

# 성대생 맞춤형 식단 추천 AI

AIP Team Proj Presentation 2: GROUP 4

전상진 강민규 김동건



# Contents

## About

---

프로젝트 배경

프로젝트 목적

## Data

---

데이터셋 구성 - 크롤링

데이터셋 구성 - 설문

## Model

---

Baseline Model

Proposed Model

## Q&A

---





# About





# 프로젝트 배경

성균관대학교 학생들은..

- 교내·주변 상권의 식당과 메뉴가 너무 많아 매 끼니마다 메뉴 선택 피로를 겪고 있음.

또한..

- "예산을 초과하지 않을까?"
- "오늘은 뭐 먹지...?"
- "내가 좋아하는 메뉴가 어디 있지?"

→ 이런 고민으로 시간을 낭비

→ 개인 맞춤형 메뉴 추천 시스템이 부재한 상황

# 프로젝트 목적

학생 특성 반영한 최적의 메뉴 추천하는 AI 시스템 구축

- 학생의 위치·선호·예산·알레르기·평점 반영

→ 기대 효과

- 정보의 홍수 속에서 메뉴 선택 고민 시간 단축 / 최적의 메뉴 추천을 통해
- 데이터 기반 합리적 맞춤형 식사 선택 제공 / 식생활 만족도 향상
- 장기적 측면에서 건강한 식습관 유도
- AI 기술을 활용한 일상생활 문제 해결로 학교 생활의 질을 향상시키는 데 기여

## 성대생 맞춤형 식단 추천 AI



# Data





# 데이터셋 구성 - 크롤링



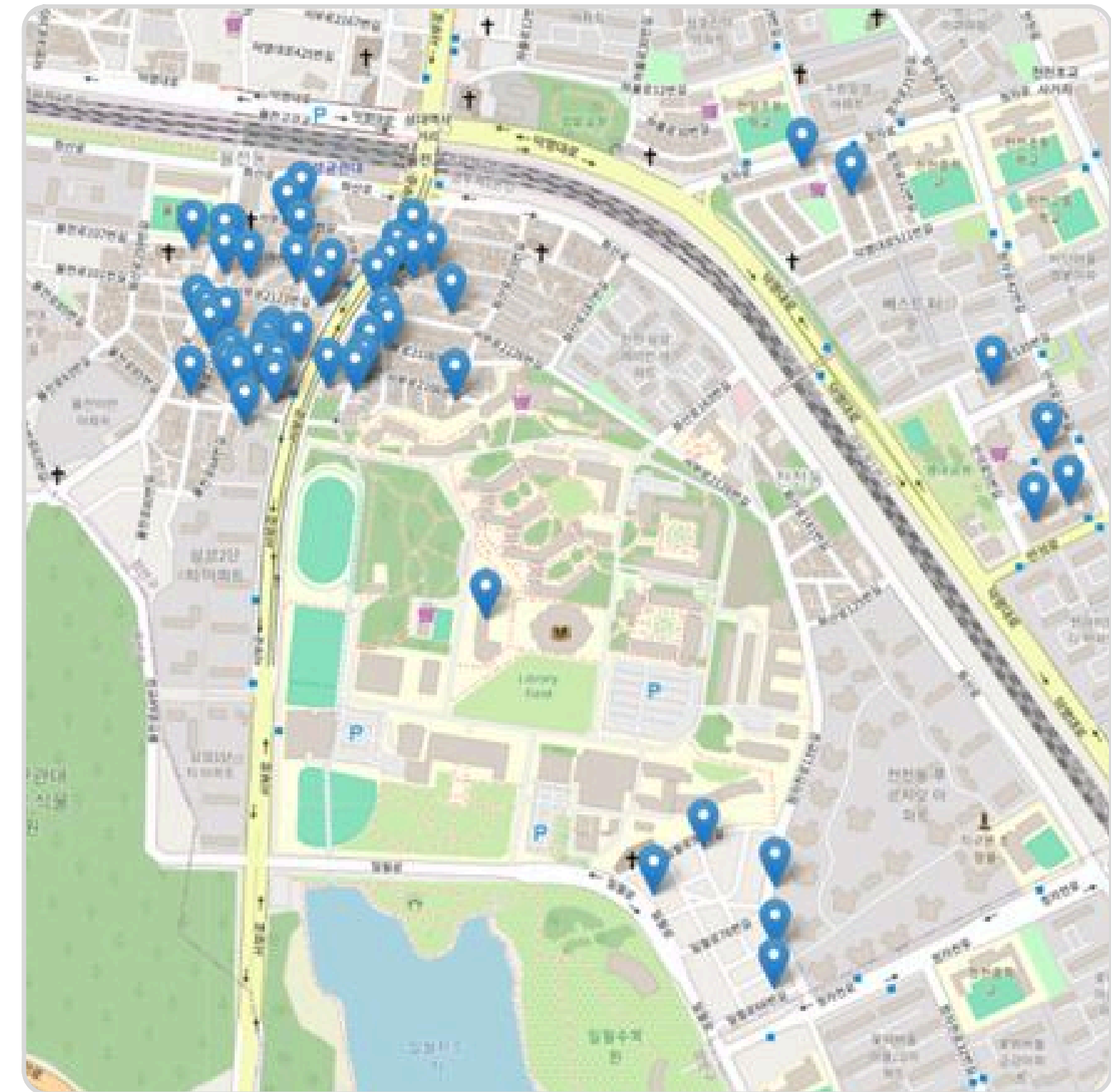
## 최종 선정 데이터

식당: 56개소

- 학교 근처 식당 한정 (화서역, 수원역 등 제거)
- 주요 칼럼 값 중 결측값 존재하는 식당 제외
- 카페와 술집 식당 제외

메뉴: 420개 항목

- 전체 수집한 메뉴 중,
  - '메뉴'가 아니거나(e.g. 어린이 특선 코스 등)
  - 중복되는 메뉴(e.g. 꺾바로우 소-꺾바로우 대, 타코야끼 8알-타코야끼 14알 등)모두 제거



# 데이터셋 구성 - 크롤링

## 1. 메뉴 데이터 (menu\_data.csv)

- 출처 | 네이버지도 + 카카오맵 크롤링
- 역할 | CB Score, Price Score 계산에 사용

1. menu\_data.csv

menu_id	menu	rest_id	price	features
1	삼겹	1	16000	육류 고기요리 한식 고기류 담백한 풍미 16000원대
num	str	num	num	str (단어는 공백으로 구분)

## 2. 식당 데이터 (rest\_data.csv)

- 출처 | 네이버지도 크롤링 + 주소  
→ 지오코딩(좌표 변환)
- 역할 | Distance Score, Avg. Rating Score

2. rest\_data.csv

rest_id	rest_name	Latitude	Longitude	rating
1	~고깃집	37.xx	126.xx	4.18
num	str	num	num	num



# 데이터셋 구성 - 설문

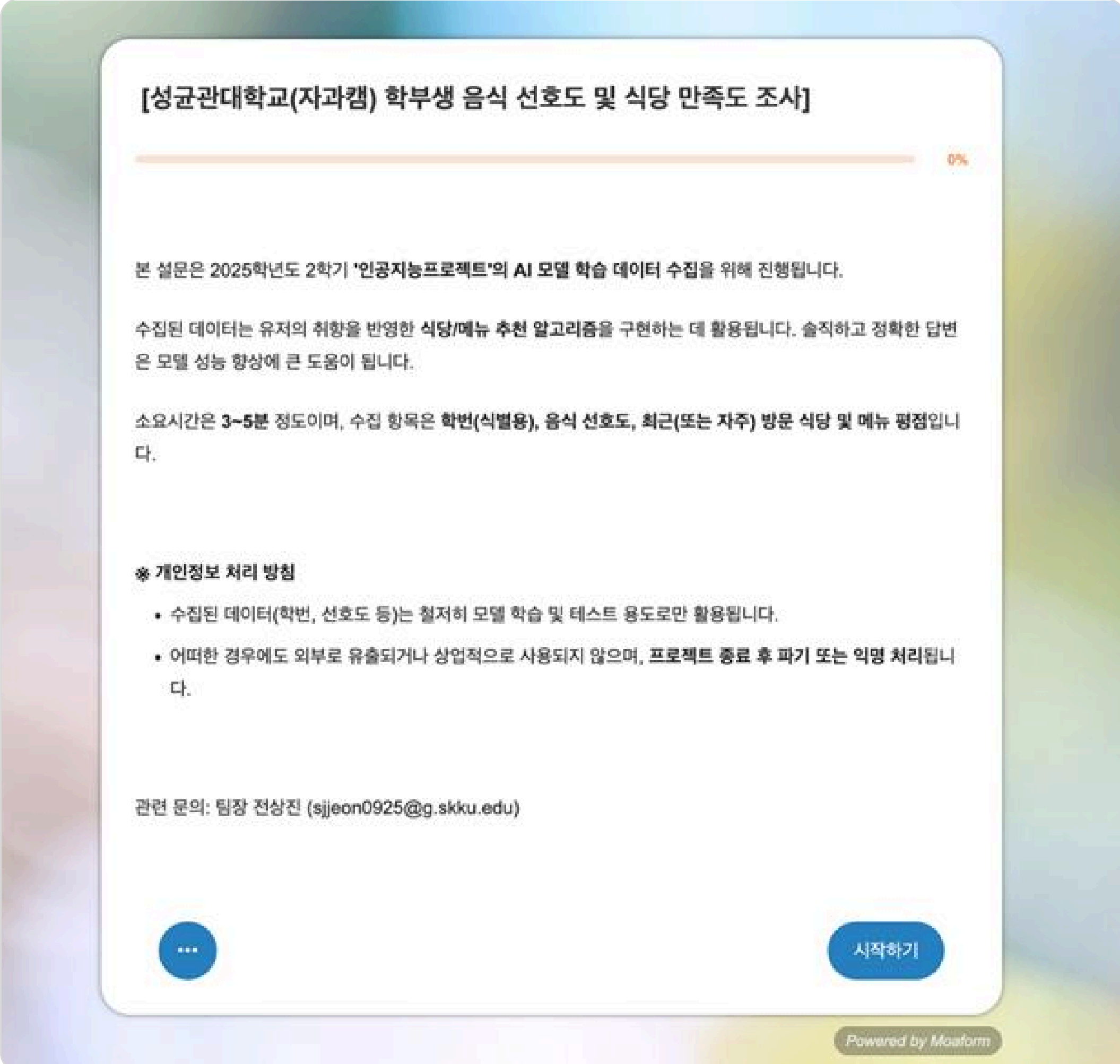
→ 사용자 설문 데이터를 반영해  
더 정확한 "맞춤형 식사 추천" 가능

: <https://moaform.com/q/xke6xR>

학생들의 개인 선호도 반영을 위해

- 사용자의 음식 카테고리
- 메뉴 스타일
- 선호하는 맛과 특징
- 선호 식당
- 선호 식당의 선호 메뉴 및 평점

조사



[성균관대학교(자과캠) 학부생 음식 선호도 및 식당 만족도 조사]

0%

본 설문은 2025학년도 2학기 '인공지능프로젝트'의 AI 모델 학습 데이터 수집을 위해 진행됩니다.

수집된 데이터는 유저의 취향을 반영한 식당/메뉴 추천 알고리즘을 구현하는 데 활용됩니다. 솔직하고 정확한 답변은 모델 성능 향상에 큰 도움이 됩니다.

소요시간은 3~5분 정도이며, 수집 항목은 학번(식별용), 음식 선호도, 최근(또는 자주) 방문 식당 및 메뉴 평점입니다.

※ 개인정보 처리 방침

- 수집된 데이터(학번, 선호도 등)는 철저히 모델 학습 및 테스트 용도로만 활용됩니다.
- 어떠한 경우에도 외부로 유출되거나 상업적으로 사용되지 않으며, 프로젝트 종료 후 파기 또는 익명 처리됩니다.

관련 문의: 팀장 전상진 (sjjeon0925@g.skku.edu)

시작하기

Powered by Moaform



# 데이터셋 구성 - 설문

## 3. 사용자 프로필 데이터 (user\_data.csv)

- 출처 | 실제 학생 설문 기반 구축
- 역할 | CB Score 계산,  
Hard Filtering (알레르기 포함 메뉴 제거)

3. user\_data.csv

user_id	preference	allergy
200	순대 순댓국 기타 부드러운 맛	새우
num	str (단어는 공백으로 구분)	str (선택 사항)

## 4. 사용자 평점 데이터 (rating\_data.csv)

- 출처 | 실제 학생 설문 기반 구축
- 역할 | CF Score 계산, GNN 모델 연산, MLP Target(Y) 학습
- 구성 | 29명의 유저가 인당 7개의 식당, 각 식당 별 3개의 메뉴에 대한 평점

4. rating\_data.csv

user_id	menu_id	rating	location	append
201	84	2.95	후문	김밥 기타 바삭한 식감 5000원대
num	num	num	str (평가 당시 위치)	str (평가 당시 추가 쿼리)



# Model





# Baseline Model

Baseline – 콘텐츠 기반(CB) + 협업 필터링(CF) + MLP 구조(Blender MLP)

## 1. 베이스라인 코드 기반 설명

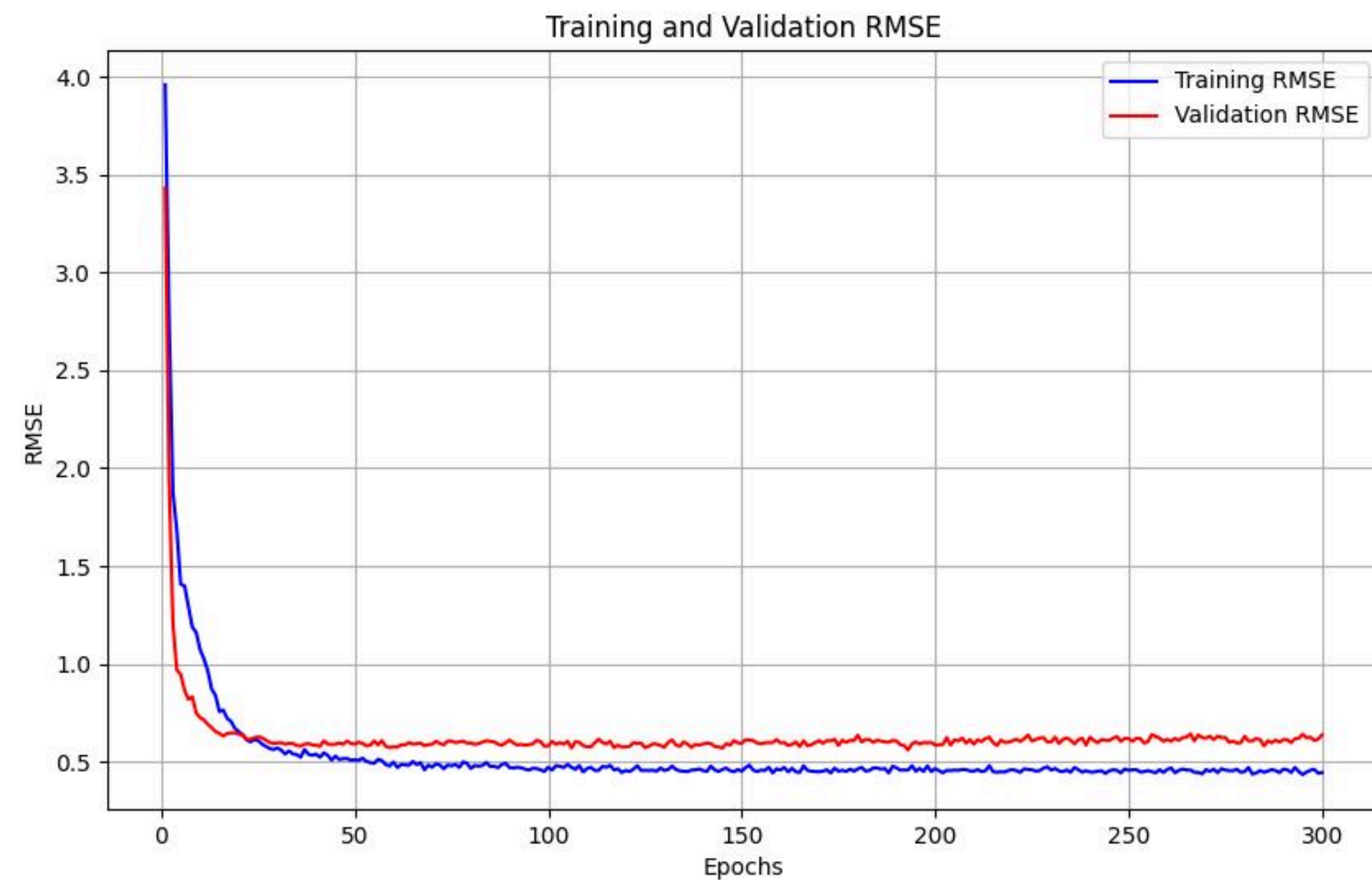
- Filtering - MLP pipelining 구조
  - 4개의 csv 파일 입력
- Filtering 부분에서 CB, CF 모델 연산
  - Distance, Price, Rating 값은 거의 그대로 사용
- 위의 데이터를 MLP 입력 X로 사용
  - CF Score / CB Score / 거리 Score / 가격 Score / 식당 평점 Score
- 타겟 Y : rating\_data.csv의 유저의 실제 rating 값

## 2. 학습 방법

- 지도학습 방식으로, 4개의 csv 파일 입력을 5종류의 데이터로 변환하여 MLP가 실제 데이터 Y값을 예측하도록 진행
- train, test 분리는 8:2 비율로 무작위 분리
  - skit-learn 자체 메소드 사용
  - from **sklearn.model\_selection** import **train\_test\_split**



# Baseline Model



최종 평가 결과 (Top-10)  
- 총 평가 유저 수: 29명  
- Average Recall : 11.77%

- 설문조사를 통해 얻은 실제 데이터로 학습 한 결과 MLP의 loss는 정상적으로 감소
- 실제 유저의 선택을 재현하는 Recall 값은 많이 낮은 것을 확인할 수 있음
  - 11.77%



# Proposed Model

## 1. Graph Model (graph\_model.py)

- 유저 Node와 메뉴 Node를 생성
- Node들 간에 연결을 학습을 통해 추론하는 GNN 구조
- GNN 단독으로 연산하여 Recall 값을 구하면 48.92%가 나옴

```
User 20 : 정답 19개 중 9개 적중 -> Recall: 47.4%
User 20 : 정답 19개 중 9개 적중 -> Recall: 47.4%
User 20 : 정답 19개 중 8개 적중 -> Recall: 42.1%
User 20 : 정답 21개 중 10개 적중 -> Recall: 47.6%
User 20 : 정답 21개 중 10개 적중 -> Recall: 47.6%
-----
📊 최종 평가 결과 (Top-10)
- 총 평가 유저 수: 29명
- Average Recall : 48.92%
-----
```

Parameter	Value
EMBEDDING_STD	0.01
GNN_EPOCHS	500
GNN_LR	0.005
GNN_WD(Weight Decay)	1e-6

```
Epoch 450/500, Loss: 0.0754
Epoch 460/500, Loss: 0.0656
Epoch 470/500, Loss: 0.0489
Epoch 480/500, Loss: 0.0708
Epoch 490/500, Loss: 0.0679
Epoch 500/500, Loss: 0.0649
GNN 모델 학습 완료.
```



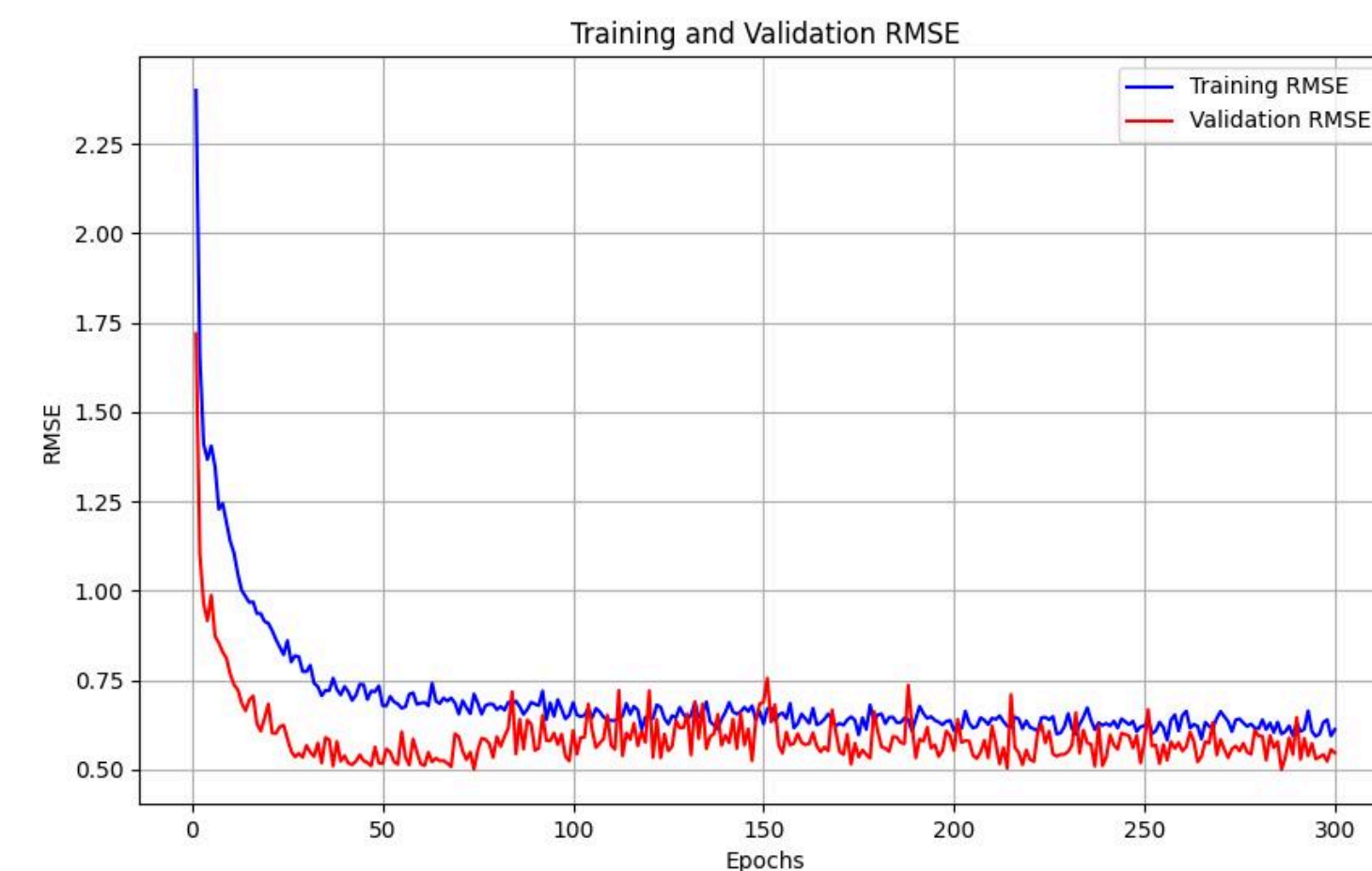
# Proposed Model

## 2. MLP 모델과 결합

- Baseline의 Filtering-MLP 모델과 결합
  - 5가지 입력 데이터에 GNN 데이터 추가
- 가치가 높은 GNN 값의 비중 증가
  - Min-Max 정규화
  - Negative Sampling 추가
- 최종 Recall = 31.62%

최종 평가 결과 (Top-10)  
- 총 평가 유저 수: 29명  
- Average Recall : 31.62%

Parameter	Value
<b>MLP_EPOCHS</b>	300
<b>MLP_BATCH_SIZE</b>	4
<b>MLP_LR</b>	0.001
<b>Dropout</b>	0.5





# Proposed Model

## 3. 단순 정확도(Recall) vs 실질적 만족도

- GNN 단독 모델의 Recall이 더 높게 측정됨.
  - GNN은 사용자의 과거 방문 패턴을 학습하여, 사용자가 '이미 먹어본(평가한) 메뉴'를 맞추는 데 탁월함.
- Recall의 한계:
  - 프로젝트 목표는 메뉴 추천 모델.
  - 높은 Recall수치는 '익숙함'을 의미할 뿐, 새로운 발견이나 현재 상황을 보장하지 않는다.
- 추천 모델의 본질은 내가 평가한 메뉴를 맞추는 것이 아니라, **가보지 않았어도 좋아할 메뉴를 제안하는 것**

- 다양한 정보를 **MLP를 통해 결합함**으로써 이러한 문제 해결
- **잠재 취향 + 상황 정보 결합:**
  - **협업 필터링:** 대중적 선호도와 유저 간의 유사성 반영.
  - **콘텐츠 기반:** 사용자가 평가하지 않은 메뉴라도, 선호하는 '맛/식감 키워드'와 유사하다면 추천.
  - **상황 정보:** 거리, 가격, 가게 평점 등 의사결정에 필수적인 추가 요인을 가중치로 학습.
- 단순 과거 데이터 재현(Recall)에서 **미래 만족도 예측(Prediction)**으로의 전환
  - 만족도 예측 성능 측정은 최종 모델에 대한 유저 만족도 조사가 필요하기에 생략함



# Baseline vs Proposed Model

	Baseline	Proposed
모델 아키텍처	MLP (5-Input) (SVD + CB + Meta)	MLP (6-Input) (Baseline + GNN)
데이터 처리	Raw Data 학습	GNN Score 정규화
학습 전략	Positive(방문한 곳) 위주 학습	Negative Sampling 적용 (안가본 곳의 평점을 0으로 확실히 구분)
성능 (Recall)	11.77%	31.62%

## 관계 정보의 추가 (GNN Integration):

- Baseline(5차원)은 **나의 기록**만 보거나, SVD 연산으로 관계를 단순히만 파악하는 반면,
- Proposed(6차원)는 **나와 취향이 비슷한 그룹의 기록(GNN Score)**을 6번째 판단 근거로 추가하여 정보의 밀도를 높임.

## 학습 완성도 향상 (Negative Sampling & Normalization):

- GNN 점수를 그대로 쓰지 않고 0~1로 정규화하여 MLP가 더 잘 이해하도록 가공함.
- Negative Sampling을 통해, MLP가 **GNN 점수가 낮으면(=관계가 없으면) 추천하지 말아야 한다**는 판단력을 갖게 함.





# Baseline vs Proposed Model

	Baseline	Proposed
모델 아키텍처	MLP (5-Input) (SVD + CB + Meta)	MLP (6-Input) (Baseline + GNN)
데이터 처리	Raw Data 학습	GNN Score 정규화
학습 전략	Positive(방문한 곳) 위주 학습	Negative Sampling 적용 (안가본 곳의 평점을 0으로 확실히 구분)
성능 (Recall)	11.77%	31.62%

## Cold Start 문제:

- 방문 기록 데이터가 아예 없는 신규 유저의 경우에는 GNN이 작동하지 않아 CB에만 의존해야 함.

## 계산 비용:

- GNN 학습 시간이 Baseline보다 올리기에 실시간 학습에는 한계가 있음.





# Conclusion Qualitative analysis

레몬크림통새우	상초마라 성균관대수원
죽배기소불고기	최가네
제육덮밥	성대찌개고을
상하이 파스타	몽키키친
고르곤졸라 M	롤링파스타 수원성균관
치킨부리또	도스마스 성균관대점
꿔바로우	수해복마라탕 성대점
시그니처 꽃삼겹	육가든
로제떡볶이	신전떡볶이 수원성대점

baseline

돈코초라멘	면식당 수원성균관대점
냉라멘	미가라멘 성대본점
샤브샤브국수	면식당 수원성균관대점
K-불고기피자	롤링파스타 수원성균관대점
베이컨김치필라프	아늑
동백	한솔도시락 수원성대점
진달래	한솔도시락 수원성대점
까르보나라	롤링파스타 수원성균관대점
신전떡볶이	신전떡볶이 수원성대점

proposed

한식 일식 중식 육류 치킨/닭고기  
면 얼큰한/매콤한 담백한 푸짐한

위와 같은 preference를 가진 유저에 대한 메뉴 추천 결과

## Why Hybrid?

- 해당 유저는 Proposed 모델 기준으로 Recall 값이 20.0%로 낮은 유저
- Proposed(Hybrid) 모델에서 유저가 선호하는 카테고리과 부합하는 메뉴 추천이 더 많이 일어난 것을 확인할 수 있음.
- 단순히 Recall 수치뿐만 아니라 구체적인 취향 등을 잘 반영하였음을 알 수 있다.



# Q&A

