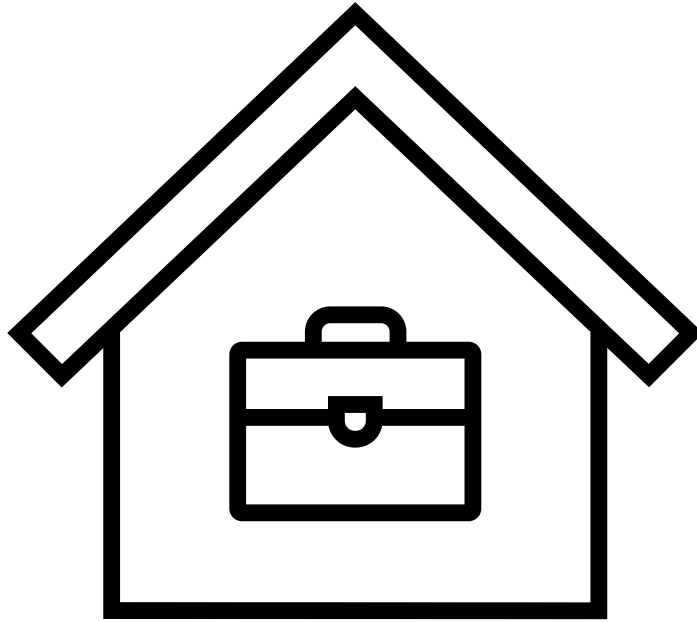
- 활용 예시 3 : 네이버쇼핑 / 카카오 선물하기
 - 개인별 카테고리 추천 등



추천 시스템은 어떤 문제를 풀고 있을까?

추천시스템

- 예시 : 100가지 종류의 아이스크림을 파는 가게



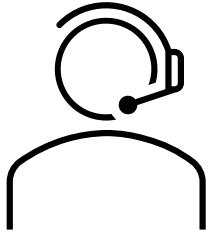
추천시스템

5월 18일

5월 19일

5월 20일

5월 21일(오늘)



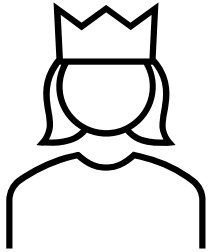
체리쥬빌레



엄마는외계인



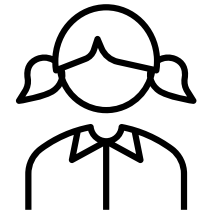
민트초콜릿칩



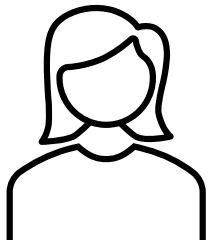
뉴욕치즈케이크



슈팅스타



아몬드붕붕



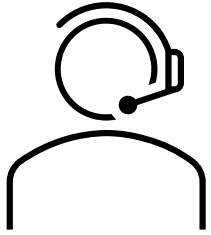
추천시스템

5월 18일

5월 19일

5월 20일

5월 21일(오늘)



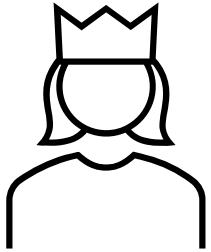
체리쥬빌레



엄마는외계인



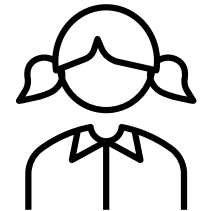
민트초콜릿칩



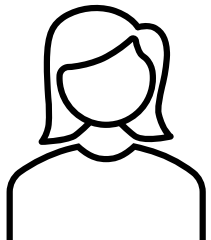
뉴욕치즈케이크



슈팅스타



아몬드봉봉

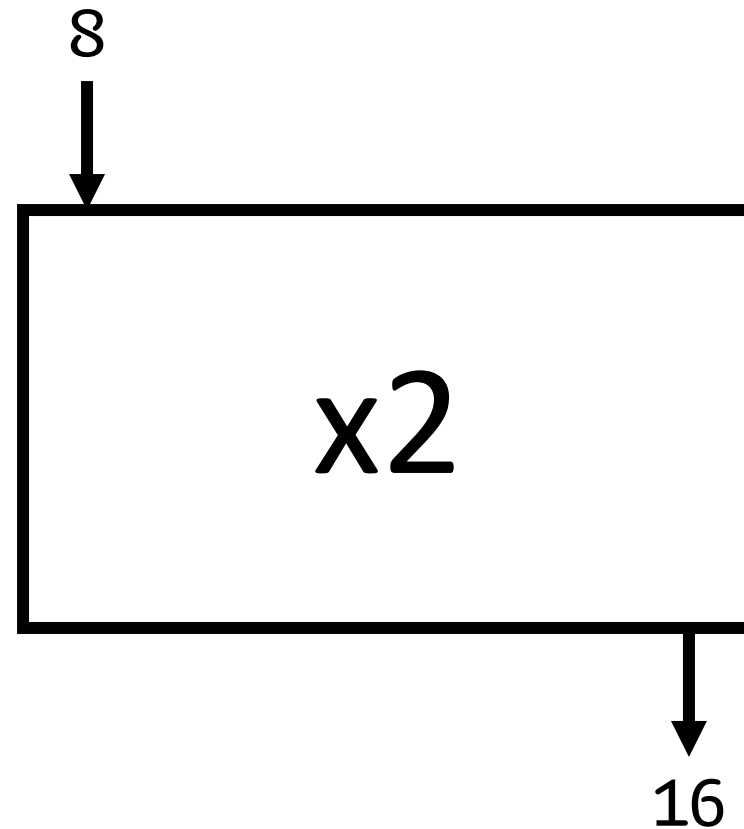
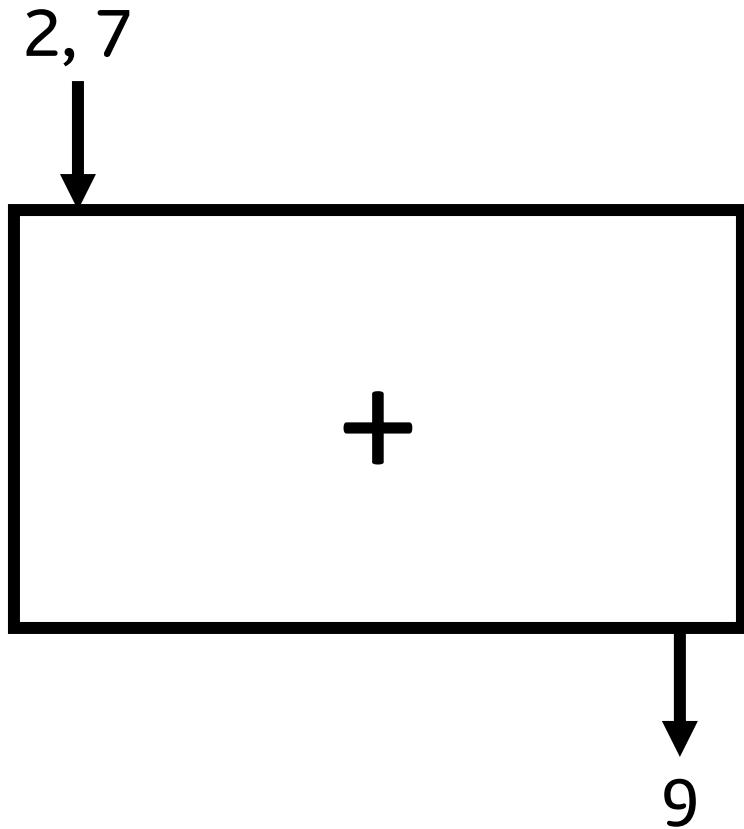


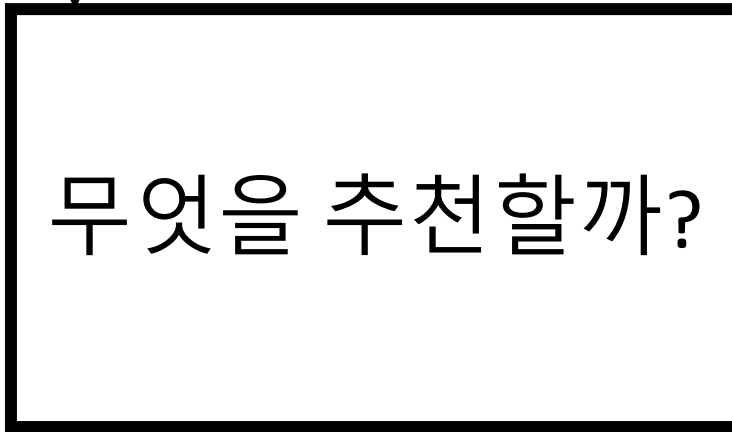
?

간단하게 추천 시스템을 만들어 보자.

추천시스템

- 함수의 개념과 비슷





사용자 정보 &
사용자 이력정보

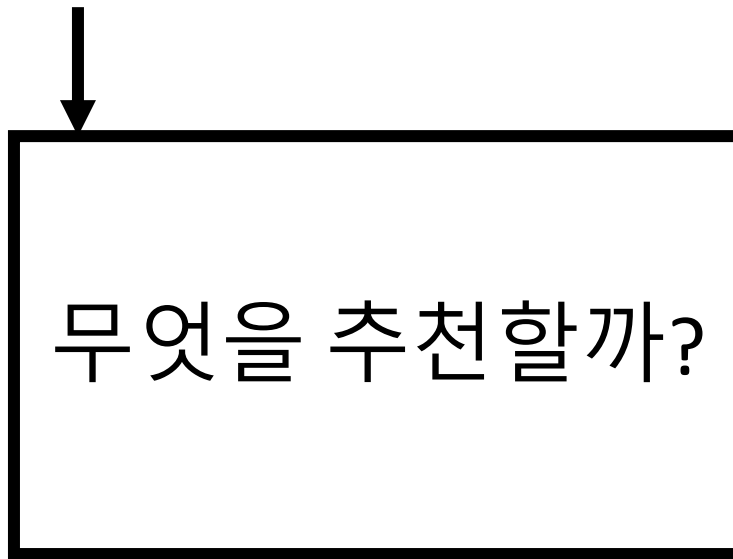


무엇을 추천할까?

1.  2.  3.  ...



추천할 수 있는 아이템 집합
& 아이템에 대한 정보

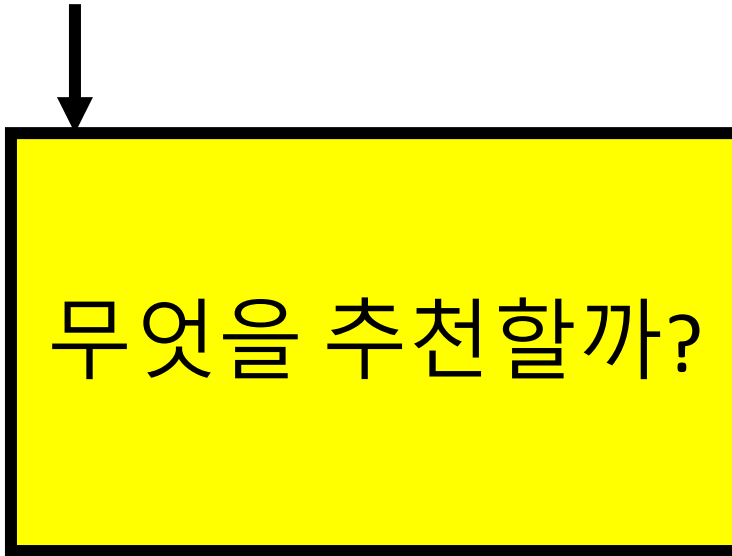




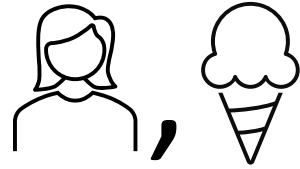
무엇을 추천할까?

사용자가 좋아할 만한 아이템 리스트

1. 
2. 
3.  ...



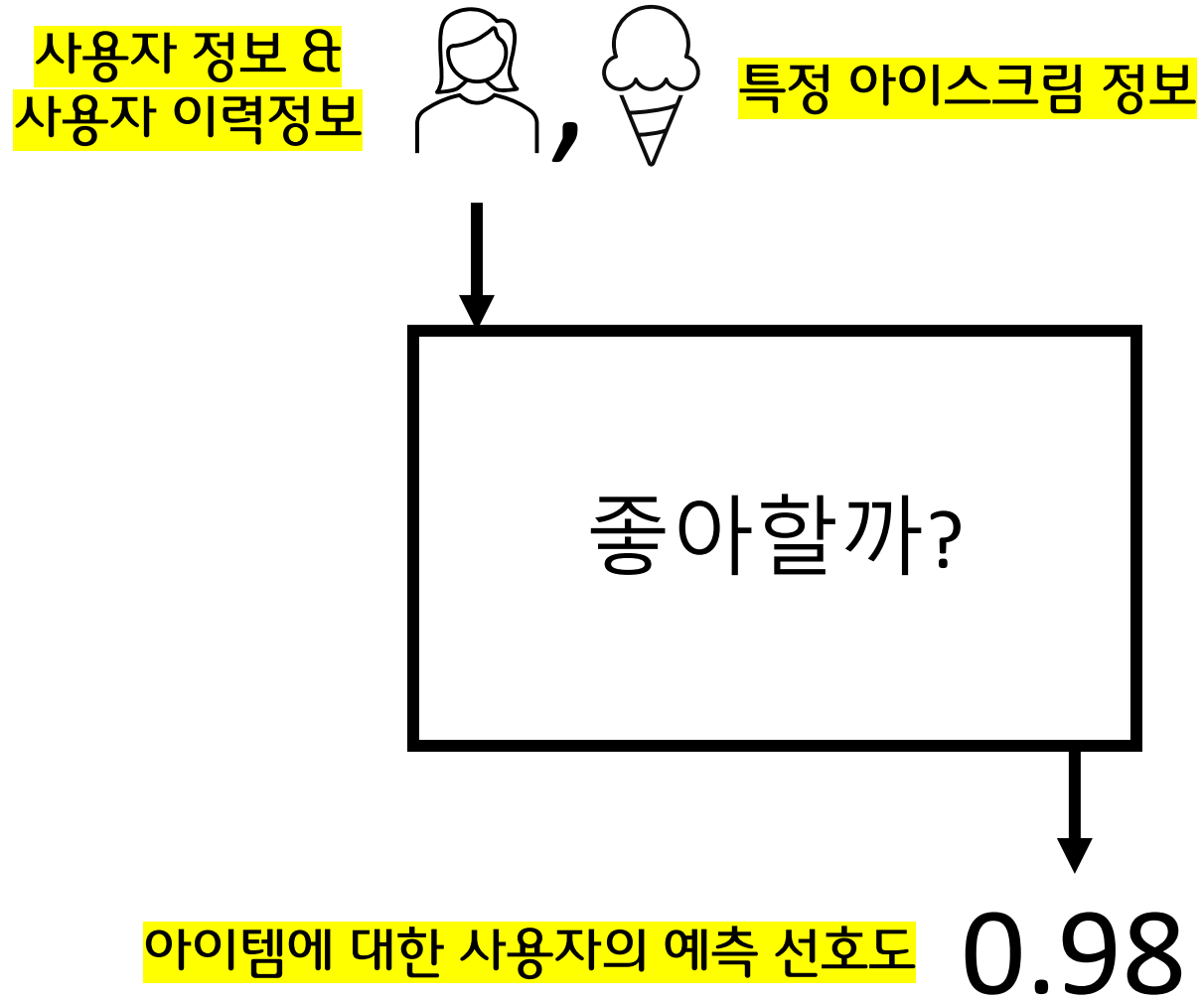
추천 시스템은
아이템에 대한 사용자의 선호도를 예측하는 시스템



좋아할까?



0.98



추천시스템

- 선호도 예측에 사용할 수 있는 알고리즘
 - Collaborative Filtering / Content-based Filtering
 - Click Through Rate Prediction
 - Sequential Recommendation

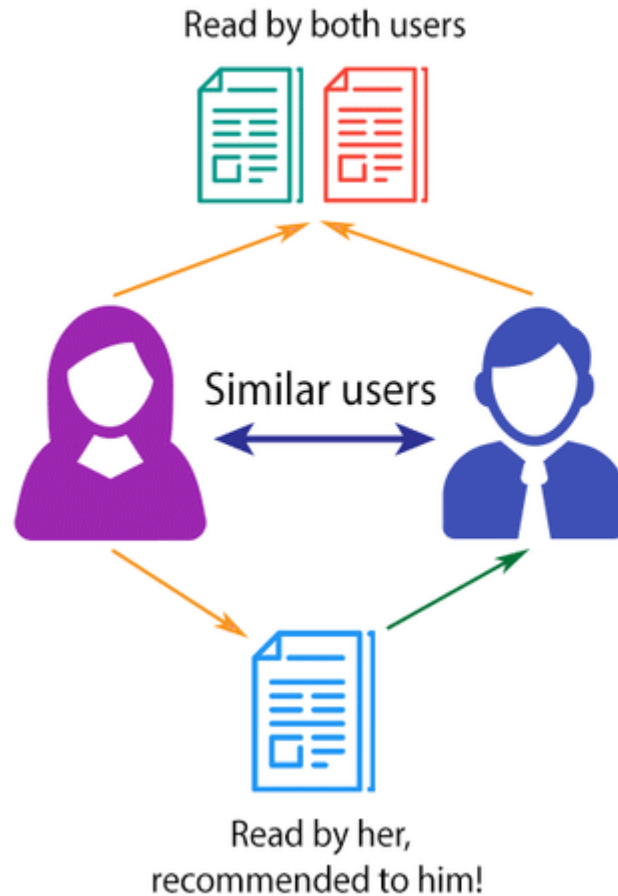
추천시스템

- 선호도 예측에 사용할 수 있는 알고리즘

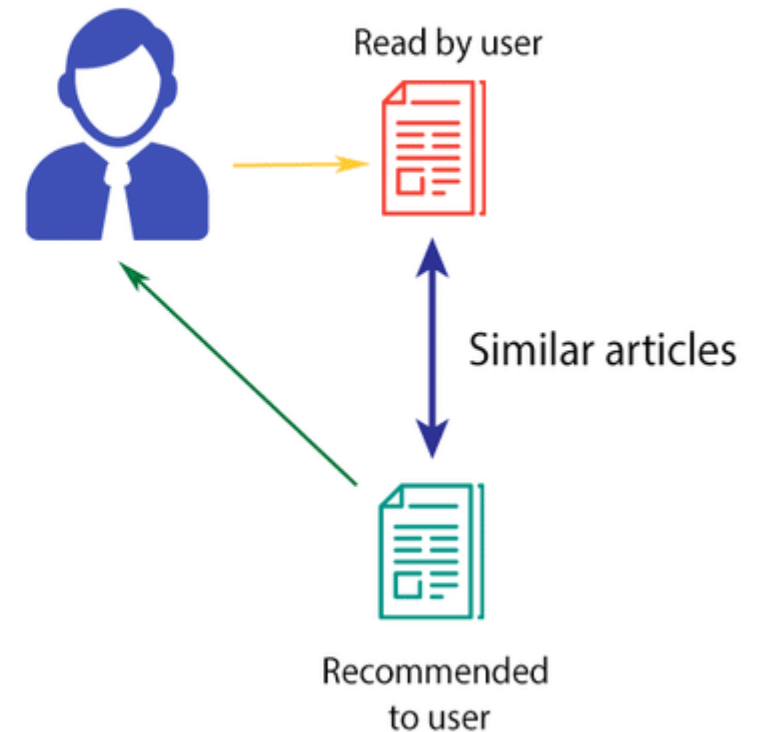
- **Collaborative Filtering / Content-based Filtering**

- Click Through Rate Prediction
- Sequential Recommendation

COLLABORATIVE FILTERING



CONTENT-BASED FILTERING



추천시스템










- 선호도 예측에 사용할 수 있는 알고리즘
 - **Collaborative Filtering** / **Content-based Filtering**
 - 나와 비슷한 취향의 사람들이 좋아하는 것은 나도 좋아할 가능성이 높다
 - 많은 사용자로부터 얻은 취향 정보를 활용
 - 사용자의 취향 정보 : 집단 지성
 - 축적된 사용자들의 집단 지성을 기반으로 추천
 - 예를 들어 A 상품을 구매한 사용자가 함께 구매한 다른 상품들 추천
 - Memory-based approach
 - 유사한 사용자나 아이템을 사용
 - 최적화 방법이나, 매개변수를 학습하지 않고 단순한 산술 연산만 사용
 - 쉽게 만들 수 있고 결과의 설명력이 좋을 뿐만 아니라 도메인에 의존적이지 않은 장점
 - 데이터가 축적이 안되고 sparse한 경우 성능이 낮고, 데이터가 너무 많아지면 속도가 저하되는 단점
 - Model-based approach
 - 기계학습을 통해 추천
 - Sparse한 데이터도 처리가능한 장점이 있으나 결과의 설명력이 낮은 단점이 있음

추천시스템

- 선호도 예측에 사용할 수 있는 알고리즘

- Collaborative Filtering

- Memory-based approach

		Item-based Similarity			
		 라면	 삼각김밥	 김치	 콜라
User-based Similarity	유저 A 	✓	✓	✓	✓
	유저 B 	✓	✓	✓	
	유저 C 	✓	✓		
	유저 D 	✓			
	유저 E 		✓		✓

추천시스템

- 선호도 예측에 사용할 수 있는 알고리즘
 - **Collaborative Filtering** / Content-based Filtering
 - Cold Start Problem : 한 사용자에게 대한 충분한 데이터가 부족한 경우 선호도 예측 불가능
 - First Rater : 새로운 아이템이 등장하여 평점 점수가 부족한 경우 추천 불가능
 - Grey Sheep Problem : 일관성이 없는 의견을 가진 사용자들의 데이터는 추천에 혼란을 줌
 - Shilling Attack : 악의적으로 평가 점수를 긍정/부정으로 입력하는 경우 추천에 방해가 됨

추천시스템

- 선호도 예측에 사용할 수 있는 알고리즘
 - Collaborative Filtering / Content-based Filtering
 - 콘텐츠 기반으로 분석하여 추천해주는 방식
 - 사용자가 관심 분야에 대해서 직접 입력한 정보나, 구매 내역, 평점 등을 기반
 - 다른 유저의 데이터가 필요하지 않고, 추천할 수 있는 아이템의 범위가 넓은 장점
 - 새로운 아이템이나 인기 없는 아이템도 추천 가능
 - 추천하는 이유를 제시할 수 있음
 - 적절한 feature를 찾기가 어려우며 새로운 유저를 위한 추천이 어려운 단점
 - 과도한 특수화 (over specialization)
 - 선호하는 특성을 가진 항목을 반복 추천함
 - 유저의 다양한 취향을 반영하기 어려움

추천시스템

- 선호도 예측에 사용할 수 있는 알고리즘
 - Click Through Rate (CTR) Prediction
 - 추천된 아이템을 유저가 클릭할 확률을 의미
 - 클릭률(CTR)은 광고 또는 무료 제품 등록정보를 본 사용자가 해당 광고를 클릭하는 빈도를 나타내는 비율
 - CTR을 사용하면 키워드 및 광고와 무료 등록정보의 실적을 파악 가능
 - CTR은 광고가 클릭된 횟수를 광고가 게재된 횟수로 나눈 값
 - $\text{클릭수} \div \text{노출수} = \text{CTR}$
 - 예를 들어 클릭수가 5회, 노출수가 100회인 경우 CTR은 5%
 - 대부분의 추천시스템은 CTR의 최대화를 목표로 하고 있음

추천시스템

- 선호도 예측에 사용할 수 있는 알고리즘
 - Click Through Rate (CTR) Prediction

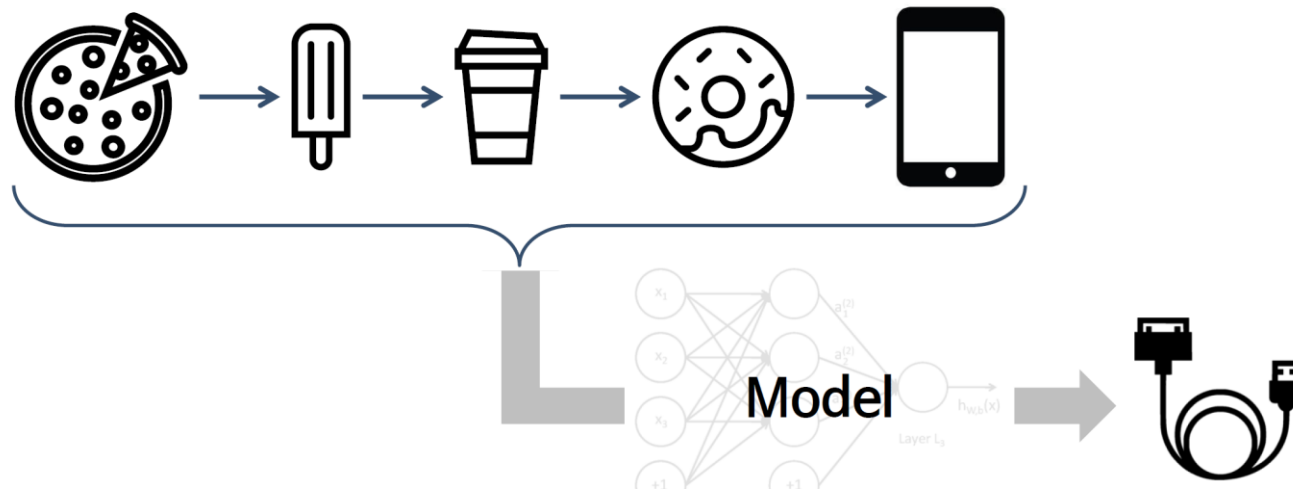
광고 번호	광고 위치	광고 카테고리	유저 디바이스	유저 나이	유저 성별	클릭 여부 (정답)
1	홈 지면	패션의류	아이폰	19	남자	1
2	검색 지면	디지털/가전	안드로이드	26	여자	0
3	홈 지면	기타	아이폰	22	남자	1
...
3203	검색 지면	패션의류	아이폰	42	여자	1

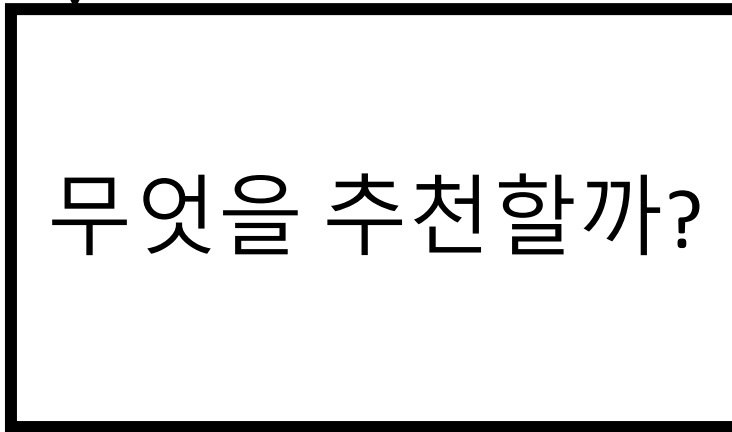
- CTR 예측을 진행하려면 기본적으로 노출 & 클릭 이력이 필요
- 실제로는 예측 모델을 만드는 과정보다 전처리 하는 과정에서 많은 시간을 소모
- 차원 축소 후 예측하는 경우가 많음

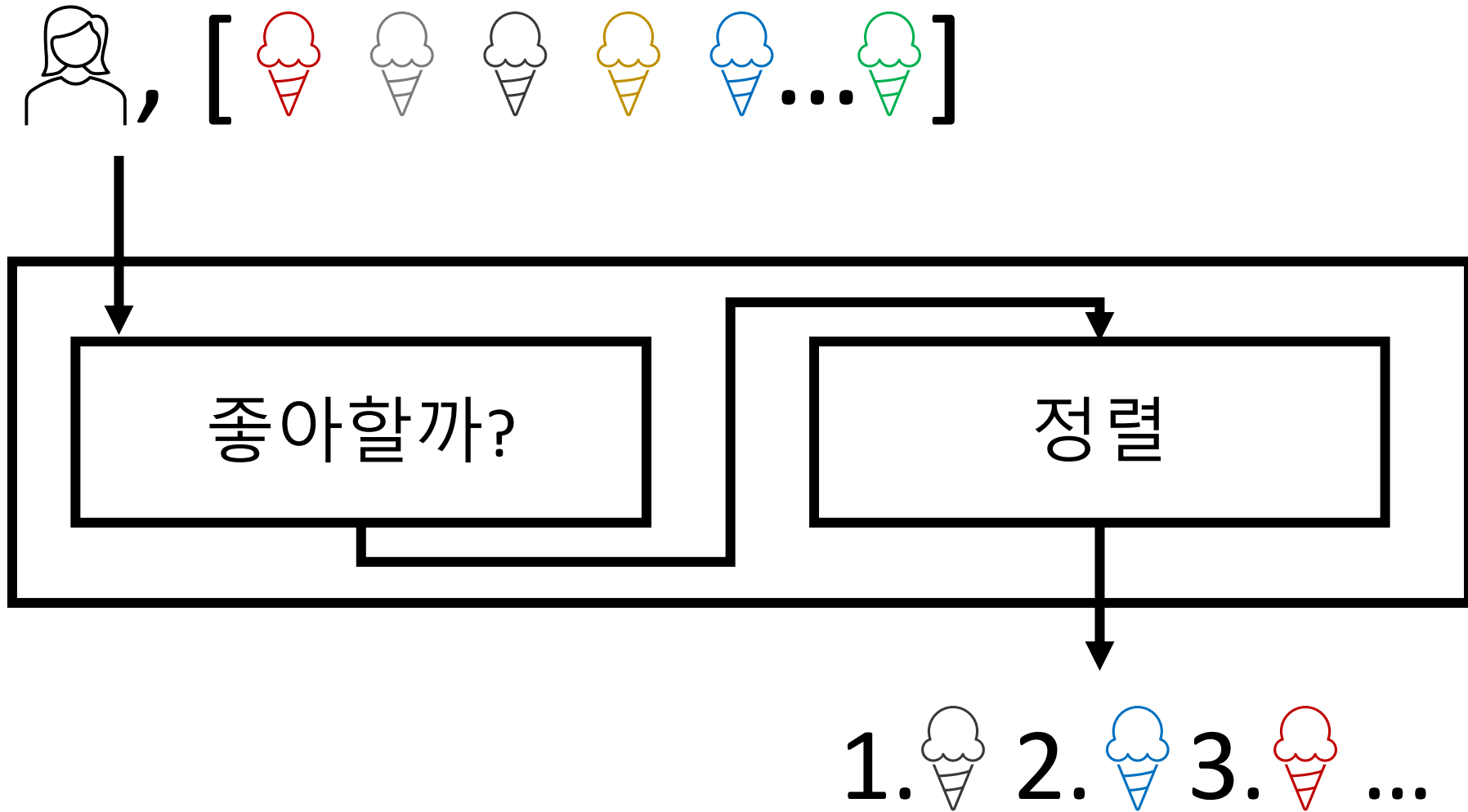
추천시스템

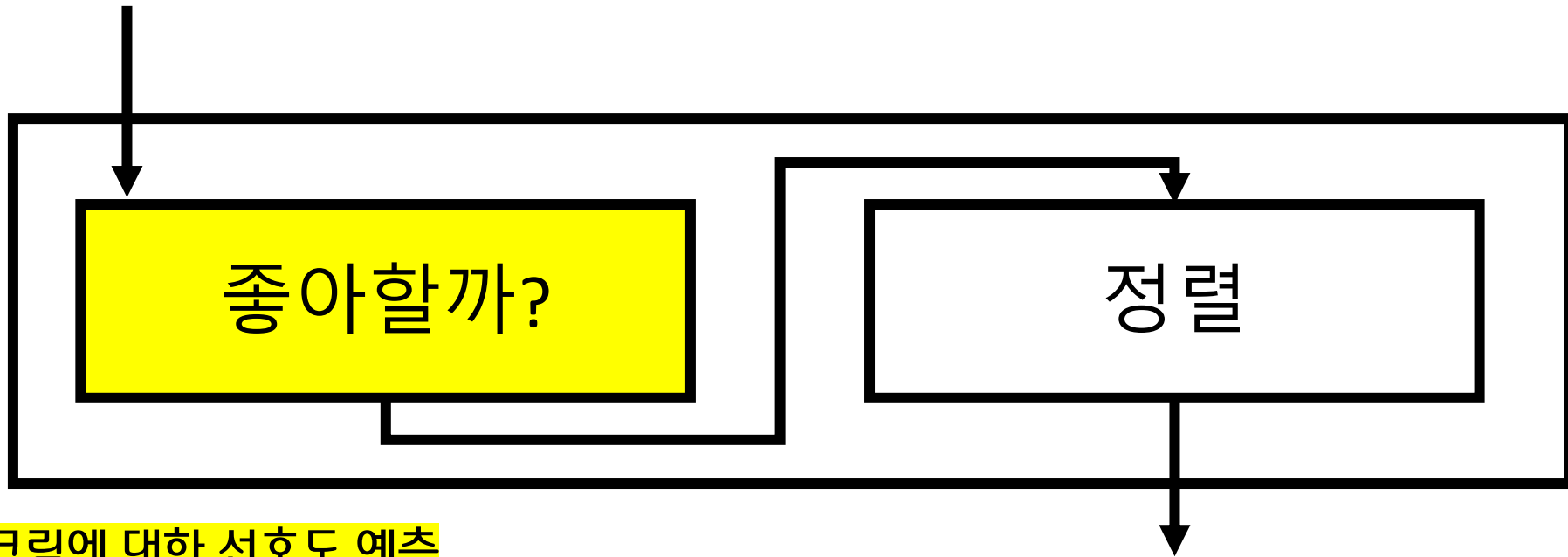
- 선호도 예측에 사용할 수 있는 알고리즘
 - Sequential Recommendation
 - 사용자의 과거 아이템 선택의 정보가 동일하게 중요하다는 기본 가정에서 출발한 추천시스템 알고리즘
 - 실제로 사용자가 선택을 할 때에는 과거 구매 정보가 동일하게 중요할까?
 - 사용자의 선호도, 관심은 끊임없이 변화하고 발전한다는 아이디어에서 시작
 - 조금 더 실제 사용자의 관심사를 잘 반영하는, 변화의 패턴까지도 잡아낼 수 있는 모델을 만들자.
 - 과거 행동에서 유의한 순차적인 패턴을 찾고, 최근 아이템에 더 집중하여 이를 기반으로 가장 선호할

다음 아이템을 추천







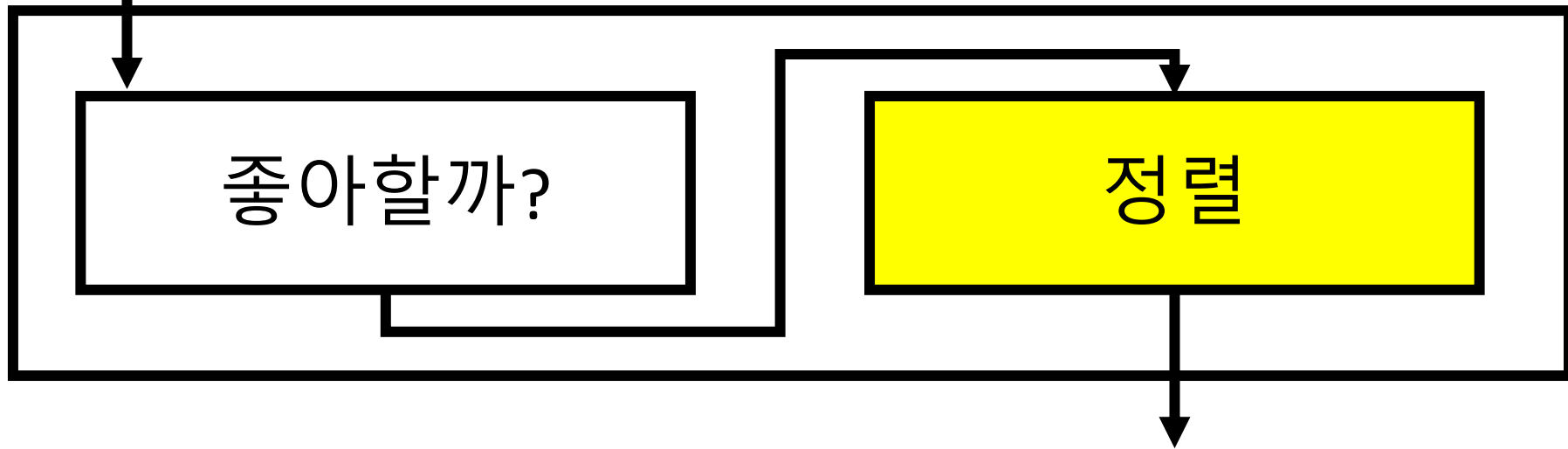


모든 아이스크림에 대한 선호도 예측

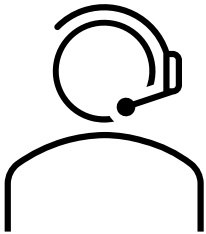
- 1.
- 2.
- 3.
- ...



가장 예측 선호도가 높은 아이스크림 추천 결과로 선정



1.  2.  3.  ...



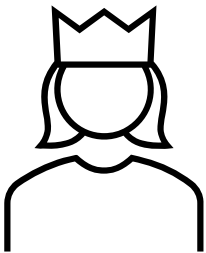
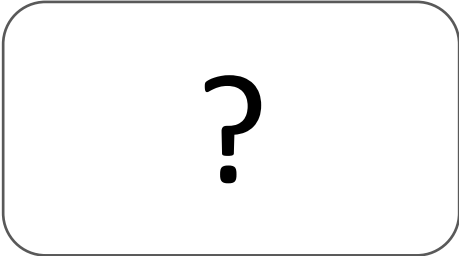
체리쥬빌레



엄마는외계인



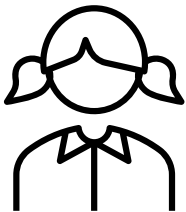
민트초콜릿칩



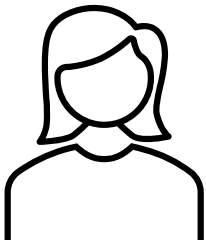
뉴욕치즈케이크



슈팅스타



아몬드붕붕

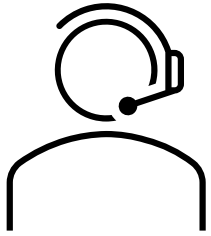


5월 18일

5월 19일

5월 20일

5월 21일(오늘)



체리쥬빌레




엄마는외계인



민트초콜릿칩

?

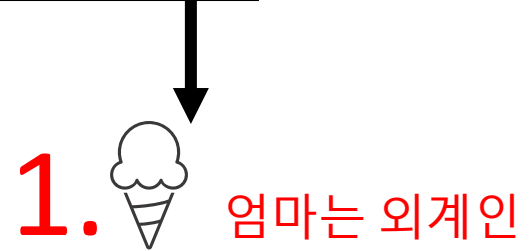
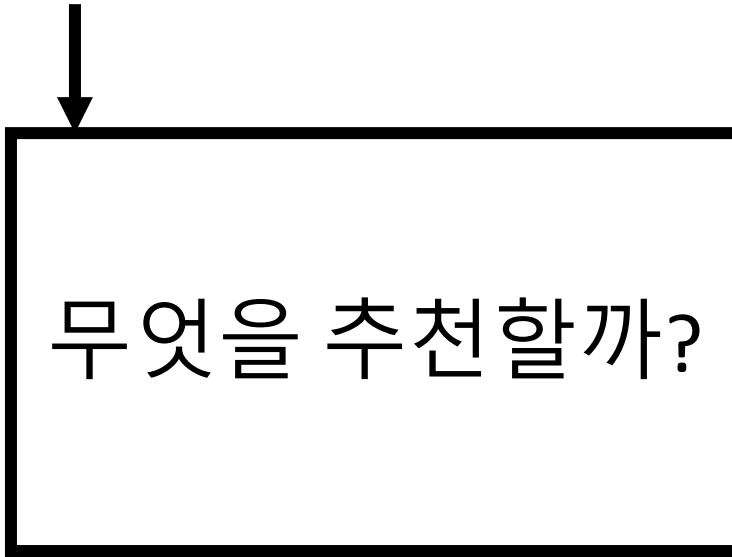
좋아할까? [ , ] = 0.7

좋아할까? [ , ] = 0.91 

엄마는외계인

좋아할까? [ , ] = 0.45

...



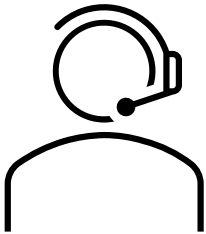
추천시스템

5월 18일

5월 19일

5월 20일

5월 21일(오늘)



체리쥬빌레



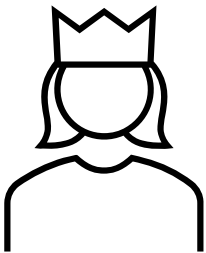
엄마는외계인



민트초콜릿칩



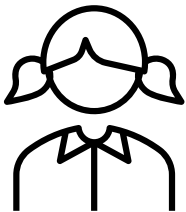
엄마는외계인



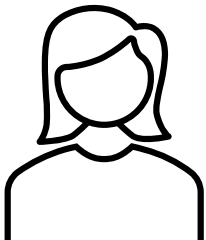
뉴욕치즈케이크



슈팅스타



아몬드붕붕

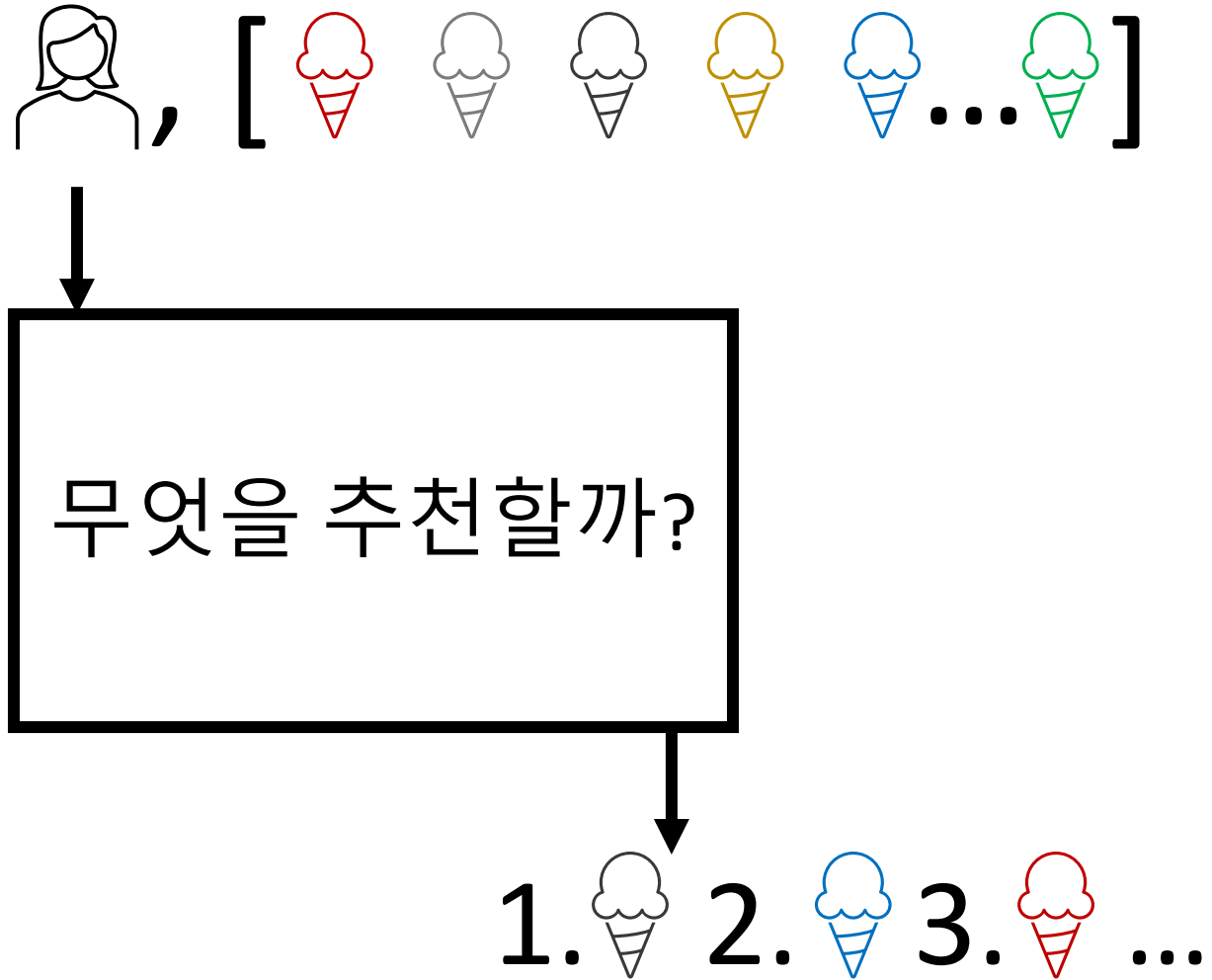


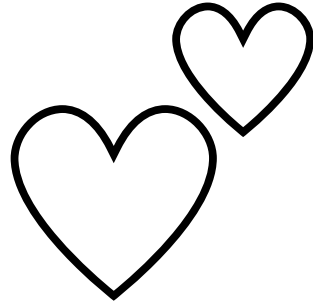
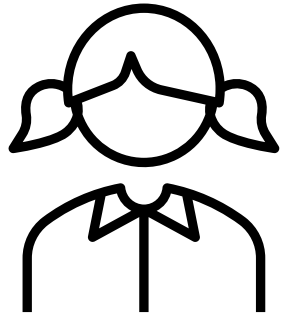
추천시스템

- 추천 시스템의 ‘성과’는 무엇일까?
 - 추천 시스템 성능이 좋다/나쁘다는 어떻게 측정할 수 있을까?
 - **단기 사용자 만족도**
 - 인당 클릭수: 사용자 하루 평균 클릭 수
 - 체류 시간: 사용자가 서비스에 머무는 시간

추천시스템

- 아이스크림 3개를 추천하려면?





첫 번째



아몬드붕붕

예측 선호도 점수 : 0.97

두 번째

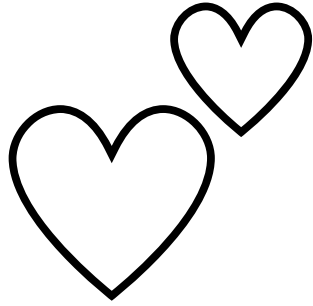
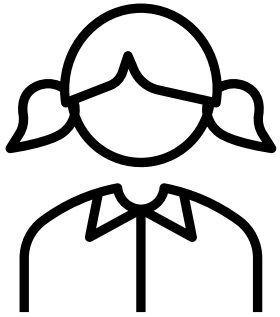
?

세 번째

?

추천시스템

- 단순히 예측 선호도가 높은 아이тем들로 채우면?



첫 번째



아몬드붕붕

예측 선호도 점수 : 0.97

두 번째



엄마는외계인

예측 선호도 점수 : 0.89

세 번째

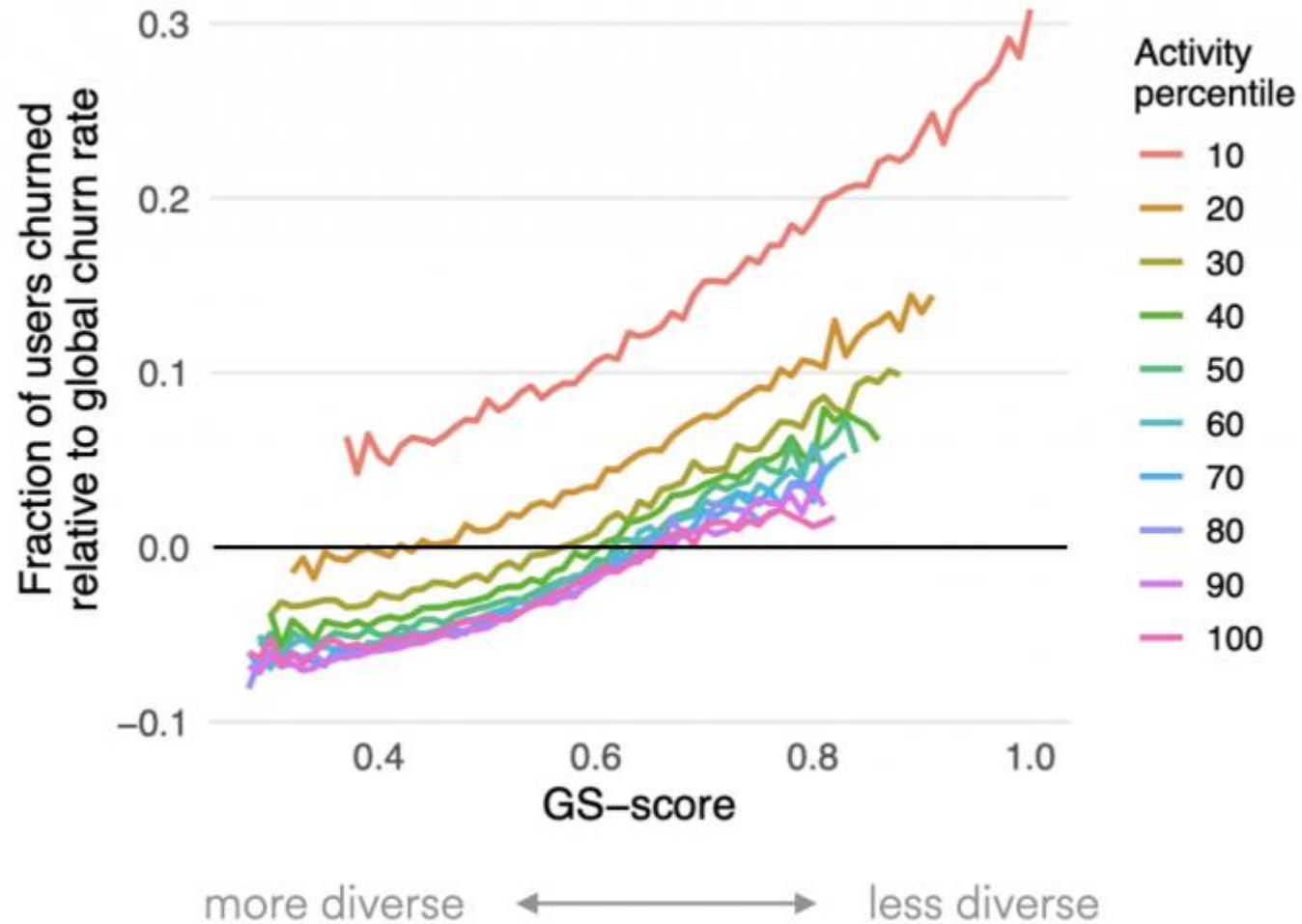


31올스타

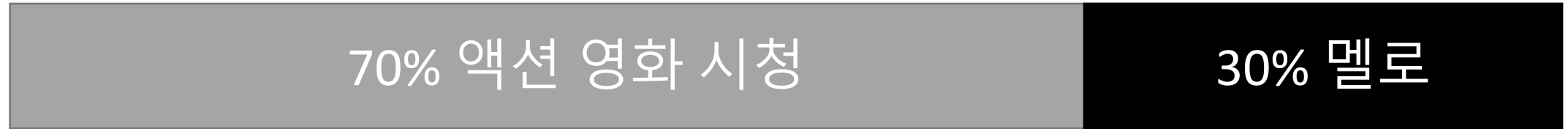
예측 선호도 점수 : 0.85

추천시스템

- 다양한 콘텐츠를 소비하는 사용자일수록 서서히 이탈율이 낮은 경향이 있음



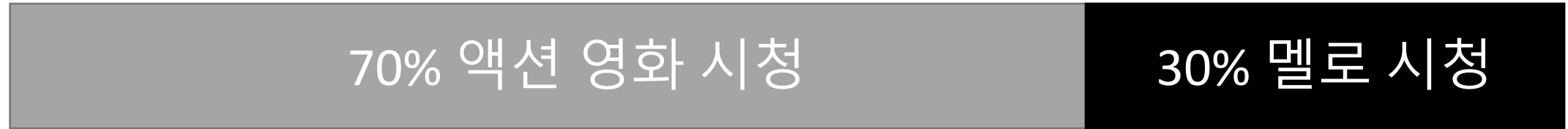
특정 사용자의 시청 이력



추천 결과 (시청 이력 데이터만 이용해서 추천 시)



특정 사용자의 시청 이력



추천 결과 (calibration 적용 시)



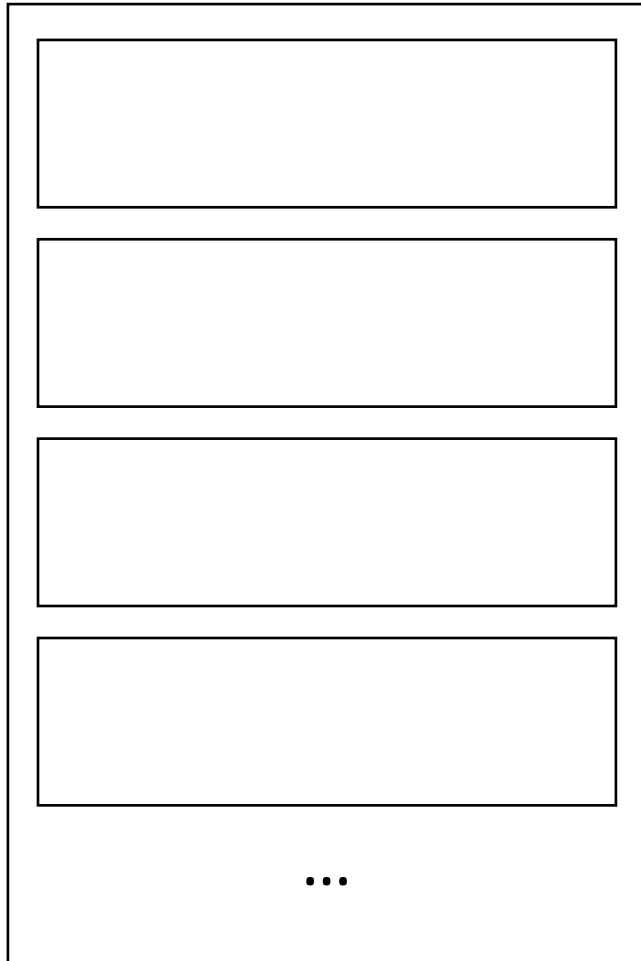
추천시스템

- 다양화 (Diversification)



추천시스템

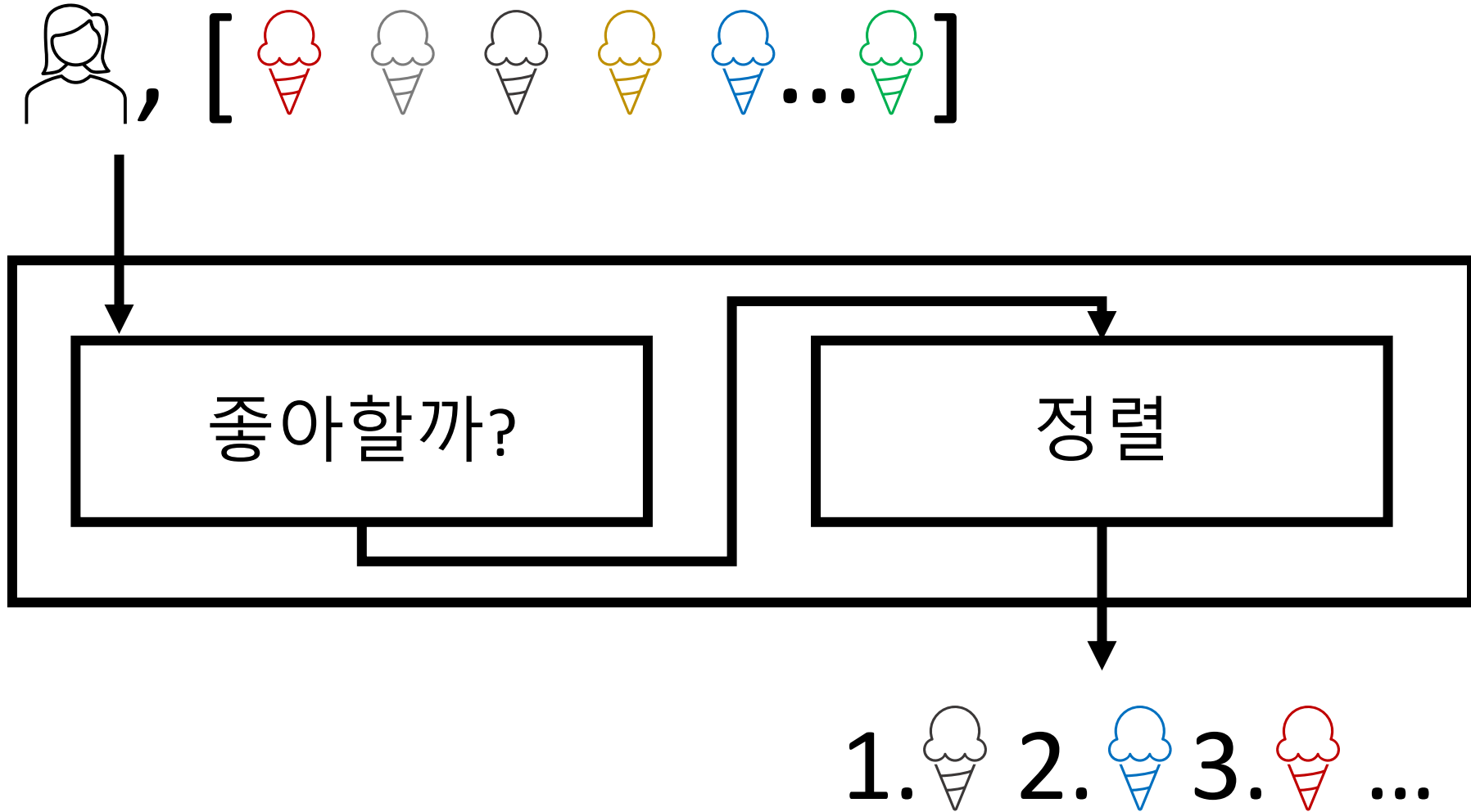
- 다양화 로직을 적용하면 클릭률이 떨어지지 않을까?
- 여러 개를 보여주고 스크롤 가능한 상황이라면 다양화 로직을 적용했을 때

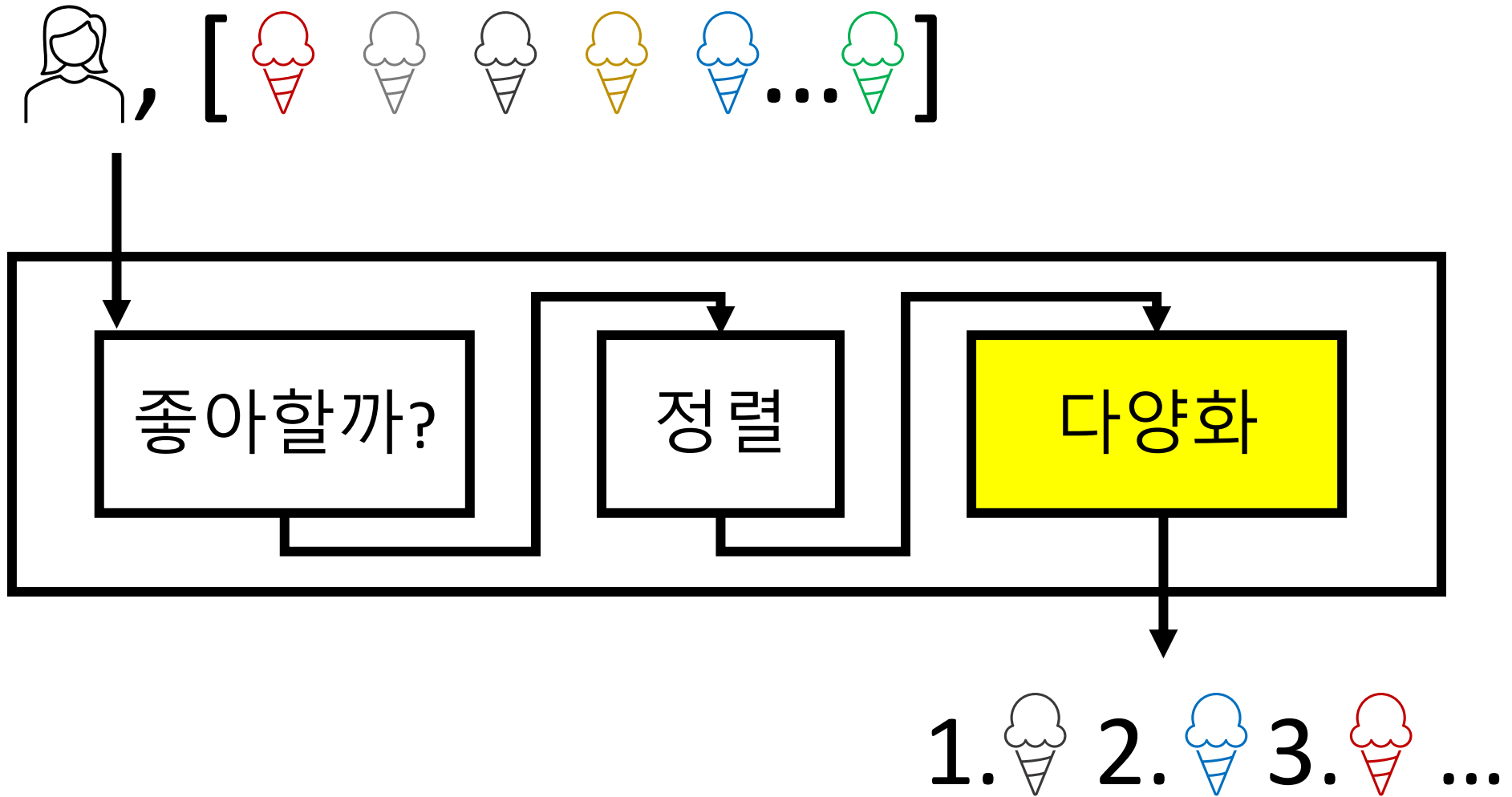


아이템별 클릭 률 ↓

스크롤 수 ↑

인당 클릭 수 ↑

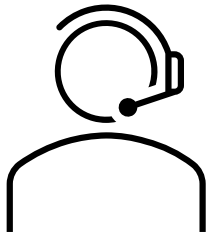




추천시스템

- 이 사람이 내일 또 올 때, 어떤 아이스크림을 추천해야 될까?

5월 18일



체리쥬빌레

5월 19일



엄마는외계인

5월 20일



민트초콜릿칩

5월 21일



엄마는외계인

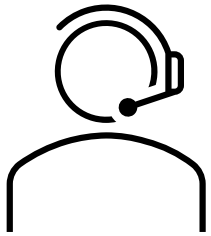
5월 22일(오늘)

?

추천시스템

- 이 사람이 내일 또 올 때, 어떤 아이스크림을 추천해야 될까?
- 이 사람에게 매일 초코 아이스크림만 추천해도 될까?

5월 18일



체리쥬빌레

5월 19일



엄마는외계인

5월 20일



민트초콜릿칩

5월 21일



엄마는외계인

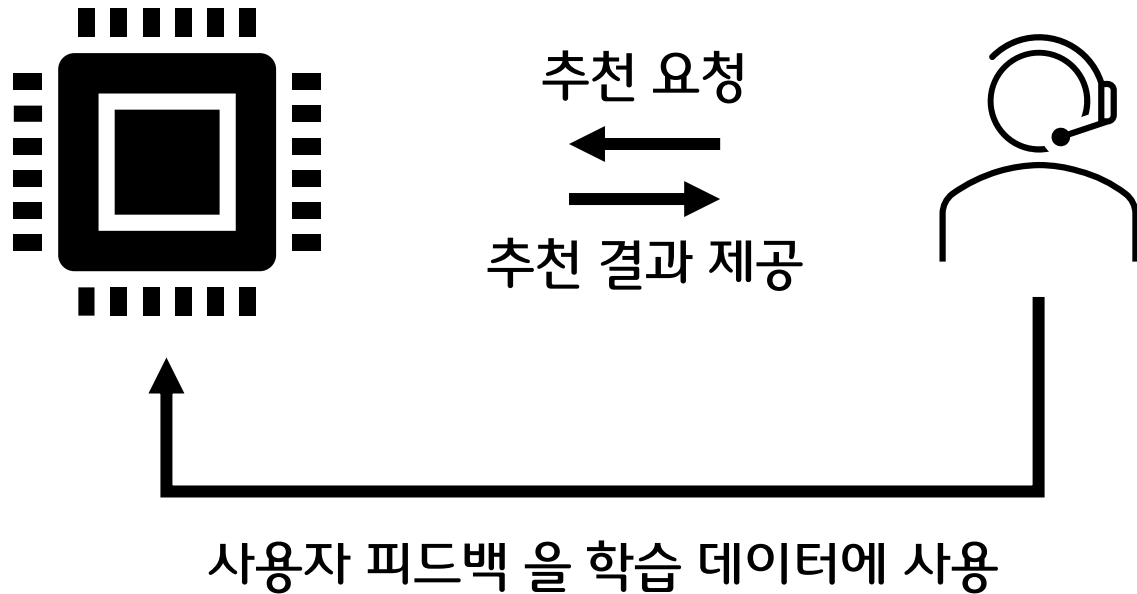
5월 22일(오늘)



아몬드붕붕

추천시스템

- 피드백 루프 (Feedback Loop)
 - 특정 시스템의 결과물 일부 혹은 전부가 미래의 학습 데이터로 사용 되는 구조를 가질 경우
 - 필터 버블 이슈: 개인화 추천 엔진이 유저의 과거 이력에 기반한 추천 결과를 내보내면서 유저가 새로운 유형/견해의 콘텐츠에 노출되지 않고 특정 유형/견해를 갖는 콘텐츠에 고립되는 현상 발생



단기적으로 사용자들이 만족하더라도
소비 다양성이 줄어들면 서비스 이탈율은 높아짐



추천시스템은 사용자의 소비 다양성을 높여서
장기적인 서비스 만족도를 높이는 방향으로
발전하고 있음

추천 플랫폼의 장기적인 목표는
사용자가 현재 원하는 것을 충족시켜주는 것뿐만 아니라
미래에 그들이 더 자주 추천 플랫폼에 방문하도록 만드는 것

Google, 2021

페이스북에서 가장 중요하게 생각하는 가치는
사용자들에게 장기적인 가치 (long-term value)를 주는 것

Facebook, 2021

감사합니다

kimtwan21@dongduk.ac.kr

김 태 완