

특화 프로젝트

빅데이터

맛집 추천 시스템

A405

구영지 이용석 이재욱 한우석 홍지희

목차

01

빅데이터란?

빅데이터
이해하기

02

프로젝트 소개

프로젝트 주제, 목표
그리고 개요

03

추천 시스템

협업 필터링
컨텐츠 기반 필터링

04

향후 계획

파트별
작업 계획

CHAPTER. 1

빅데이터란?

- 벡터이터 이해하기

01




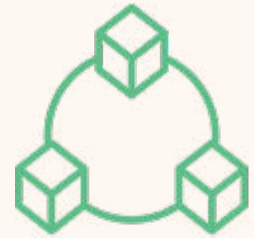
01. 빅데이터란?

빅데이터 이해하기

The Amazon logo, featuring the word "amazon" in a bold, black, lowercase sans-serif font. Below the text is a thick, orange curved arrow that starts under the 'a' and points towards the 'n', resembling a smile.

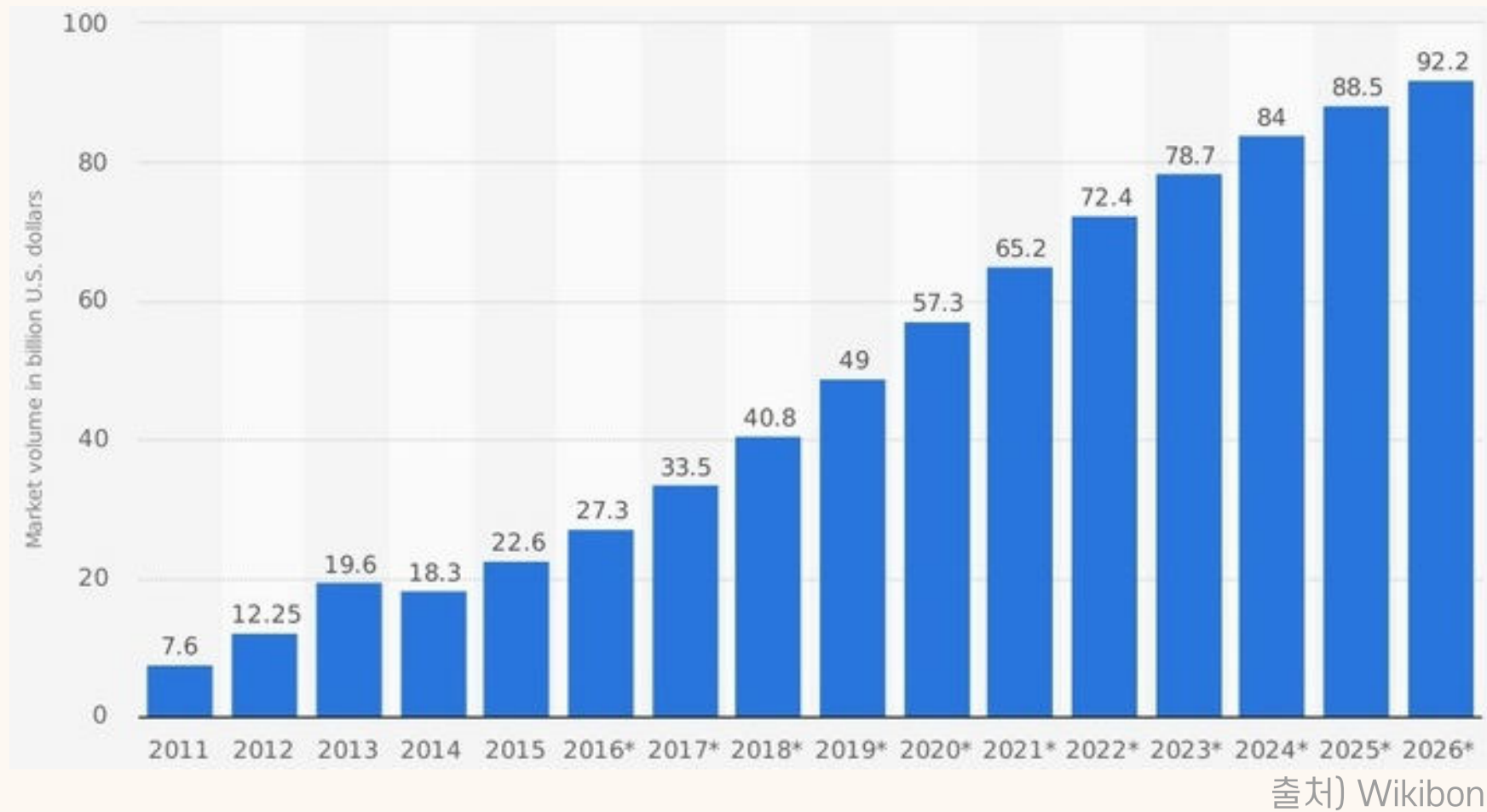
빅데이터 이해하기

 가승우 · 1주 전(수정됨)
오늘도 알수없는 유튜브 알고리즘은 나를
이곳으로 불렀다...

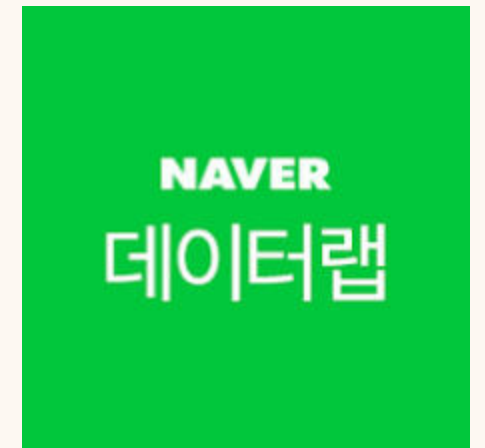
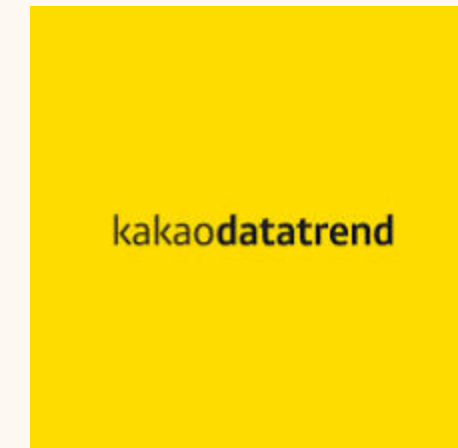


빅데이터

: 대규모의 데이터. 데이터로부터 가치를 추출하고 결과를 분석하는 기술



세계 빅데이터 시장 규모 전망 (단위: 10억 달러)



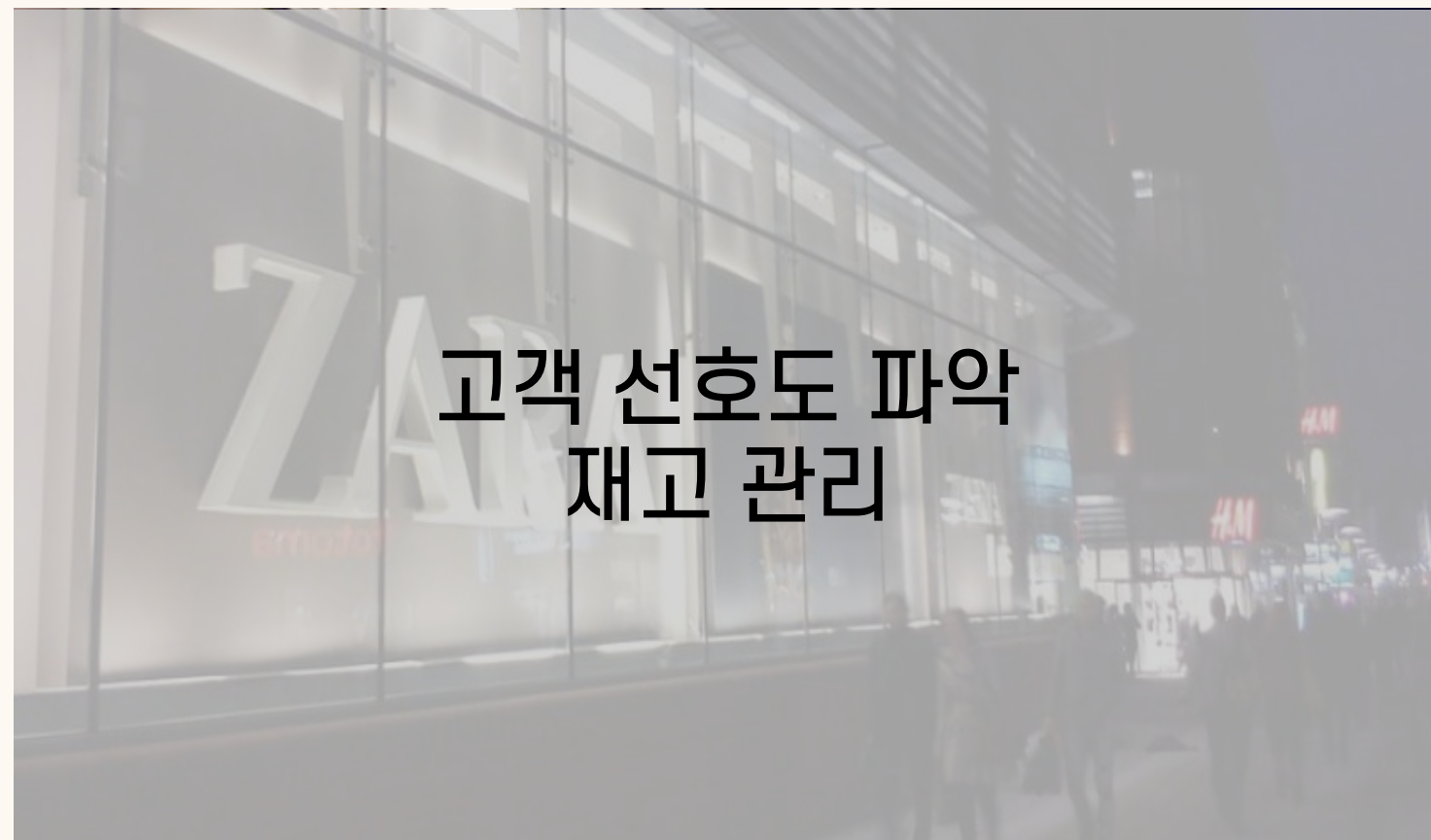
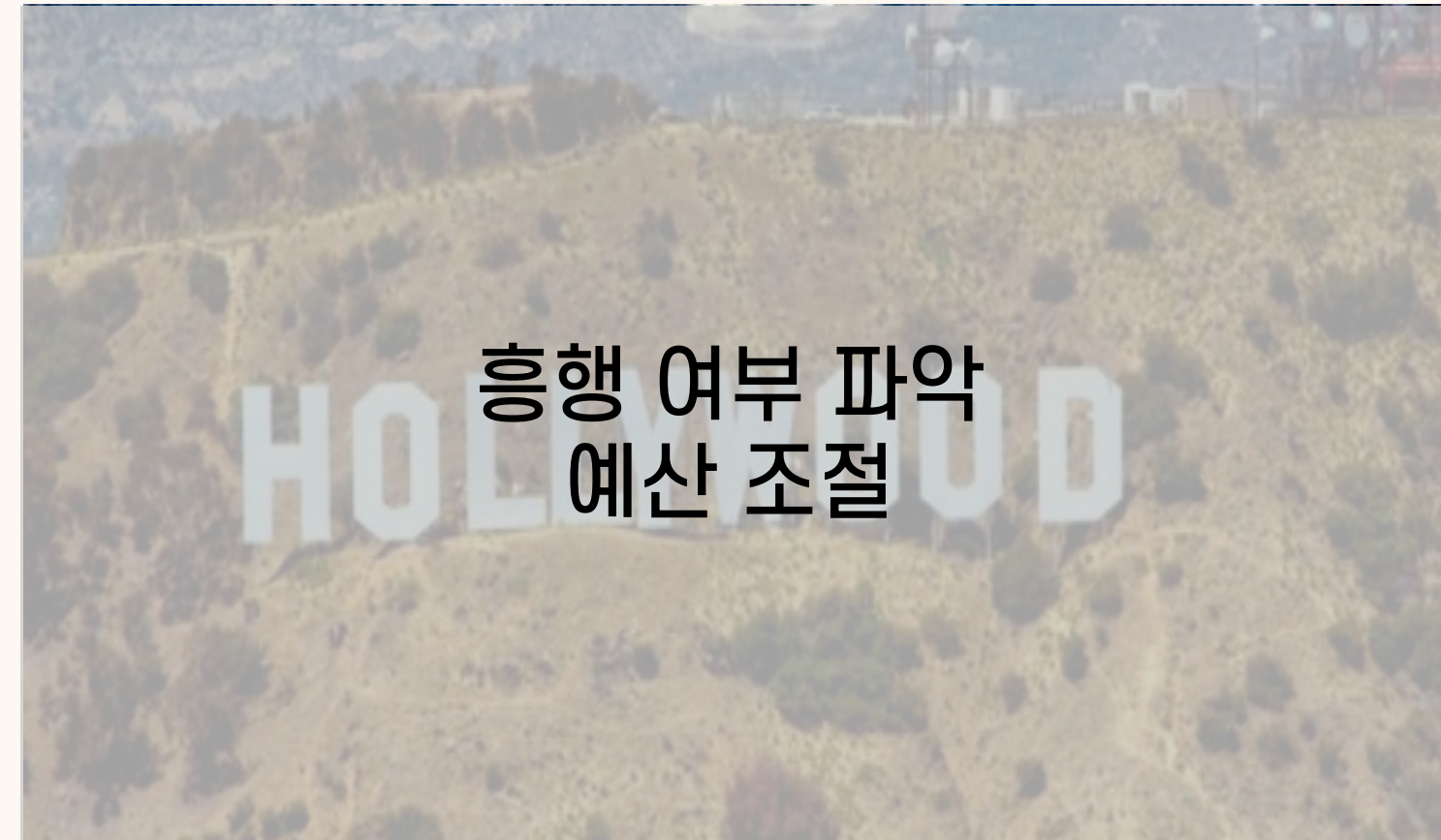
데이터센터 설립 붐...시장 선점 위해 적과도 손잡는다

디지털전환 핵심 '데이터센터'...'3년 내 24곳' 추가로 세운다

발행일 : 2021.03.16 16:00

01. 빅데이터란?

빅데이터 이해하기





02

CHAPTER. 2

프로젝트 소개

- 프로젝트 주제
- 프로젝트 목표
- 프로젝트 개요



사용자 리뷰 기반의 맛집 추천 서비스

데이터 전처리
추천 알고리즘 구현

+

차별화된 UI/UX

||

개인별 맞춤 맛집 추천 서비스





데이터 분석 및 정리

데이터 가공 및 저장
사용자-아이템 행렬 생성



기본 추천 시스템 구축

협업 필터링
컨텐츠 기반 필터링



서비스 고도화

하이브리드 추천 시스템 구축
성능측정 모니터링 시스템



마이그레이션(Migration)?

: 현재 사용하는 운영환경에서 더 나은 다른 운영환경으로 시스템을 전환하는 과정

ex) 윈도우 -> 리눅스



데이터베이스(Database) 마이그레이션(Migration)

: 한 종류의 데이터베이스에서 다른 종류의 데이터베이스로 데이터를 옮기는 것을 의미

02. 프로젝트 소개

프로젝트 개요

```
category_list": [{"category": "카페"}], "menu_list": [], "bhour_list": [], "review_cnt": 0, "review_list": [], {"id": 3, "name": "Battered Sole", "branch": null, "area": "이태원", "tel": "02-749-6867", "address": "서울특별시 용산구 이태원동 118-9", "latitude": "37.535032", "longitude": "126.991664", "category_list": [{"category": "피쉬앤칩스"}, {"category": "편"}, {"menu_list": [{"menu": "메인 (피쉬앤칩스, 오리 가슴살 등)", "price": 14000}, {"menu": "스터터 보드(피테 등)", "price": 18000}, {"menu": "디저트 (초콜렛 풍당 등)", "price": 8500}, {"menu": "생과일 캥데일", "price": 9000}, {"menu": "생맥주", "price": 7000}], "bhour_list": [{"type": 1, "week_type": 1, "mon": 0, "tue": 0, "wed": 0, "thu": 0, "fri": 0, "sat": 0, "sun": 1, "start_time": "17:30:00", "end_time": "23:30:00", "etc": ""}, {"type": 3, "week_type": 1, "mon": 1, "tue": 0, "wed": 0, "thu": 0, "fri": 0, "sat": 0, "sun": 0, "start_time": "00:00:00", "end_time": "00:00:00", "etc": ""}, {"type": 1, "week_type": 1, "mon": 0, "tue": 1, "wed": 1, "thu": 1, "fri": 0, "sat": 0, "sun": 0, "start_time": "17:30:00", "end_time": "23:30:00", "etc": ""}, {"type": 1, "week_type": 1, "mon": 0, "tue": 0, "wed": 0, "thu": 0, "fri": 1, "sat": 1, "sun": 0, "start_time": "17:30:00", "end_time": "24:00:00", "etc": ""}], "review_cnt": 0, "review_list": [], {"id": 4, "name": "Chakyounge", "branch": null, "area": "달맞이길", "tel": "051-756-5566", "address": "부산광역시 해운대구 중2동 1509-5", "latitude": "35.158587", "longitude": "129.175004", "category_list": [{"category": "레스토랑"}, {"category": "카페"}, {"menu_list": [], "bhour_list": [], "review_cnt": 0, "review_list": [], {"id": 5, "name": "Delabobo", "branch": null, "area": "발산역", "tel": "02-2667-9854", "address": "서울특별시 강서구 등촌동 689", "latitude": "37.559904", "longitude": "126.840512", "category_list": [{"category": "디저트카페"}, {"category": "디저트"}, {"menu_list": [], "bhour_list": [], "review_cnt": 0, "review_list": [], {"id": 6, "name": "Gogoss Pub", "branch": null, "area": "하남", "tel": "031-792-9289", "address": "경기도 하남시 덕풍동 412-18", "latitude": "37.539046", "longitude": "127.203933", "category_list": [{"category": "편"}, {"menu_list": [{"menu": "크리스피류라이드치킨", "price": 16000}], "bhour_list": [], "review_cnt": 0, "review_list": [], {"id": 7, "name": "Madeleine", "branch": null, "area": "만촌동", "tel": "053-759-7947", "address": "대구광역시 수성구 만촌1동 1356-7", "latitude": "35.869088", "longitude": "128.636058", "category_list": [{"category": "마들렌"}, {"category": "베이커리"}, {"menu_list": [{"menu": "레몬향", "price": 3500}, {"menu": "아메리카노", "price": 3000}, {"menu": "핫초코", "price": 3500}], "bhour_list": [{"type": 1, "week_type": 1, "mon": 1, "tue": 1, "wed": 1, "thu": 1, "fri": 1, "sat": 1, "sun": 1, "start_time": "09:00:00", "end_time": "23:00:00", "etc": ""}], "review_cnt": 0, "review_list": [], {"id": 8, "name": "MesCafe", "branch": null, "area": "남포동", "tel": "051-243-1125", "address": "부산광역시 중구 광복동2가 25", "latitude": "35.100110", "longitude": "129.032478", "category_list": [], "menu_list": [], "bhour_list": [], "review_cnt": 0, "review_list": [], {"id": 9, "name": "Red Camp", "branch": null, "area": "진주", "tel": "070-7786-9258", "address": "경상남도 진주시 평거동 792-3", "latitude": "35.170957", "longitude": "128.064554", "category_list": [{"category": "고깃집"}, {"category": "고기집"}, {"menu_list": [], "bhour_list": []}]
```

stores										
	id	store_name	branch	area	tel	address	latitude	longitude	category	image
0	1	Agal	None	종대	010-6689-5886	서울특별시 마포구 동교동 170-13	37.556062	126.926666	아구찜 포장마차	
1	2	Assisy	None	광주	062-367-0700	광주광역시 서구 농성동 631-33	35.150746	126.890062	카페	
2	3	Battered Sole	None	이태원	02-749-6867	서울특별시 용산구 이태원동 118-9	37.535032	126.991664	피쉬앤칩스 밥	
3	4	Chakyounge	None	달맞이길	051-756-5566	부산광역시 해운대구 중2동 1509-5	35.158587	129.175004	레스토랑 카페 디저트	
4	5	Delabobo	None	발산역	02-2667-9854	서울특별시 강서구 등촌동 689	37.559904	126.840512	디저트카페 디저트	
...
459983	459996	팔복부대찌개	None	죽양동	031-877-2410	경기도 의정부시 죽양동 154-6	37.757939	127.032758		
459984	459997	팔복부대	None	안중읍	031-682-1733	경기도 양주시 안중읍 용성리 718-3	37.002743	126.944991		
459985	459998	팔복분식	None	화천면	031-8071-2668	경기도 양주 화천면 도수리 611-6	37.471677	127.315744		
459986	459999	팔복분식	None	소주동	055-312-3713	경상남도 양산시 소주동 147-1	35.407860	129.158564		
459987	460000	팔복분식	None	의정부동	031-772-6624	경기도 의정부시 의정부동 479-7	37.739845	127.038074		
459988 rows × 10 columns										
reviews										
	store_id	user_id	score	content				reg_time		
0	15	68632	5	전포 찰밥에 새콤 상간! 포주에서 온 웨프가 직접 요리하는 포주식 레스토랑!				1970-01-01 00:00:00		
1	18	389728	5	샌드위치 내용물이 알차게 들어있고 맛있었어요 가성비 추천				1970-01-01 00:00:00		
2	19	68716	4	홍콩러스 1층 매장 푸드코트 내 있는 매장에서 제한된 음식은 저를보고 시켜봤는데 사진...				1970-01-01 00:00:00		
3	37	774353	2	전 여기 5년전에 가봤었는데 그때 기억은 별로였어요!난채순님이 맛있고, 자려지는건...				1970-01-01 00:00:00		
4	38	115682	3	친구들끼리 술 간단하게마시러 갈..스끼다시 권장자인 회 알아 조금 부족한 느낌..맛...				2019-03-15 22:16:47		
...		
91393	360499	17371	5	완월이 소스가 매운데 중독성있는 맛입니다. 마지막에 볶음밥 꼭드세요				1970-01-01 00:00:00		
91394	360505	198050	4					1970-01-01 00:00:00		
91395	360514	190766	5	완월이 소스가 매운데 중독성있는 맛입니다. 마지막에 볶음밥 꼭드세요				1970-01-01 00:00:00		
91396	360514	201564	4	고기가 엄청많이 먹고싶을 때 중독 될리는 곳이에요!~파워처럼 살러드바에 사이드며뉴가...				1970-01-01 00:00:00		
91397	360515	811078	5					1970-01-01 00:00:00		
91398 rows × 5 columns										

store									
	id	name	address	area	branch	category	area	branch	category
1	1	서울특별시 마포구 동교동 170-13	종대	[NULL]	아구찜 포장마차				
2	2	광주광역시 서구 농성동 631-33	광주	[NULL]	카페				
3	3	서울특별시 용산구 이태원동 118-9	이태원	[NULL]	피쉬앤칩스 밥				
4	4	부산광역시 해운대구 중2동 1509-5	달맞이길	[NULL]	레스토랑 카페 디저트				
5	5	서울특별시 강서구 등촌동 689	발산역	[NULL]	디저트카페 디저트				
6	6	경기도 하남시 덕풍동 412-18	하남	[NULL]	밥				
7	7	대구광역시 수성구 만촌1동 1356-7	만촌동	[NULL]	마들렌 베이커리				
8	8	부산광역시 중구 광복동2가 25	남포동	[NULL]					
9	9	경상남도 진주시 평거동 792-3	진주	[NULL]	고깃집 고기집				
10	10	서울특별시 서대문구 창천동 30-33	신촌	[NULL]	중식당 레스토랑				
11	11	경기도 서울시 팔곡동 1006-3	팔곡	[NULL]	국수 고기국수				
12	12	서울특별시 강남구 역삼동 644-10	역삼역	[NULL]	역삼역점				
13	13	부산광역시 수영구 광안동 116-4	수영역	[NULL]	cafe				
14	14	대전광역시 유성구 상대동 469-4	유성	[NULL]	오정 방앗간				
15	15	부산광역시 부산진구 전포동 229-13번지 2층	전포동	[NULL]	호주 레스토랑				
16	16	서울특별시 용작구 사당동 1003-28	사당	[NULL]	에스프레소				
17	17	서울특별시 광진구 자양동 11-11	건대	[NULL]					
18	18	전라북도 진주시 완산구 룡자동1가 652 상산고등학교	진주	[NULL]	샌드위치 주스				
19	19	경기도 화성시 향남읍 향정리 487-1	홍남리	[NULL]	향남점				
20	20	경기도 용인시 기흥구 서현동 9-3	서현동	[NULL]					
21	21	경기도 안성시 신건지동 16-11	신건지동	[NULL]	카페 분류원 카페				
22	22	서울특별시 서초구 방배동 480-3	방배동	[NULL]					
23	23	경기도 의왕시 내손동 622-11	내손동	[NULL]					
24	24	서울특별시 서초구 방배동 480-3	사당	[NULL]	샌드위치 연어샌드위치				
25	25	경상남도 창원시 외창구 용호동 18-12	가로수길	[NULL]	핸드드림 카페				
26	26	경기도 성남시 분당구 율동 318	율동	[NULL]					
27	27	경기도 안산시 상록구 사동 1475	사동	[NULL]					
28	28	경기도 용인시 처인구 포곡읍 금어리 106-1	처인구	[NULL]					
29	29	서울특별시 용산구 한남동 743-29	한남동	[NULL]	bar				
30	30	경기도 화성시 남양읍 남양리 2070-9	화성남양	[NULL]	고기집				
31	31	경기도 부천시 막대동 144-45	막대동	[NULL]					
32	32	서울특별시 용산구 한남동 271-4	한남동	[NULL]	업그레이드 카페				
33	33	서울특별시 도봉구 방촌1동 423-92	막성역대	[NULL]	분류				
34	34	대구광역시 북구 노원동3가 227-7	노원동3가	[NULL]					
35	35	대구광역시 북구 노원동3가 384-1	노원동3가	[NULL]					

파싱(parsing)* 이전의
.json 형태의 미가공 데이터

데이터 프레임* 형태로
가공된 데이터

AWS DB로 데이터 전송

* 파싱(parsing): 구문 분석. 데이터를 원하는 형태로 조립하고 다시 빼내는 것

* 데이터 프레임(data frame): 행과 열로 구성된 데이터 형태

CHAPTER. 3

추천 시스템

- 협업 필터링
- 콘텐츠 기반 필터링

03



협업 필터링(Collaborative Filtering)

: 대규모의 기존 사용자로부터 모은 데이터(평점, 구매 패턴 등)를 기반으로 사용자와 비슷한 성향의 항목을 추천하는 기술

Memory-based Methods

사용자 기반 협업 필터링(User-based Collaborative filtering)

사용자 간의 선호도를 분석하여 나와 유사한 성향의 사용자가 좋아한 상품/콘텐츠를 추천하는 기법

ex) 이 책을 구입하신 분들이 함께 산 책', '이 드라마를 시청하신 분들이 많이 시청한 또 다른 드라마' 등

유사한 선호도(구매)을 가진 사용자들의 상품/컨텐츠를 추천

아이템 기반 협업 필터링(Item-based Collaborative Filtering)

사용자들의 선호도를 바탕으로 아이템 간의 유사도를 계산하고, 특정 사용자가 어떤 아이템을 구매

하거나 좋다고 평가하면 그와 유사한 아이템을 추천해주는 방식 -> **대상이 사람이 아닌 아이템**

Model-based Methods

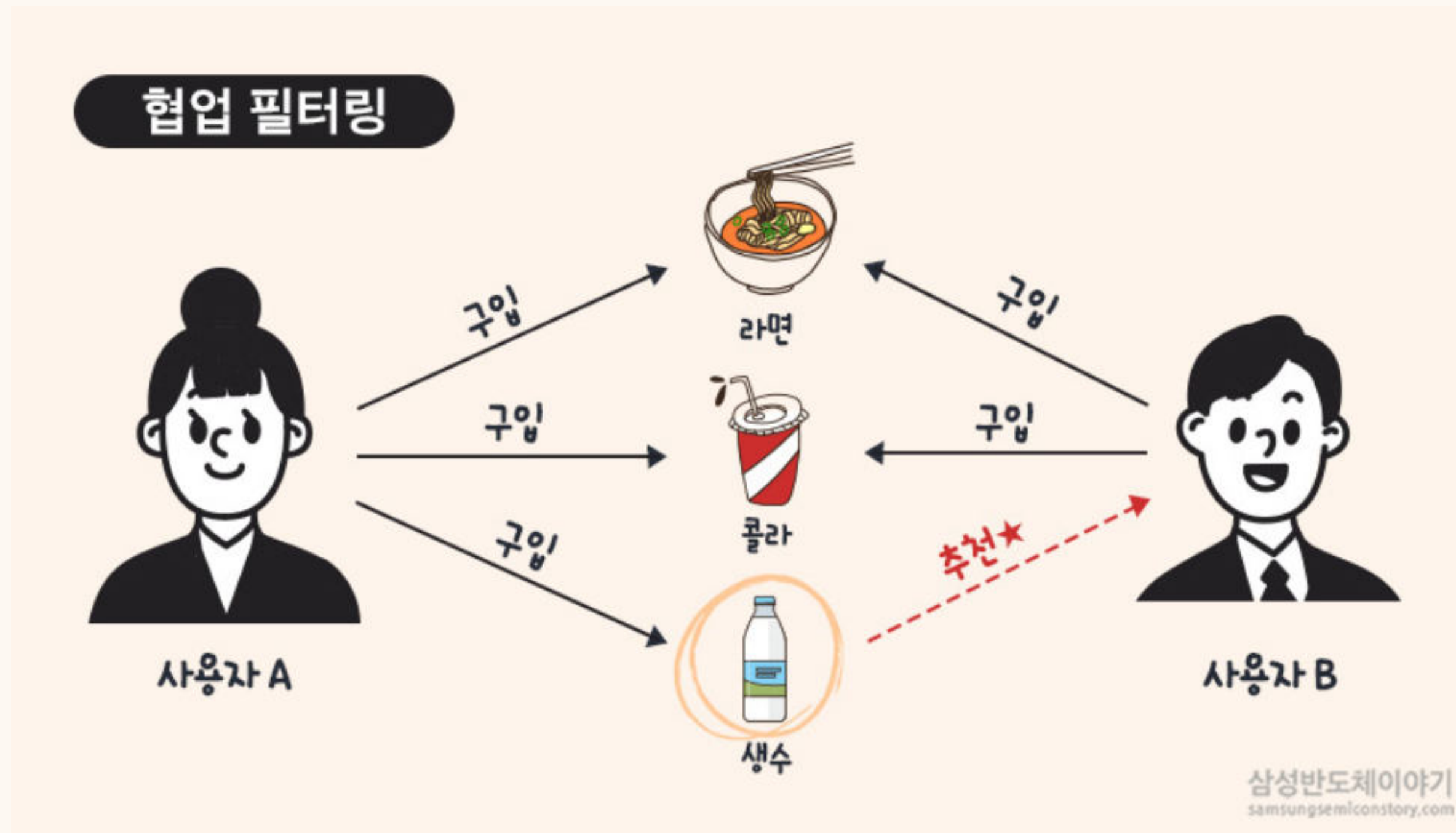
머신러닝을 이용해 평점을 예측할 수 있는 모델을 만드는 방식

과거의 사용자 평점 데이터를 이용해 모델을 만들었기 때문에

평점 정보가 없다하더라도 특정 아이템에 대한 사용자의 평점을 예측할 수 있음

협업 필터링(Collaborative Filtering)

: 대규모의 기존 사용자로부터 모은 데이터(평점, 구매 패턴 등)를 기반으로 사용자와 비슷한 성향의 항목을 추천하는 기술



협업 필터링

(Collaborative Filtering)

장점

1. 높은 적중률

직관적인 알고리즘 결과

2. 구체적 특징 불필요

항목 자체를 파악할 필요가 없음

단점

1. Cold Start

기존 데이터에 완전히 포함되지 않는 새로운 유형의 데이터가 입력될 경우 추천 불가

2. 계산 효율 저하

계산량이 비교적 많은 알고리즘이므로 사용자 수가 많은 경우 계산이 길어짐

3. Long Tail

소수의 인기 콘텐츠가 전체 추천 콘텐츠의 많은 비율을 차지 -> 데이터 비대칭 쏠림 현상

컨텐츠 기반 필터링(Contents-based Filtering)

: 항목 자체를 분석하여 사용자에게 추천 -> **아이템 분석 알고리즘이 필수**

ex) 음악을 추천하기 위해서는 음악 자체를 분석하고, 상품이라면 상품 설명을 분석해 유사한 항목을 추천



컨텐츠 분석 - 비정형 데이터로부터 관련 정보 수집

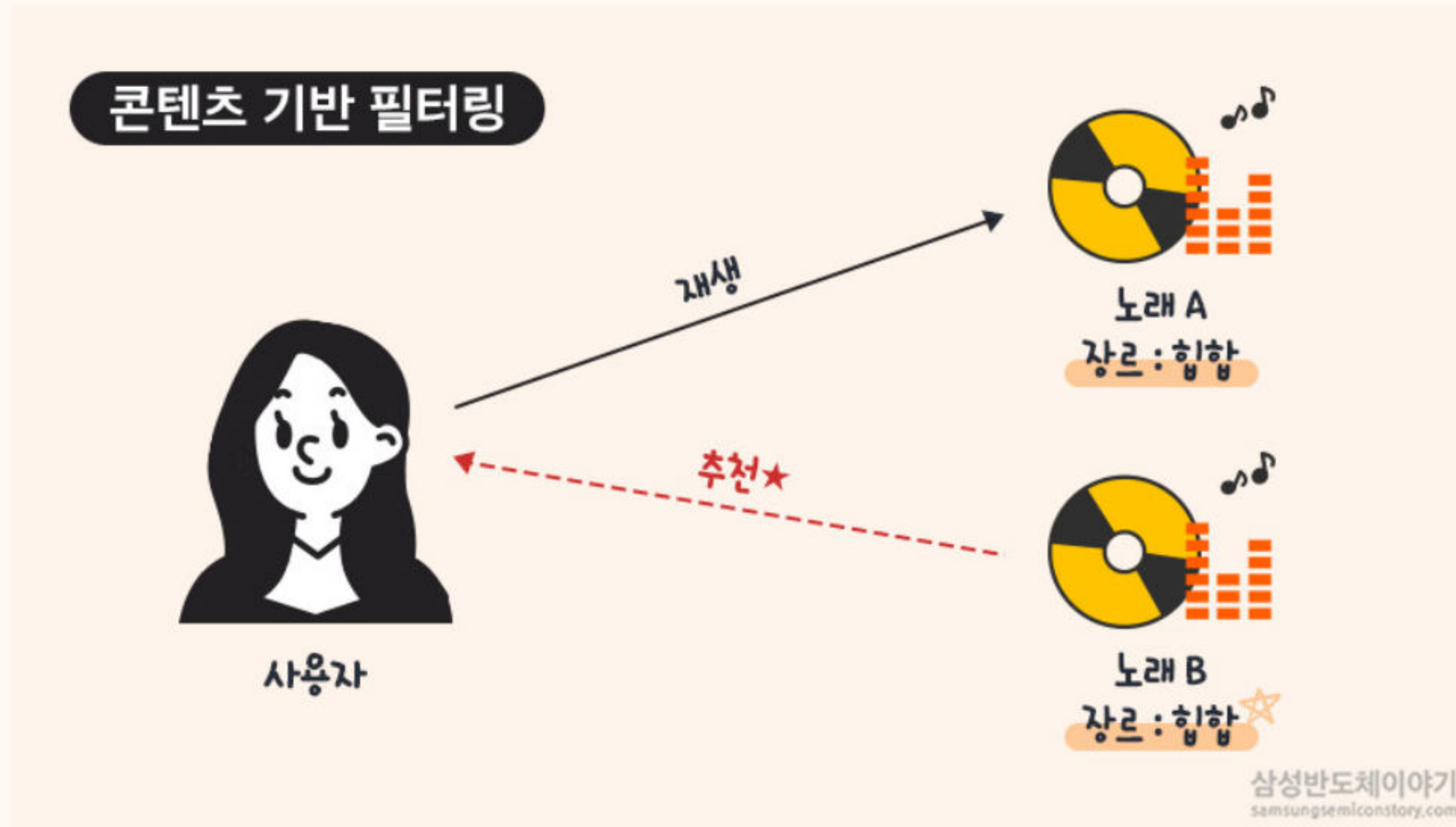
유저 프로필 파악 - 유저 성향과 아이템 취향 파악

유사 아이템 선택 - 아이템 중 분석한 유저의 프로필과 매칭하는 아이템 선택

콘텐츠 기반 필터링(Contents-based Filtering)

: 항목 자체를 분석하여 사용자에게 추천 -> **아이템 분석 알고리즘이 필수**

ex) 음악을 추천하기 위해서는 음악 자체를 분석하고, 상품이라면 상품 설명을 분석해 유사한 항목을 추천



컨텐츠 기반 필터링

(Contents-based Filtering)

장점

1. 타 사용자 데이터 불필요

아이템에 대한 정보 자체만으로 추천하여
다른 사용자에게 대한 정보가 불필요함

2. 다양한 아이템 추천

새로운 아이템이나 비인기 아이템도
추천이 가능함

3. 명확한 추천의 근거

아이템 특징 분석에 따른 추천으로 명시적인
추천 이유 존재

단점

1. 특징 설정의 어려움

아이템 특징을 선정할 때 어려움이 존재함

2. 메타 정보 한정성

아이템 정보를 함축하는 데 한계 존재

04

CHAPTER. 4

향후 계획

- 파트별 작업 계획





추천 알고리즘 생성

크롤링을 활용한 필요 데이터 수집

유저, 가게, 리뷰를 기반으로 한 추천 시스템 구축

효과적인 UI/UX 적용

데이터에 적합한 화면 구성

기존 맛집 추천 서비스와의 차별화 된 시각화



감사합니다



Q&A