



사용자의 기억을 담아
새로운 여행으로 이끌어 주는
AI 앱 “기억의 지도”

Team. 기억의 지도

김승순

April 15, 2024





*"Provide a new travel Experiences through
Nostalgia of Travel"*

Table of Contents

- 01. 팀 소개 (R&R)
- 02. 프로젝트 및 과정 소개
- 03. 데이터 소개
- 04. AI 모델 소개
- 05. Application Demo 시연
- 06. Lessons & Learns

기억의 지도 팀 소개

01. 팀 소개 (R&R)

Team Leader

- 앱 서비스 기획
- 앱 UI/UX 디자인
- 관광지 이미지 검색 및 추천 모델 개발
- Flutter를 활용한 앱 개발
- Firebase 기반 서버 구축



02

프로젝트 소개

*"Travel makes one modest. You see what
a tiny place you occupy in the world." –
Oscar Wilde*



02. 프로젝트 소개

"The world is a book, and those who do not travel read only one page." - Saint Augustine

"I have always found that travel expands my mind, challenges my belief, and forces me to re-evaluate my life." - Anthony Bourdain

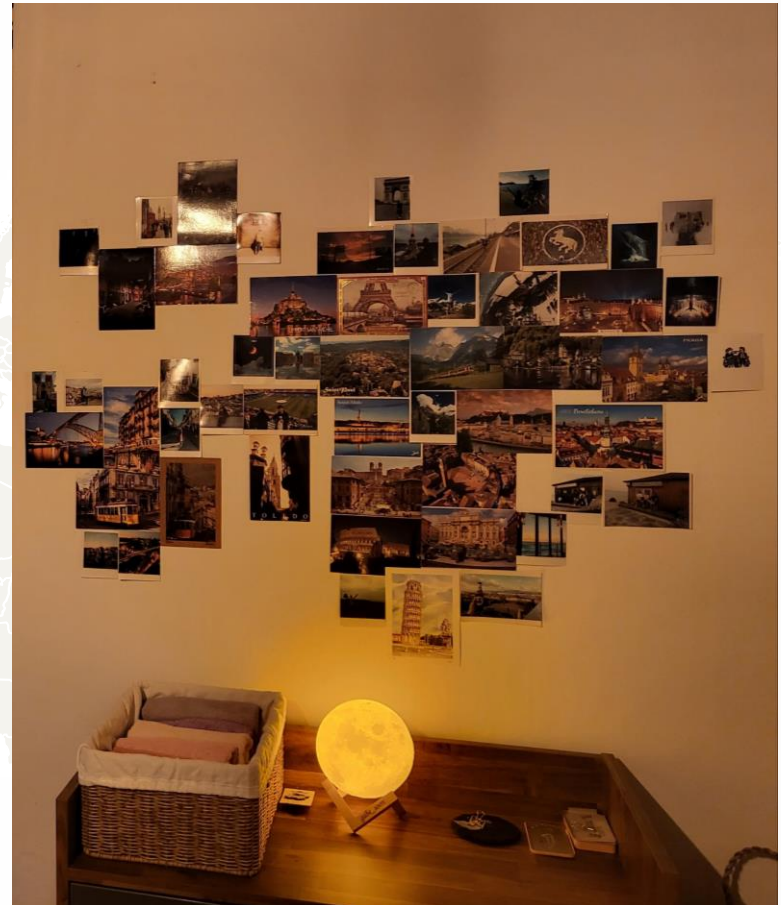
"It is not the distance you travel but the memories you create that matter" - Jane Austen

"Travel has a certain magic that sets free the imagination and sets the spirit soaring" - Francis Bacon

"Travel makes one modest. You see what a tiny place you occupy in the world." - Oscar Wilde

"Travel brings identity and purpose to life." - Rumi

"The farther we travel the more we see, the more we see the more we know, and the more we know the more we realize how little know." - Mark Twain



02. 프로젝트 소개

시장동향

Customized Travel

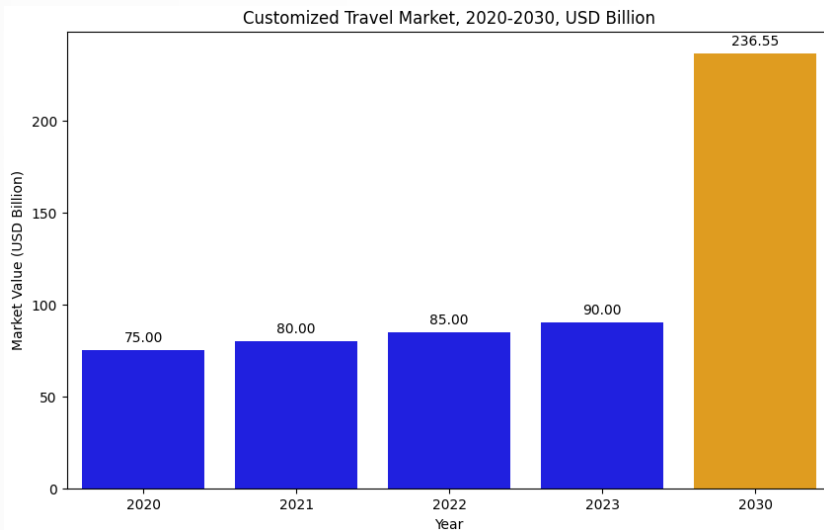
- 맞춤형 여행에 대한 선호도는 2030년까지 약 236.55 USD Billion 까지 성장할 것으로 전망

R.O.U.T.E

- R : 쉼이 있는 여행
- O : 원포인트 여행
- U : 나만의 명소 여행
- T : 스마트 기술 기반 여행
- E : 모두에게 열린 여행

Opinion Leader

- 맞춤형 투어의 증가
- 소셜 살롱과 커뮤니티 기반 여행
- 고급화와 특수 목적 여행
- AI와 직거래를 통한 자동화
- 여행 멤버십과 구독 서비스
- 콘텐츠 중심의 비즈니스 모델



Resource : 한국관광협회중앙회, 2023

2024 국내 관광 트렌드 R.O.U.T.E.



회, 여가, 소비, 환경, 노동분야의 주요 이슈를 토대로 2024년 유망 여행 테마 5가지 선정함

경쟁업체 서비스 특징

	기억의 지도	트리플	여기어때	마이리얼트립
서비스 컨셉	<ul style="list-style-type: none"> 감성적인 여행 추억을 시각화 	<ul style="list-style-type: none"> 자유 여행 스케줄링 지원 	<ul style="list-style-type: none"> 숙박 시설 및 상품 서비스 빅데이터 기반의 여행정보 제공 	<ul style="list-style-type: none"> 현지 가이드 및 투어 매칭 여행 상품 종합결제 기능
개인화 여행 추천	<ul style="list-style-type: none"> 사용자 여행 기록 및 피드백 기반의 개인화된 맞춤형 여행지 추천 	<ul style="list-style-type: none"> 빅데이터 기반 보편적 트렌드 정보 제공 	<ul style="list-style-type: none"> 빅데이터 기반 보편적 트렌드 정보 제공 	<ul style="list-style-type: none"> 빅데이터 기반 보편적 트렌드 정보 제공
추억 관리기능	<ul style="list-style-type: none"> 사진과 일기로 여행 기록 관리 	<ul style="list-style-type: none"> 고객 후기와 여행기 보관 기능 	<ul style="list-style-type: none"> 숙박 이용 후기 관리 	<ul style="list-style-type: none"> 가이드와 여행자의 소통 기능 가이드 후기 관리
AI 기술 활용	<ul style="list-style-type: none"> CLIP모델 활용한 이미지 분석 NLP활용한 여행기록 분석 LightGCN 기반 여행지 추천 	<ul style="list-style-type: none"> 빅데이터 분석 기반 여행지 & 여행상품 추천 	<ul style="list-style-type: none"> 빅데이터 분석 기반 여행지 & 여행상품 추천 	<ul style="list-style-type: none"> 빅데이터 분석 기반 가이드 & 여행지 추천
시장 트렌드 반영	<ul style="list-style-type: none"> 개인화된 여행지 추천 정보 제공 	<ul style="list-style-type: none"> 편리한 예약 및 관리 기능 제공 	<ul style="list-style-type: none"> 할인과 특가상품을 통한 가성비 여행 경험 제공 	<ul style="list-style-type: none"> 차별화된 가이드 투어를 통한 색다른 여행 경험 제공

고객 페르소나 설정

김현우

남성 34세, 마케팅 매니저

"맞춤 여행 추천을 받을 수 있다면 효율적으로 여행 계획을 세울 수 있을 것 같아요"



관심사

- 감성적인 여행지를 방문해 사진을 남긴다
- 여가 시간에는 새로운 여행지를 모색한다
- 개인화된 경험을 중시한다.

시나리오

회사의 다양한 이벤트를 기획하는 현우씨. 일에 치여 자주 여행을 가지는 못하지만, 한번의 여행이 소중한만큼 나에게 맞는 나만의 여행을 통해 추억을 쌓고 싶어하며, 이러한 경험을 소중히 여긴다.

이지은

여성 29세, 여행 블로거

"여행을 다니며 찍은 사진들을 기반으로 새로운 곳을 찾아볼 수 있으면 좋겠어요."



관심사

- 여행지마다 특별한 사진을 남긴다.
- 자신의 블로그에 여행후기를 작성한다.
- 새로운 여행지 발견에 관심이 많다.

시나리오

주말마다 새로운 장소를 방문해 사진을 찍는 지은씨. 자신의 블로그를 관리하며 독특한 여행기록을 남기고 있다.

장민아

여성 26세, 대학원생

"여행 친구들과 공유할 수 있는 여행 사진 기반 추천 앱이 있으면 좋겠어요."



관심사

- 건축물을 찾아다니며 사진을 찍는다.
- 답사 경험과 기록을 중요하게 생각한다.
- 여행 경험을 친구들과 공유한다.

시나리오

전공과 관련된 건축물을 찾아 여행하는 민아씨. 대학원 연구로 바빠 본인의 여행 기록을 관리하는데 어려움을 겪고 있다.

박준석

남성 38세, 시스템 엔지니어

"여행 계획을 세우고 사진으로 기록을 남길 수 있는 것이 여행의 큰 즐거움이죠"



관심사

- 가족 여행 시 많은 사진을 남긴다.
- 휴가 때마다 새로운 여행지를 찾아다닌다.
- 아이들에게 여행지에 대한 정보를 알려주며 교육을 하는 것을 선호 한다.

시나리오

가족과 함께 여행을 떠나는 것을 좋아하는 준석씨. 여행 사진과 기록을 잘 보관하고 싶고, 아이들에게 여행지에 대한 다양한 정보를 알려주며 교육이 되는 여행을 시켜주고 싶다.

02. 프로젝트 소개

앱 주요 기능 소개



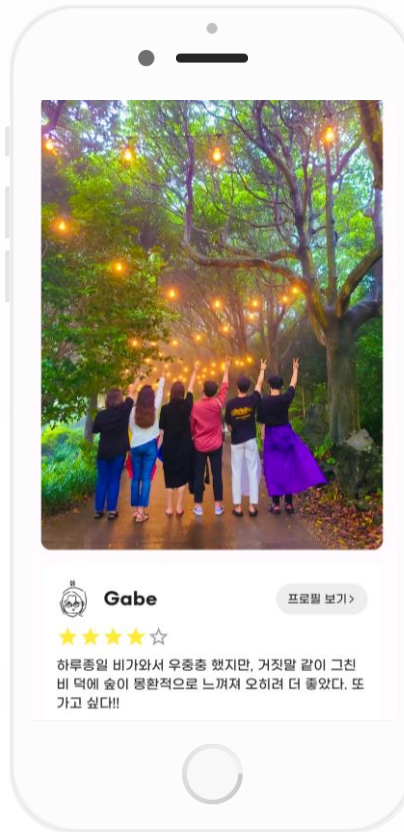
사진으로 여행지 찾기

사용자가 촬영한 이미지를 유사이미지 검색기능을 통해 특정하여 여행지를 찾아줍니다.



여행지 상세정보 제공

사용자가 등록한 여행지에 대한 다양한 정보를 제공합니다.



"기억의 지도는 단순한 여행 정보 제공을 넘어, 사용자가 직접 만드는 여행의 기록장이자 미래 여행의 나침반 역할을 합니다."

여행지에서의 소중한 순간들을 AI가 기록하고 그 데이터를 바탕으로 '나도 몰랐던 내 취향의 여행지'를 발견하게 해주는 여행 도우미 입니다."



여행 기록

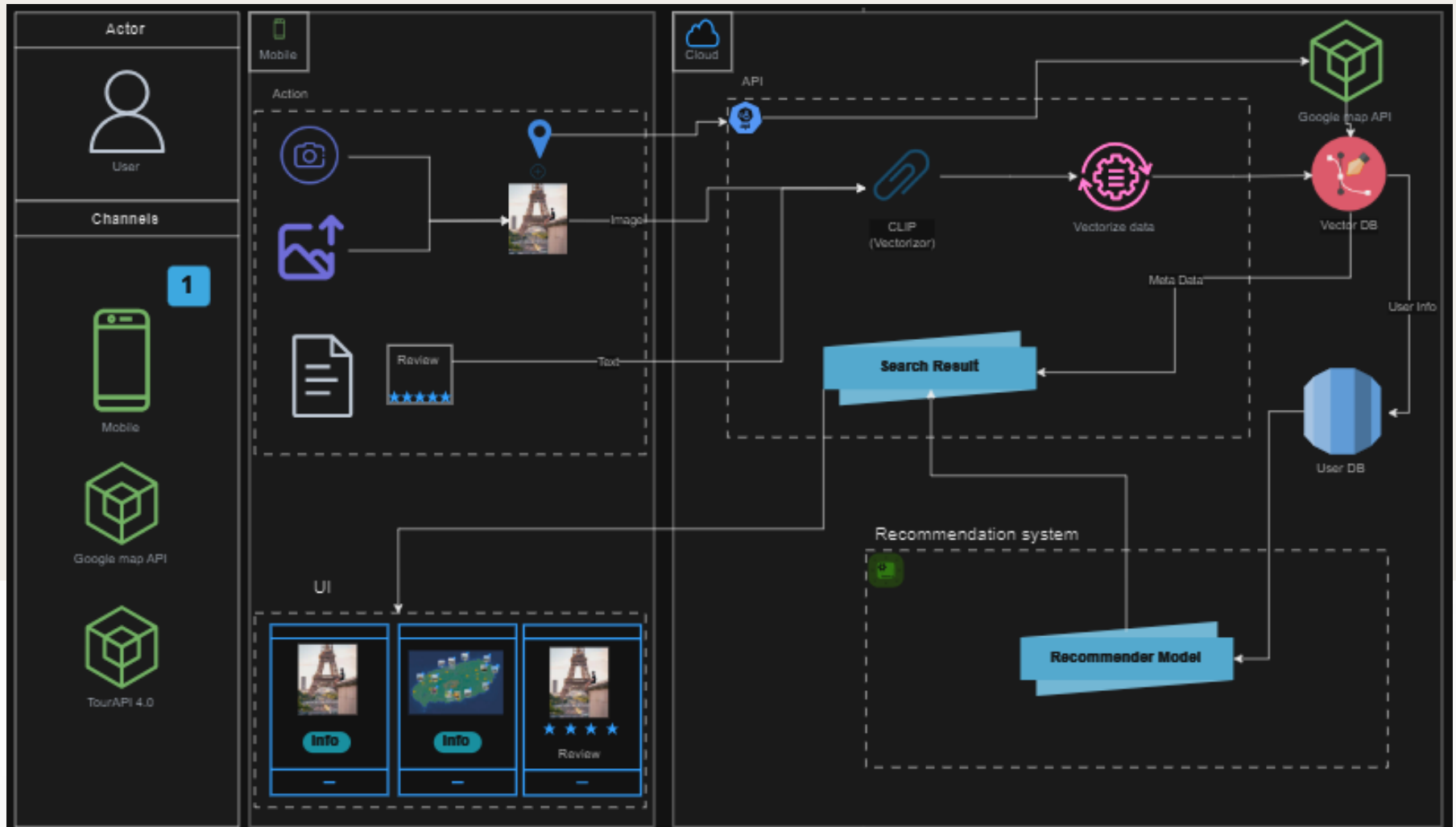
사용자의 여행을 앨범으로 기록하고 추억할 수 있도록 돕습니다.



여행지 추천

사용자의 여행기록을 기반으로 개인화된 "나도 모르는 나만의 맞춤형 여행지 추천"을 제공

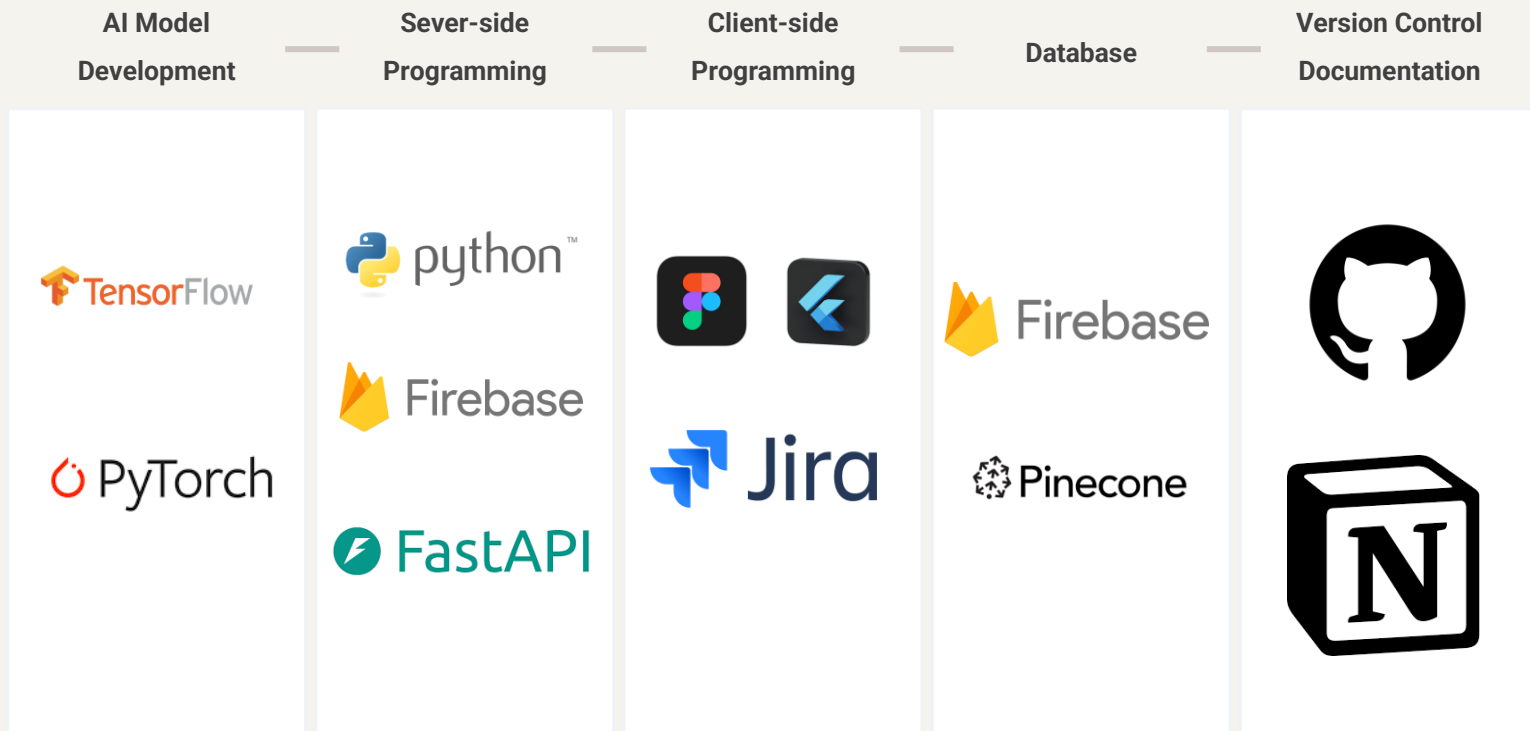
기억의 지도 시스템 구조도



주차별 작업 내용

구분	기간	활동	비고
사전 기획	• 2/22(목) ~ 2/29(목)	<ul style="list-style-type: none"> 프로젝트 기획 및 주제 선정 기획안 작성 	• 아이디어 선정
데이터 수집	• 3/4(월) ~ 3/8(금)	<ul style="list-style-type: none"> 필요 데이터 및 수집 절차 정의 외부 데이터 수집 	
Vector DB 및 좌표 처리 기능 개발	• 3/11(월) ~ 3/15(금)	<ul style="list-style-type: none"> 제주 및 도서지역 데이터로 초기벡터DB구축 Google Map API로 주소 추출 기능 개발 	
유사이미지 검색 및 개인 추천 시스템 개발	• 3/18(월) ~ 3/22(금)	<ul style="list-style-type: none"> Maskrcnn 기반 Segmentation 및 CLIP기반 이미지 데이터 벡터화, 유사 이미지 검색 기능 구현 유사이미지 시나리오별 성능 실험 및 최종 안 결정 더미 데이터 활용한 추천 모델 연구 시작 	
시스템 통합 및 모바일 앱 개발	• 3/25(월) ~ 3/29(금)	<ul style="list-style-type: none"> 모바일 앱 기초 화면 구성 FastAPI 엔드포인트 정의 및 테스트 데모앱과 서버 연동 테스트 	
모바일 앱 개발 및 추천 모델 개선	• 4/1(월) ~ 4/12(금)	<ul style="list-style-type: none"> Flutter 앱 개발 및 기능 구현 추천알고리즘별 기능 구현 및 성능 비교, 최종 모델 선정 최종 앱 및 시스템 구축 및 테스트 	
총 개발기간	• 3/4(월) ~ 4/12(금)(총 6주)		-

Technology Stack



02. 프로젝트 소개

작업방식



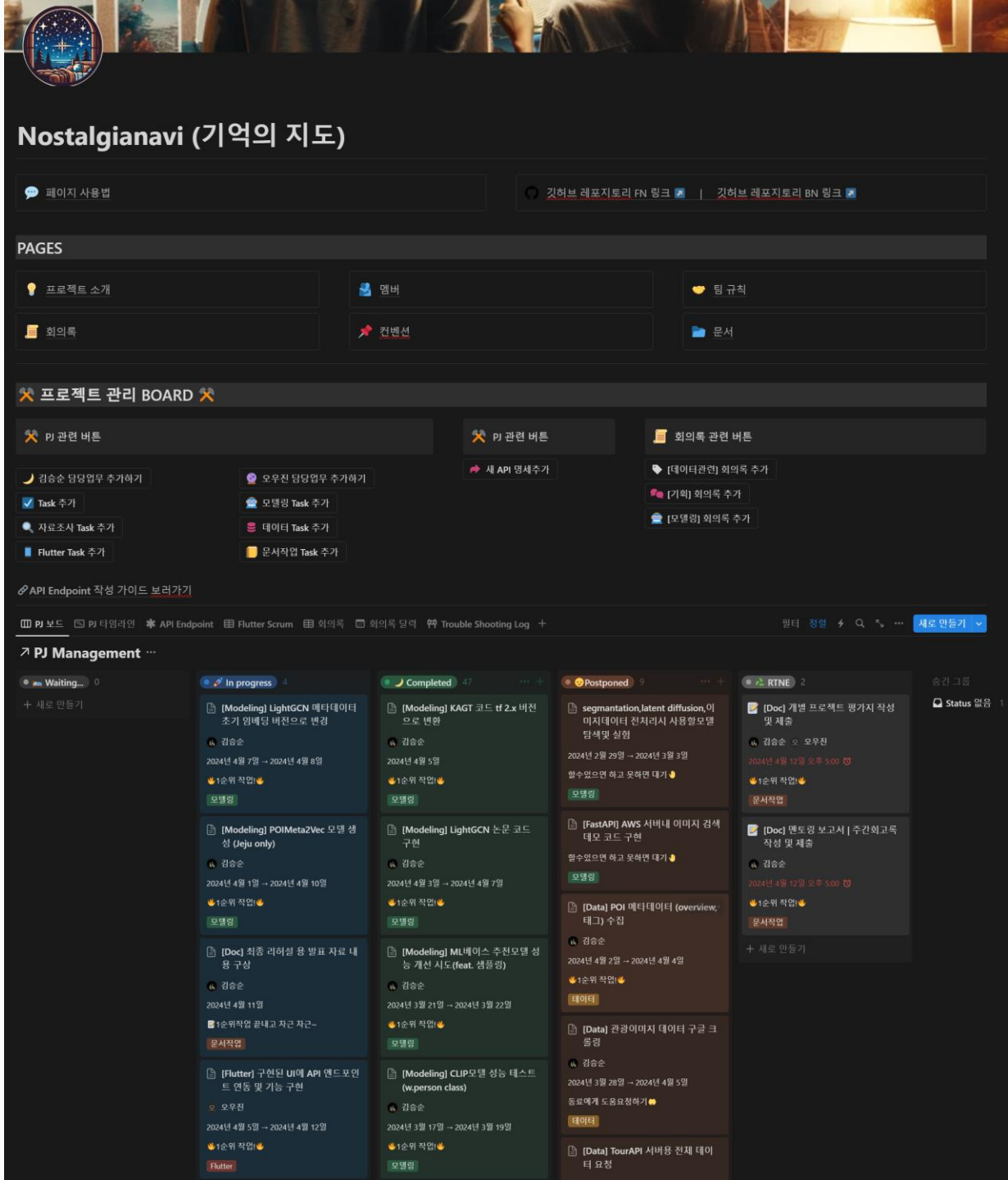
Agile 방식 응용

- Notion을 활용한 프로젝트 스프린트별 작업 상황 정리
- Task 추가시 작업 목적 및 작업 계획을 작성해 작업 외 인원이 확인해도 내용 파악이 가능하도록 함
- 1일 1회 작업 내용 정리 및 피드백
- 마감기한, 우선순위 설정해 작업 집중도 향상 도



자료조사 내용 카테고리별 정리

- 프로젝트에 필요한 자료들을 분야별로 구분해서 정리하고 상시 공유
- 공유된 자료의 출처 및 상세 내용을 기록



02. 프로젝트 소개

작업방식



Agile 방식 응용

- Notion을 활용한 프로젝트 스프린트별 작업 상황 공유
- Task 추가시 작업 목적 및 작업 계획을 작성해 작업 외 인원이 확인해도 내용 파악이 가능하도록 함
- 1일 1회 작업 내용 공유 및 피드백
- 마감기한, 우선순위 설정해 작업 집중도 향상 도모



자료조사 내용 카테고리별 정리

- 프로젝트에 필요한 자료들을 분야별로 구분해서 정리하고 상시 공유
- 공유된 자료의 출처 및 상세 내용을 기록

[Data] 로그 데이터 & 관광데이터 병합처리

Status	Completed
Representative(담당자)	김승순
Task_Period	2024년 3월 26일 ~ 2024년 4월 1일
분류	데이터
긴급도	Yes!!!!
중요도	Yes!!!!
작업우선순위	1순위 작업!
속성 4개 추가	

댓글 추가

Task 내용

I 현재 해야 하는 업무 내용 및 목적 작성

작업 목표:

- 한국 관광공사에서 제공하는 관광지 메타데이터인 travel_df와 외부 기업이 수집한 여행객 로그 데이터인 travel_log를 병합하여 관광지 추천 시스템을 만드는 것

문제 상황:

- 두 데이터프레임을 병합하기 위해서는 공통된 키(key)가 필요한데, travel_df의 contentid와 travel_log의 POI_ID가 이에 해당.
 - 그러나 travel_log의 POI_ID 컬럼에는 null 값이 존재
 - travel_log의 장소명(VISIT_AREA_NM, POI_NM 등)이나 주소(ROAD_NM_ADDR, LOTNO_ADDR 등)가 travel_df의 해당 정보(title, addr1 등)와 일치하지 않는 경우가 있음
 - travel_log에서 동일한 장소를 나타내는 데이터라도 VISIT_AREA_NM이 서로 다르게 입력된 경우가 있음
 - 좌표 정보(X_COORD, Y_COORD)는 모바일 앱에서 수집된 사용자 위치 좌표로, travel_df의 관광지 좌표와 정확히 일치하지 않음.
 - 또한 변경 내에 여러 관광지가 존재할 수 있어 활용이 어려움

해결 방안:

- travel_log의 POI_ID가 null인 경우, 장소명이나 주소 정보를 이용하여 travel_df의 contentid와 매칭할 수 있을 것이란 가정
- 주소 정보를 기반으로 장소를 매칭하되, 한국의 주소 체계가 도로명 주소(ROAD_NM_ADDR)와 지번 주소(LOTNO_ADDR) 두 가지임을 고려
 - travel_df의 addr1이 도로명 주소 체계를 따르므로, travel_log의 ROAD_NM_ADDR과 비교하는 것이 적절함
- 주소 문자열에 대해 전처리(불필요한 공백, 특수문자 제거 등)를 수행한 후, 유사도 측정 알고리즘(Levenshtein distance, Jaro-Winkler distance 등)을 적용하여 유사도 점수를 계산
- 유사도 점수가 일정 임계값 이상인 경우, 해당 장소들을 동일한 장소로 간주하고 travel_log의 POI_ID에 travel_df의 contentid 값을 할당
- 매칭되지 않은 장소들에 대해서는 추가적인 전처리나 규칙 기반의 매칭 로직, 사용자 확인 요청 등의 방법을 사용할 수 있음

Task 계획

I Task 수행을 위한 계획을 간단하게 작성합니다.

- ☒ 전처리 코드 작성
- ☒ 샘플 테스트를 통해 POI 매핑
- ☒ 결과 확인
 - 변경된 POI의 코드로 조회된 장소가 실제 방문지 이름과 동일인지 등
- ☒ 전체 데이터에 대한 변경 처리
- ☒ 결과 확인

02. 프로젝트 소개

추천 시스템 구현과정

"Travel makes one modest. You see what
a tiny place you occupy in the world."
- Oscar Wilde

06 운영 & 유지

1. 피드백을 통한 모델 및 앱서
비스 업데이트 적용

05 System 화

1. 모델 이미지 생성
2. 페르소나 / 타겟별 배치 형성
3. 서버 확장 비용 산정
4. 실시간 서버 세팅

04 A/B Test

1. 내부 테스트
2. 외부 테스트
3. 현업 KPI 컨센서싱
4. 사용자 피드백 수집
5. A/B테스트 & 모니터링



01 추천 기획

1. 가설 설정
2. 고객 페르소나 형성
3. 추천 타겟 설정
4. 현업 KPI 컨센서싱

02 Data Mart

1. 가설별 필요 데이터 맵핑
2. 데이터 수집
3. 데이터 전처리

03 Model R&D

1. 모델 리서치
2. 모델링
3. 모델 이슈체크
4. 사용자 피드백 수집
5. 추가 모델링

02. 프로젝트 소개

추천 시스템 구현과정

"Travel makes one modest. You see what
a tiny place you occupy in the world."
- Oscar Wilde

04 Prototype 앱 개발

1. Figma 기반 앱 UI기획
2. API Endpoint 설계 및 테스트
3. Flutter 기반 앱 구현 및 API
연결

03 Model R&D

1. 모델 리서치
2. 모델링
3. 성능 검증 및 피드백
4. 모델 추가 연구

01 추천 기획

1. 앱 컨셉 설정
2. 서비스 타겟 사용자 및 고객
페르소나 정의

02 Data Mart

1. 데이터 확보
2. 기획내용과 확보된 데이터
맵핑(DB Schema)
3. 데이터 분석
4. 데이터 전처리



03

데이터 소개

*"The farther we travel the more we see, the
more we see the more we know, and the
more we know the more we realize how
little know." – Mark Twain*



03. 데이터 소개



관광 메타 데이터

1. AI Hub의 'KVQA 데이터셋'

- 지식정보(text)-시각정보(image)-질의응답(text)으로 이뤄진 다중쌍 데이터셋
- POI* 데이터 : 390,000 개
- 이미지 데이터 : 2,332,500 장

2. TourAPI 데이터셋

- 한국 관공공사 데이터셋
- 관광지 카테고리별 상세 정보 제공(POI포함 총 327종)
- **POI 데이터 52,338 개**

사용대상 : 벡터DB구축, 추천알고리즘



관광 로그 데이터

1. AI Hub의 '국내 여행로그 데이터'

- **여행객 16,000명의 여행 기록** 데이터
- 상세 데이터 총 138종
- 여행사진 726,552장
- POI 7,442,866 곳 데이터

사용대상 : 추천알고리즘



크롤링 데이터

1. POI 이미지

- TourAPI의 여행지명과 지역명을 결합해 구글 이미지에서 검색

2. POI Overview & Tag

- TourAPI의 여행지명과 지역명을 결합해 구글지도, 카카오맵, 네이버 백과사전에서 수집

사용대상 : 벡터DB구축, 추천알고리즘

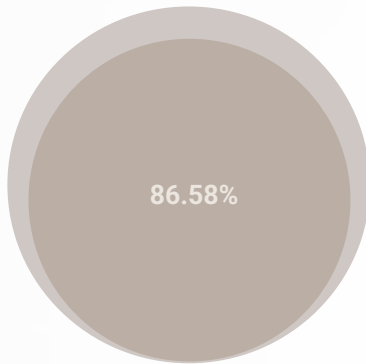
※ 약어 설명

KVQA : Knowledge-Aware Visual Question Answering

POI : Point of Interest

데이터 문제점 및 해결 방안

"The best and most beautiful things in the world
cannot be seen or even touched. They must be felt
with the heart."



Garbage data

수집 데이터의 대부분이 관광지 카테고리 분류 되어 있음에도 관광지라 보기
어려운 데이터가 많음
Ex) 교회, 동네식당, 아파트 시설(초록마
을현전에버빌아파트공원), 공공기관(ex.
한국핵융합에너지 연구원) 등



신규 메타데이터 수집

- TourAPI에서 제공하는 메타데이터를 기존 데이터로 변경
- 단, 1일 1,000건 호출 건수 제한으로 일부 데이터 확보 불가능



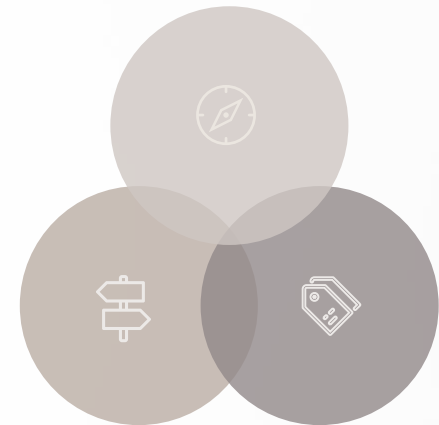
추가 데이터 수집

- 이미지 데이터에 대한 추가 수집 필요
- 메타데이터 일부(Overview)와 관광지 키워드(tag)데이터 수집



데이터 불일치

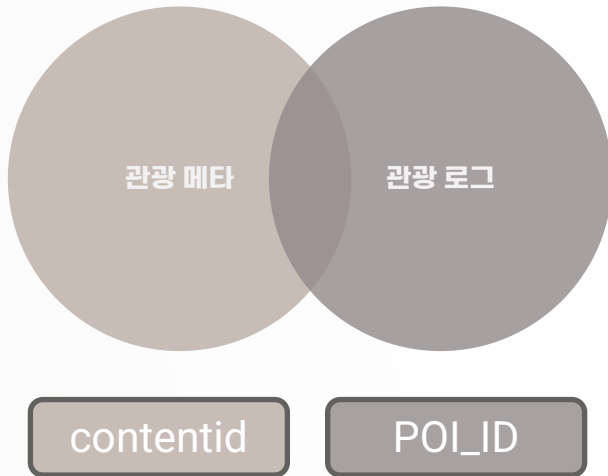
- 관광메타데이터와 관광로그 데이터간 불균형 문제로 균일화 및 데이터 맵핑 작업 수행
- NLP 기법들을 응용해 데이터 전처리 수행



관광 데이터 병합 (feat. NLP Preprocessing)



Key 데이터 불일치 문제



- 관광 메타데이터의 contentid와 관광 로그 데이터의 POI_ID는 같은 의미를 지닌 데이터
- 하지만 서로 값의 형태나 구성 내용이 서로 전혀 달라 한쪽의 체계로 통합해야 함
- 또 다른 공통 Column인 '주소'의 경우 '관광로그' 데이터의 주소가 POI의 공식 주소와 불일치
- 관광로그에서 POI_ID가 null인 데이터가 전체 데이터의 30%로 누락되는 데이터가 너무 많음
- 두 데이터의 '도로명 주소' 컬럼의 주소 텍스트를 **FuzzyWuzzy** 라는 **문자열 비교 라이브러리**를 사용해 유사도 측정(Threshold 80%)해 같은 데이터 후보군으로 간주하되 가장 높은 값을 가진 데이터로 병합
- 가장 높은 유사도를 가진 데이터의 경우 장소명을 FuzzyWuzzy로 유사도를 측정(Threshold 90%)해 2차 검증해서 최종적으로 동일 데이터로 선정
- 수행 결과 전처리전 병합시 **전체 5.2만개의** 데이터가 **1만개로 감소**하던 데이터가 **4.7만개로 데이터를 보존**

※ FuzzyWuzzy 사용 예

- 오타가 있는 문자열 비교 : 사용자의 입력오류, 데이터 불일치 교정
- **데이터 정제** : 서로 다른 데이터 소스에서 비슷한 이름이나 용어 일치시킬때 활용
- 자동 완성 기능 : 사용자가 입력하는 동안 가장 비슷한 단어, 문구 제안할 때 사용

관광 데이터 병합 (feat. NLP Preprocessing)

POI_ID	VISIT_AREA_TYPE_CD	VISIT_AREA_NM	ROAD_NM_ADDR	LOTNO_ADDR
NaN	13	구미에코랜드	경북 구미시 산동읍 인덕1길 195	경북 구미시 산동읍 인덕리 280
POI0100000006W0H2H	6	에코랜드테마파크	제주특별자치도 제주시 조천읍 번영로 1278-169	제주특별자치도 제주시 조천읍 교래리 385-1
NaN	1	제주 에코랜드	번영로 1278-169	NaN
NaN	6	에코랜드 테마파크	NaN	NaN

- 여행객이 방문한 여행지에 대한 기록, 한명의 여행객별로 다양한 방문지가 있으나, POI 별 ID 나 장소명, 주소등이 관광 메타 데이터와 불일치 하거나, Null 인 값이 대다수.

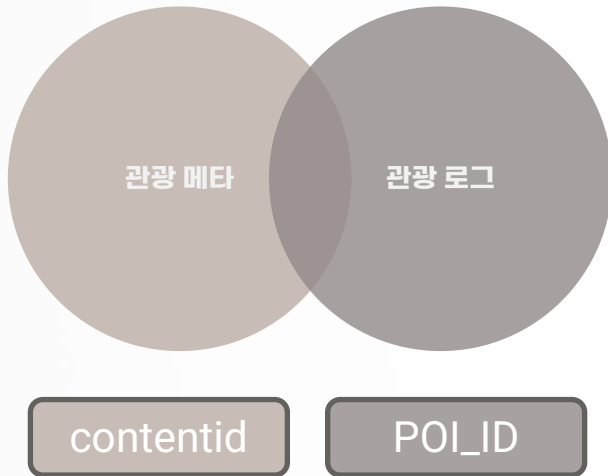


Travel Log
여행 기록 데이터

관광 데이터 병합 (feat. NLP Preprocessing)



Key 데이터 불일치 문제



- 관광 메타데이터의 contentid와 관광 로그 데이터의 POI_ID는 같은 의미를 지닌 데이터
- 하지만 서로 값의 형태나 구성 내용이 서로 전혀 달라 한쪽의 체계로 통합해야 함
- 또 다른 공통 Column인 '주소'의 경우 '관광로그' 데이터의 주소가 POI의 공식 주소와 불일치
- 관광로그에서 POI_ID가 null인 데이터가 전체 데이터의 30%로 누락되는 데이터가 너무 많음
- 두 데이터의 '도로명 주소' 컬럼의 주소 텍스트를 **FuzzyWuzzy** 라는 **문자열 비교 라이브러리**를 사용해 유사도 측정(Threshold 80%)해 같은 데이터 후보군으로 간주하되 가장 높은 값을 가진 데이터로 병합
- 가장 높은 유사도를 가진 데이터의 경우 장소명을 FuzzyWuzzy로 유사도를 측정(Threshold 90%)해 2차 검증해서 최종적으로 동일 데이터로 선정
- 수행 결과 전처리전 병합시 **전체 5.2만개의** 데이터가 **1만개로 감소**하던 데이터가 **4.7만개로 데이터를 보존**

※ FuzzyWuzzy 사용 예

- 오타가 있는 문자열 비교 : 사용자의 입력오류, 데이터 불일치 교정
- 데이터 정제 : 서로 다른 데이터 소스에서 비슷한 이름이나 용어 일치시킬때 활용
- 자동 완성 기능 : 사용자가 입력하는 동안 가장 비슷한 단어, 문구 제안할 때 사용

04

AI모델 소개

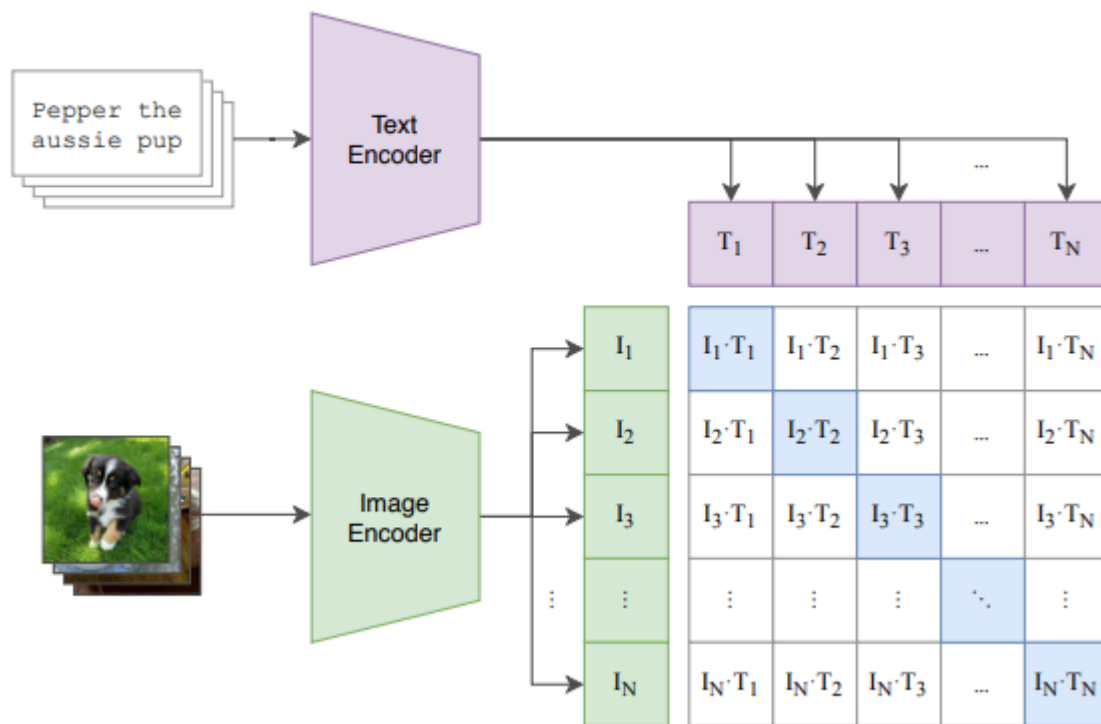
*"I have always found that travel expands
my mind, challenges my belief, and forces
me to re-evaluate my life." – Anthony
Bourdain*



04. AI모델 소개 – 이미지로 관광지 찾기(유사 이미지 검색)

- 이미지와 텍스트를 연결하는 기능이 강력한 CLIP모델을 Vectorizer로서 활용
- 단기적, 장기적 고려사항을 감안해 CLIP 모델을 선정
- 강력한 Pre-trained CV모델의 필요성
- 사진에서 텍스트를 추출해 추천모델의 데이터로 활용하기 위함
- CLIP은 Zero-shot Prediction 에서 강력한 성능을 보임

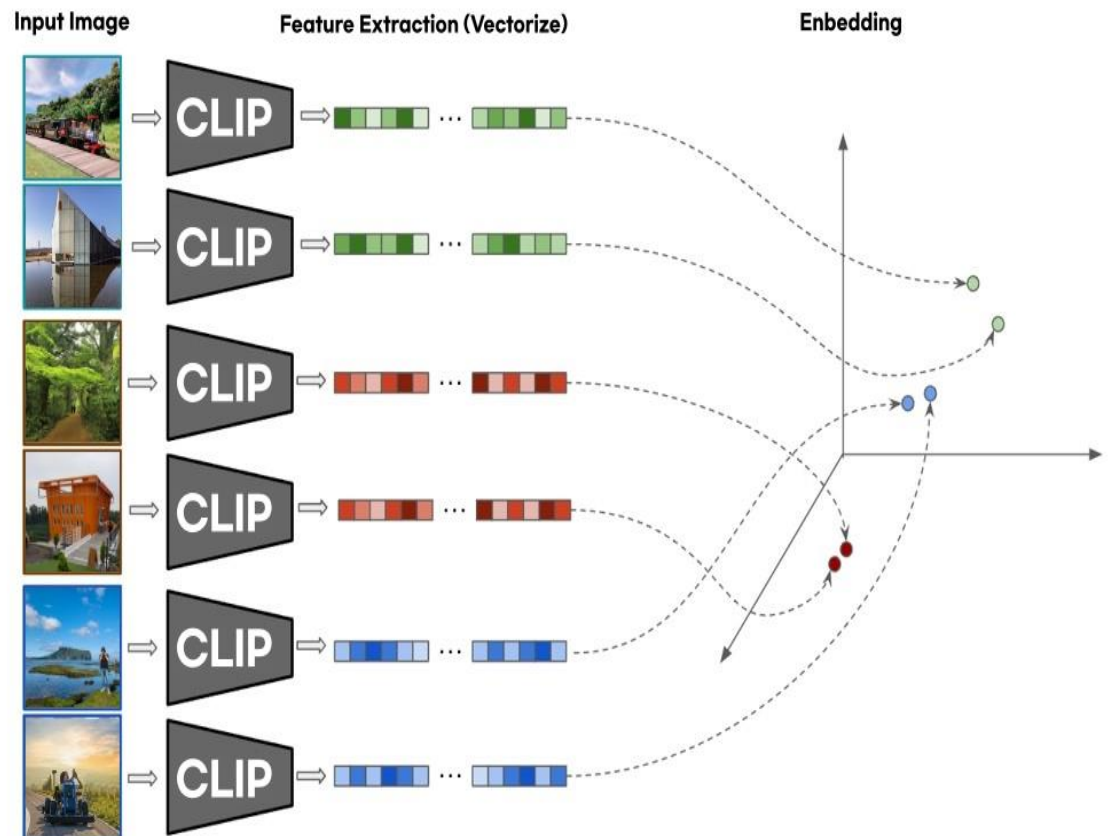
(1) Contrastive pre-training



04. AI모델 소개 – 이미지로 관광지 찾기(유사 이미지 검색)

벡터 DB 구축을 위해 관광 메타데이터와 결합하여 사용

- 관광메타데이터의 장소명으로 검색하여 크롤링한 이미지 데이터를 사용
- 이미지를 CLIP으로 벡터화
- 파일명과 동일한 이름을 가진 관광 메타 데이터를 찾아 해당 데이터의 주소, 좌표, 장소명, 원본 이미지 저장 경로를 메타데이터로 벡터DB화
- 메타데이터는 차후 유사이미지 검색 시 주소지 기반의 1차 필터링 및 관광 메타 데이터 검색을 위해 사용



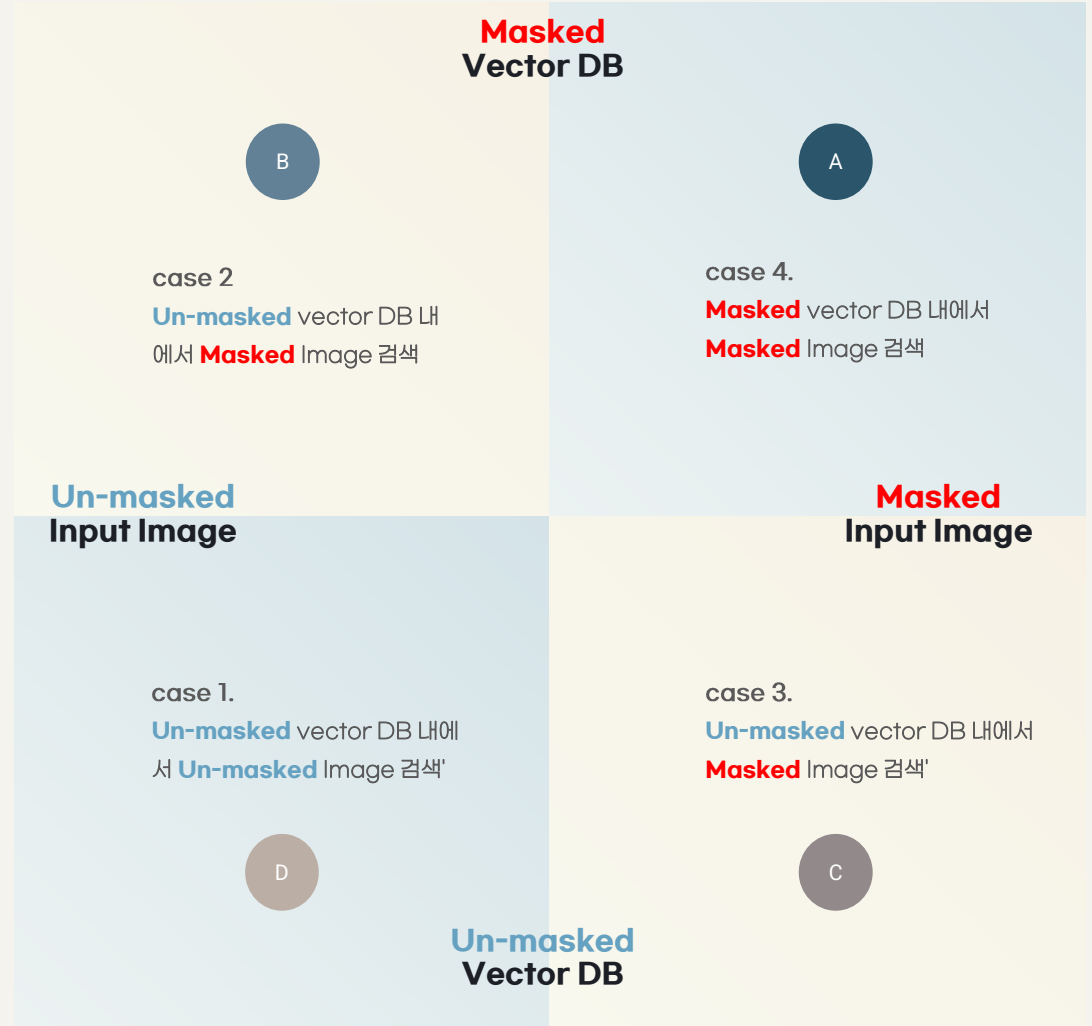
유사이미지 검색 성능 비교 실험 (1)

실험 목적

- CLIP모델과 Vector DB를 활용한 유사 이미지 검색 기능 성능 향상 테스트
- Masking 유무를 기준으로 4가지 기준으로 나눠서 실험 환경 세팅

Un-masked
Image

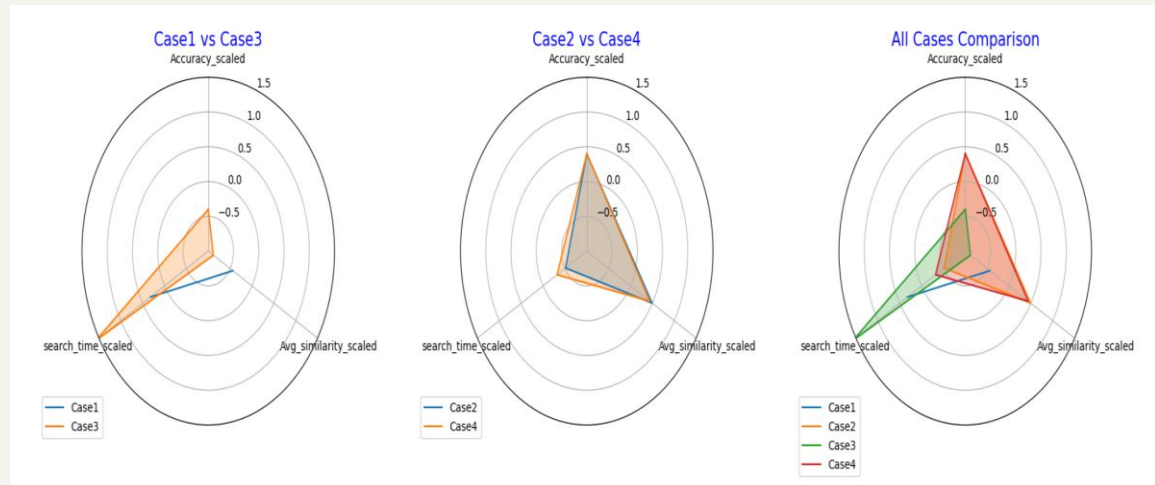
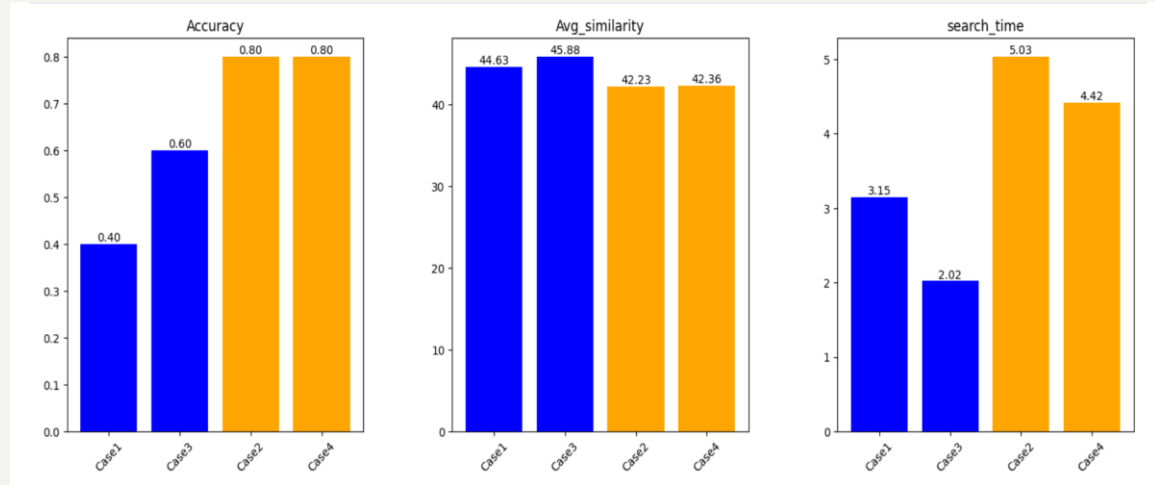
Masked Image



유사이미지 검색 성능 비교 실험 (1)

실험 결과

- **정확도 측면**에서는 Case2와 Case4가 높은 성능을 보임
- 검색 결과의 평균 유사도에는 큰 편차가 존재하지 않음
- **검색 시간**은 Case2와 Case4가 약 2배 정도의 시간이 더 소모됨
- 결과적으로 **‘정확한 검색 결과 제공’**을 목적으로 한다면 **Case2 또는 Case4를 고려**해야 함
- 반대로 **‘서비스 제공 속도’**를 중점으로 둔다면 **Case1 또는 Case3를 고려**해야 함
- **최종적으로는 Case4를 적용**해 Vector DB를 구축 할때에 Person 클래스를 masked 처리한 뒤에 벡터로 변환을 시키고, 유사 이미지 검색 시에도 사용자가 촬영한 이미지속 Person 클래스를 masked 한 뒤 벡터화 하여 유사도 비교를 수행하는 방식을 선정
- 이는 선정 기준을 ‘서비스 제공 목적’에 두었기 때문.
- **‘이미지로 관광지 찾기’ 서비스의 핵심은 ‘정확도’**



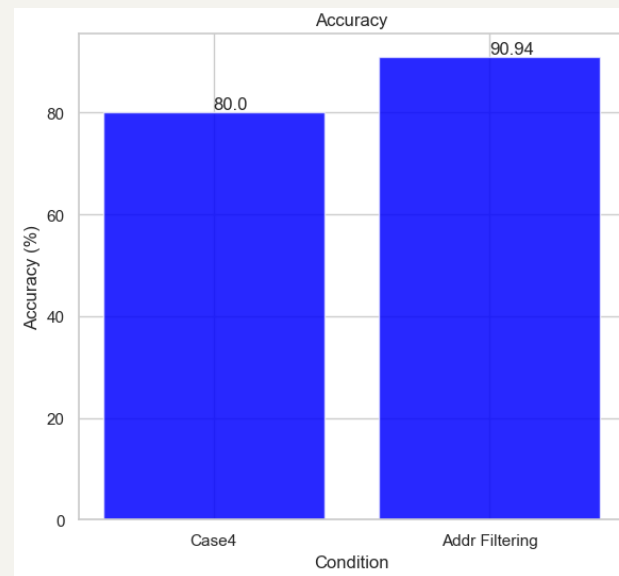
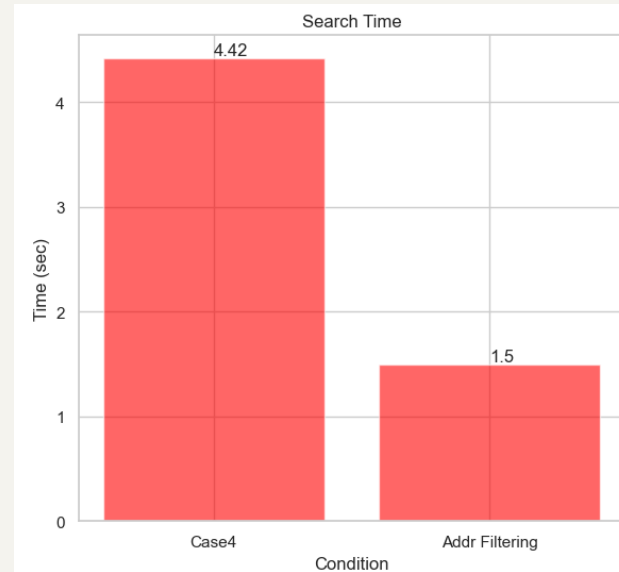
유사이미지 검색 성능 비교 실험 (2)

실험 목적

- 주소기반 필터링을 통한 유사 이미지 검색의 성능 향상 시도

실험 조건 설정

- 조건 1) 순수하게 이미지 검색
- 조건2) 주소기반으로 1차 필터링 후 이미지 검색
- 주소 검색시 이미지 데이터의 주소와, 벡터DB속 메타데이터 (주소)를 비교
- '읍면동' 단위까지 일치하는 데이터로 1차 필터링

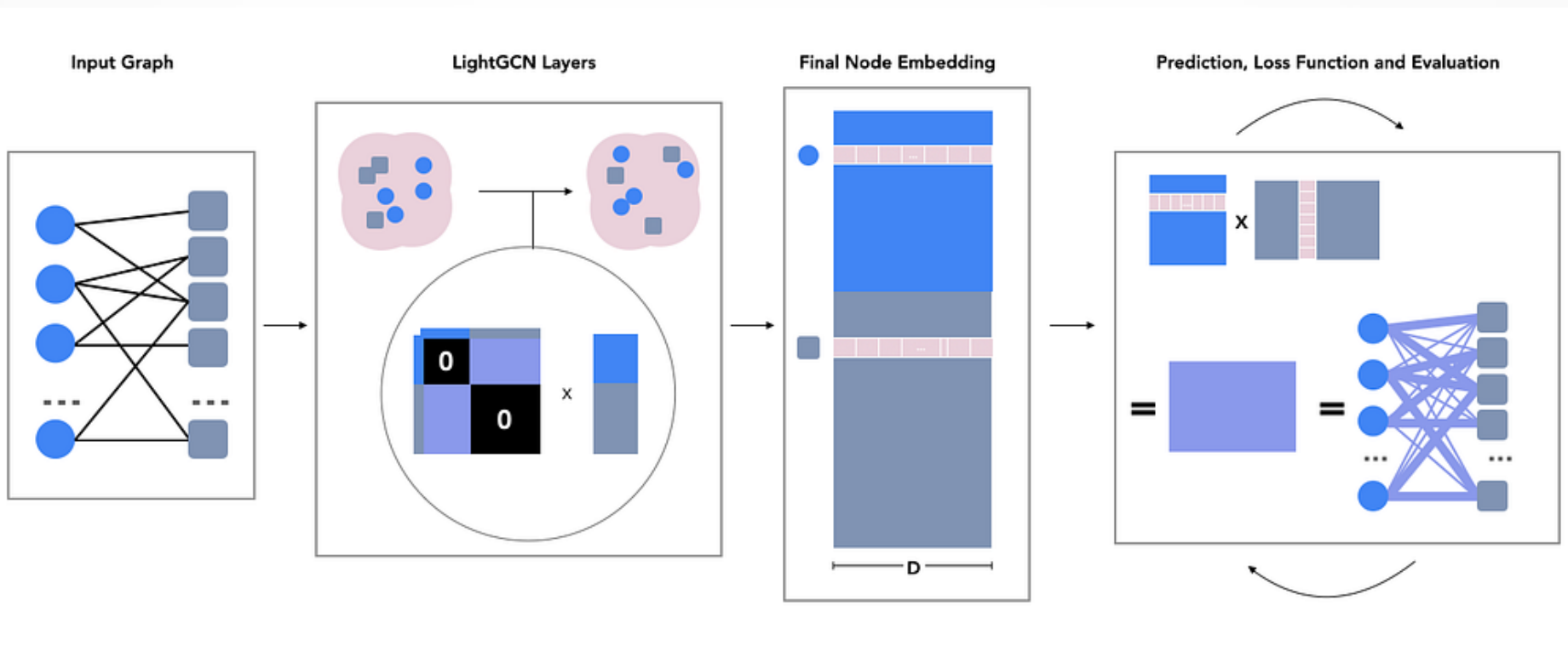


추천 모델 알고리즘

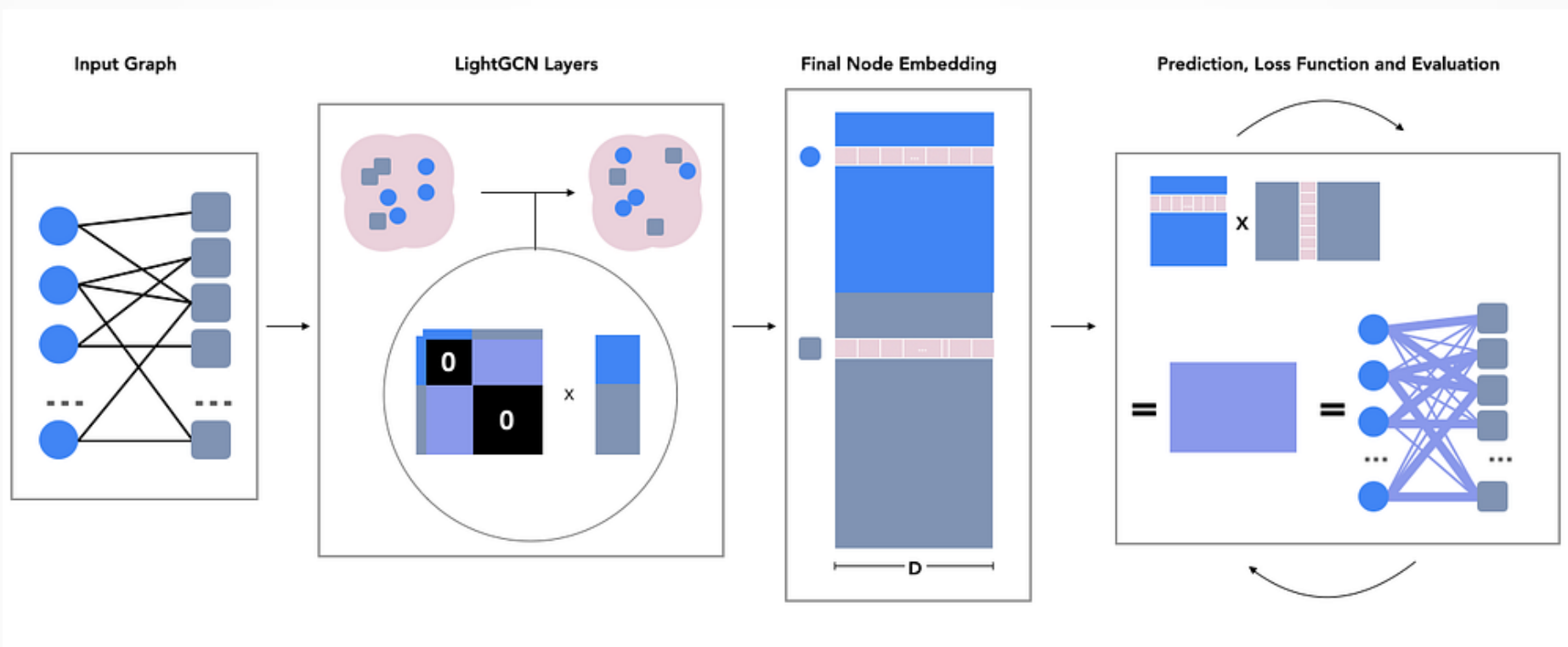
※ Cold Start : 새로운 User or Item에 대한 정보가 부족해 적절한 추천을 제공하지 못하는 현상

	CatBoost	LightGCN	LightGCN + POIMeta2Vec
모델 개요	<ul style="list-style-type: none"> 범주형 변수를 별도의 전처리 없이 자동 전처리 고차원 데이터에 우수한 성능을 보이는 Gradient Boosting 기반의 모델 	<ul style="list-style-type: none"> 그래프 구조를 활용하여 사용자와 여행지 간의 상호작용을 모델링하는 추천 알고리즘 GCN(Graph Convolution Network)를 경량화한 모델 	<ul style="list-style-type: none"> LightGCN 모델에 POIMeta2Vec을 함께 학습시키는 Customized LightGCN 모델
선정 이유	<ul style="list-style-type: none"> 가볍게 구축이 가능하며, 전처리된 데이터로 추천모델 구축이 가능한지 테스트 하기 위함 	<ul style="list-style-type: none"> 사용자와 여행지간 복잡한 상호작용을 효과적으로 캡처하고 활용하기 위해 Graph구조를 활용해 사용자와 여행지 간 고차원 연관성을 학습할 수 있어 사용자의 선호도를 효과적으로 반영 가능 경량화된 구조로 계산 효율성이 높아, 대규모 추천시스템에 적합 	<ul style="list-style-type: none"> LightGCN의 장점을 유지하면서, 여행지 메타데이터를 활용해 추천의 질과 다양성 향상 POIMeta2Vec을 활용함으로써 여행지 특성을 추천에 반영할 수 있음
특장점	<ul style="list-style-type: none"> User와 Item에 대한 다양한 범주형 데이터를 효과적으로 처리할 수 있음 Feature Importance를 제공해 추천에 대한 합리적 추론이 가능 	<ul style="list-style-type: none"> 사용자의 선호도를 반영해 개인화된 여행지 추천 제공 가능 	<ul style="list-style-type: none"> 사용자 선호도와 여행지 특성을 모두 고려한 개인화된 추천이 가능 Cold Start문제를 완화해 새로운 여행지, 상호작용 데이터가 부족한 여행지에 대해서도 추천이 가능
단점	<ul style="list-style-type: none"> Cold Start 이슈에 대응할 수 없음 	<ul style="list-style-type: none"> Cold Start 이슈에 대응할 수 없음 여행지 메타데이터를 직접적으로 활용하지 않아 추천 다양성이 제한될 수 있음 	<ul style="list-style-type: none"> 상대적으로 연산량이 많음
결과 평가	<ul style="list-style-type: none"> MAE 0.621 RMSE 0.7686 	<ul style="list-style-type: none"> Loss : 0.0203 Recall@10 : 0.2497 NDCG@10 : 0.1760 	<ul style="list-style-type: none"> Loss : 0.0152 (33.55% ▼) Recall@10 : 0.3713 (48.68% ▲) NDCG@10 : 0.2740 (55.68% ▲)

추천 모델 알고리즘(LightGCN)



추천 모델 알고리즘 (LightGCN + POIMeta2Vec)



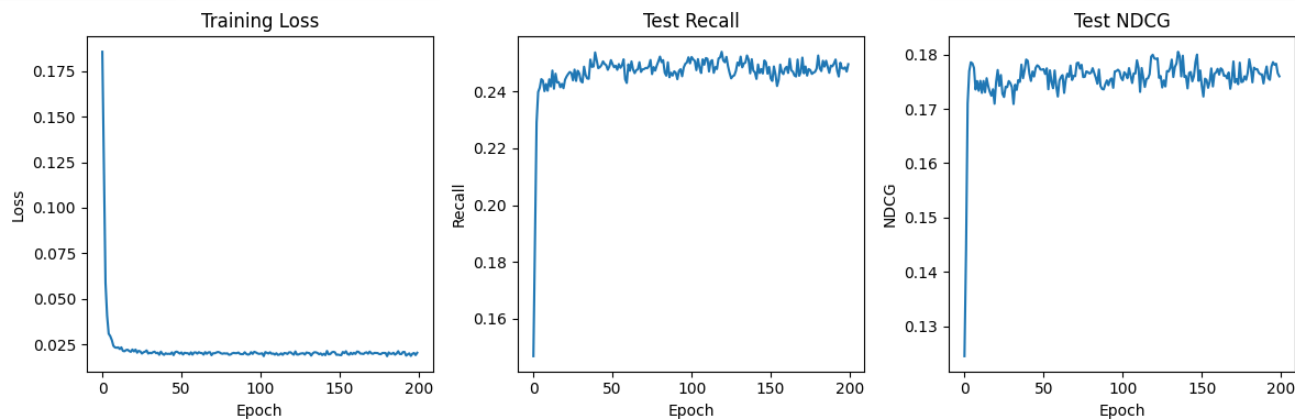
추천 모델 알고리즘 (LightGCN)

Loss
0.0203

Recall@10
0.2497

NDCG@10
0.1760

LightGCN



추천 모델 알고리즘

Loss

0.0152

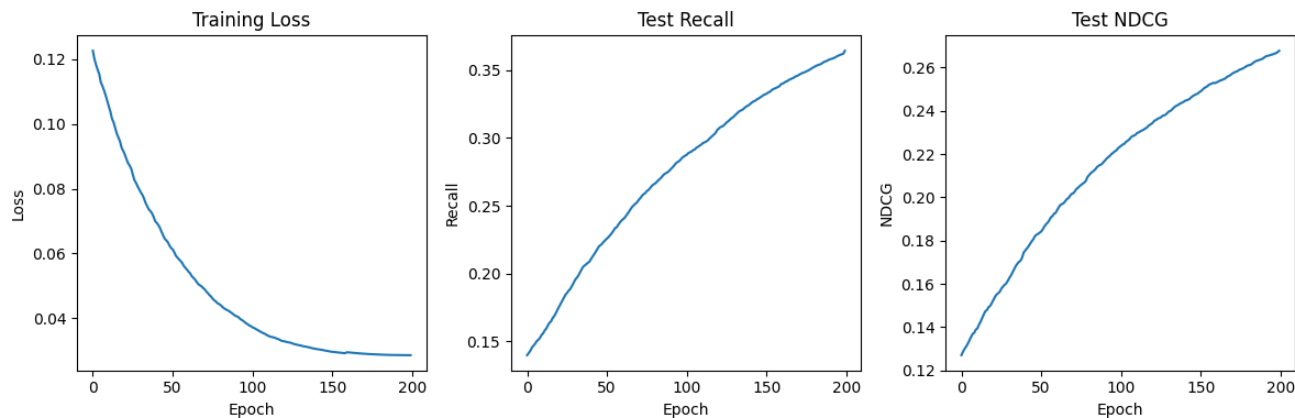
Recall@10

0.3713

NDCG@10

0.2740

LightGCN w. POIMeta2Vec



05

기억의 지도 App Demo

*"The best and most beautiful things in the
world cannot be seen or even touched.
They must be felt with the heart."*



06. Lessons & Learns

구분	내용
프로젝트 수행상 어려움 극복 사례	<ul style="list-style-type: none">수집한 2종의 관광 관련 데이터들의 관광지명, 주소등이 일치 하지 않는 이슈를 NLP의 유사도 기반 비교를 통해 해결데이터 수집과정에서 크롤링 속도 지연으로 '관광지 찾기' 기능 구현이 어려워졌고, 목표를 빠르게 변경해 미니멀한 목표 잡아서 구현(단, 제주지역 한정)추천모델의 Cold Start이슈 해결을 위해 LightGCN논문을 기반으로 Custom 모델을 구현
프로젝트에서 잘한 부분	<ul style="list-style-type: none">멀티모달을 적용한 앱서비스를 기획하고 구현한 것앱 서비스 구축을 위해 CLIP, LightGCN, CatBoost, POIMeta2Vec등 다양한 알고리즘의 논문을 탐독하고 모델을 구현해서 적용한 것
프로젝트에서 아쉬운 부분	<ul style="list-style-type: none">추천 모델의 Cold Start이슈 해결을 위해 모델을 구축했으나 신규 유입 유저에 대한 MLOps 시스템을 구현하지 못함별점 및 리뷰를 추가적으로 분석하고 모델링하는 부분을 구현해 MLOps 구축 필요AWS에서 Time-matter로 Firebase로 변경해서 백엔드 부분 1 cycle 구현 경험을 완료하지 못한 것