

2024.12.19.

Personalized Healthcare with AI

Health protector

김다은 김우신 김태현 유승찬 임과림 전지은

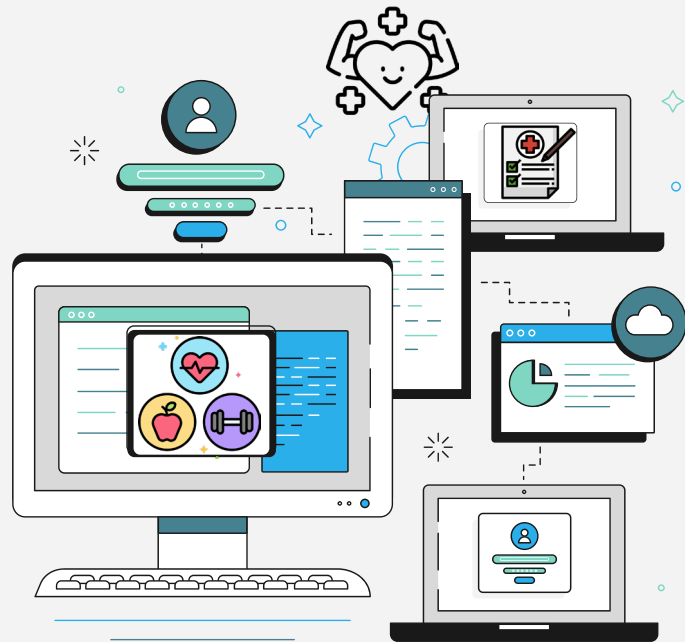


Table of contents

01

Background
& Problem
Statement

02

Data
Overview

03

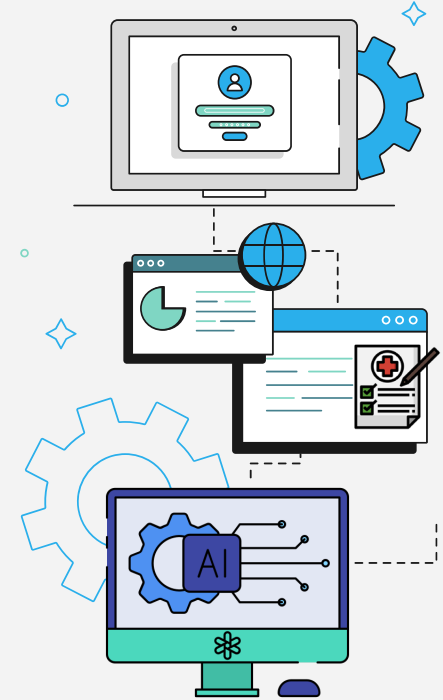
Prediction
Model

04

Recommendation
Model

05

Practical
Applications



01

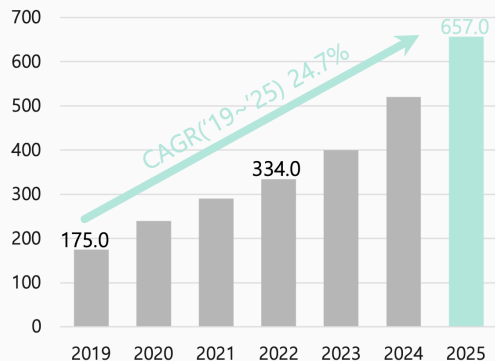
Background & Problem Statement



글로벌 디지털 헬스케어 시장 규모

[출처] ETRI

(십억달러)



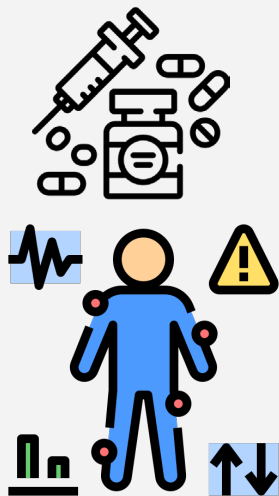
고령화 가속, 코로나 19 확산 등 **건강에**
대한 관심과 투자비용이 꾸준히 성장

000

· 국내 건강기능식품 시장 규모 ·

[출처] Quick OEM

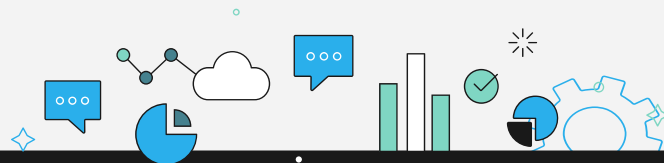




소비자들은 신뢰할 수
있는 정보 부족과
개인 맞춤형 추천
시스템의 부재로
올바른 소비결정에
어려움을 겪고 있음

Database

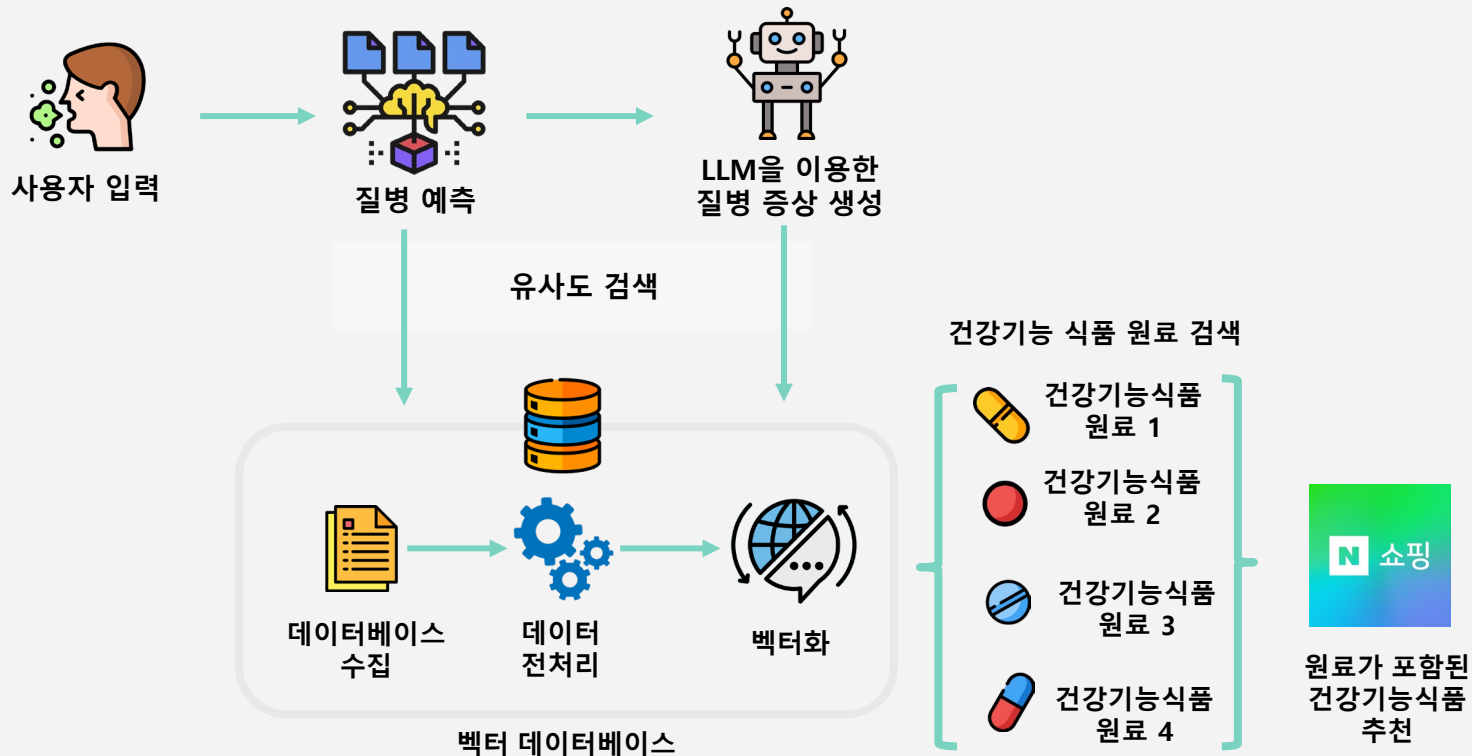
- 식품안전의약처 건강기능식품 품목분류정보 DB



- 건강기능식품
원료별 기능성,
성분명 등에
대한 속성 정보
- 461개 원료
성분에 대한
섭취 효능 및
주의사항

```
"row": [
  {
    "PRIMARY_FNCITY": "배변활동 원활에 도움을 줄 수 있음",
    "SKLL_IX_IRDNT_RAWMTRL": "아라비아검(아카시아검)식이첨유",
    "LAST_UPDT_DTM": "20200820",
    "CRET_DTM": "20080922",
    "PRDCT_NM": "아라비아검(아카시아검)",
    "IFTKN_ATNT_MATR_CN": "반드시 충분한 물과 함께 섭취할 것 (역상제외)",
    "INTK_MEMO": "",
    "DAY_INTK_LOWLIMIT": "19.99",
    "DAY_INTK_HIGHLIMIT": "20.01",
    "INTK_UNIT": "g"
  },
  {
    "PRIMARY_FNCITY": "(1) 어두운 곳에서 시각 적응을 위해 필요\n(2) 피부와 점막을 형성하고 기능을 유지하는데 필요\n(3) 상피세포의 성장과 발달에 필요",
    "SKLL_IX_IRDNT_RAWMTRL": "",
    "LAST_UPDT_DTM": "20210924",
    "CRET_DTM": "20080626",
    "PRDCT_NM": "베타카로틴",
    "IFTKN_ATNT_MATR_CN": "(1) 흡연자는 섭취 시 전문가와 상담할 것\n(2) 이상사례 발생 시 섭취를 중단하고 전문가와 상담할 것",
    "INTK_MEMO": "식물원료를 사용하여 제조가공한 경우는 1.26mg 이상(상한섭취량 없음)",
    "DAY_INTK_LOWLIMIT": "0.42",
    "DAY_INTK_HIGHLIMIT": "7",
    "INTK_UNIT": "mg"
  }
],
```

Prediction & Recommendation Model





02

Data Overview

User interviews

Q. 자신의 건강 상태를 얼마나 정확하게 알고
있나요?

* 공복 혈당, 혈압 등 포함

정확하게 알고 있다 10%

어느 정도 알고 있다 20%

잘 모르겠다 70%

- 총원: 10명
- 남성 6명, 여성 4명
- 평균 나이: 26.4세

NHIS & KNHANES

	NHIS	KNHANES
데이터 범위	전 국민의 건강검진, 진료 내역, 소득 데이터 포함	192개 지역에서 표본 추출 (약 1만 명 조사)
데이터 유형	연속형 + 범주형 (검진 수치, 의료 이용 내역)	연속형 + 범주형 (검진 수치, 생활습관, 영양 조사)
결측치	매우 많음	
사용 가능 변수	건강검진 수치, 소득, 진료 내역	건강검진 수치, 설문조사 결과 (음주, 수면, 식이행태 등)



03

Prediction Model

Prediction Model



모델 학습 : CatBoost vs LightGBM

- 두 모델 결과를 비교하여 예측 성능과 해석 가능성을 높임
- 근소하게 높은 성능을 보인 **LightGBM** 최종 선택

	CatBoost	LightGBM
데이터 처리	<ul style="list-style-type: none">• 연속형/범주형 데이터 모두 처리 가능• 데이터 전처리 부담 최소화	
결측치 처리	결측치에 대한 대응이 용이함	
학습 속도	상대적으로 느림	상대적으로 빠름
해석 가능성	Feature 중요도 해석으로 모델 결과 신뢰성 확보	
성능	분석 결과 상대적으로 낮은 성능	분석 결과 상대적으로 높은 성능



Feature Selection : RFECV (Recursive Feature Elimination with Cross-validation)

- RFECV 라이브러리를 사용하여 Feature를 제거해가며 (RFE) 성능에 대한 교차검증 진행(CV)
- 건강검진 정보 유무를 고려하여 NHIS 데이터와 KNHANES 설문조사 데이터로 나누어 모델을 학습하고 성능을 비교 검증

Model Structure

Strong Model

규칙기반 라벨링에서
유병자 판단 기준으로
라벨링 후 학습

1

위험군

0

Weak Model

규칙기반 모델링에서
의심군 판단 기준으로
라벨링 후 학습

1

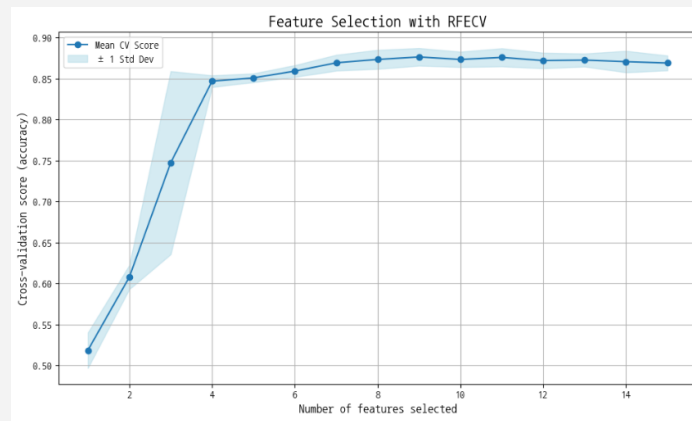
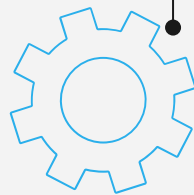
0

안전군

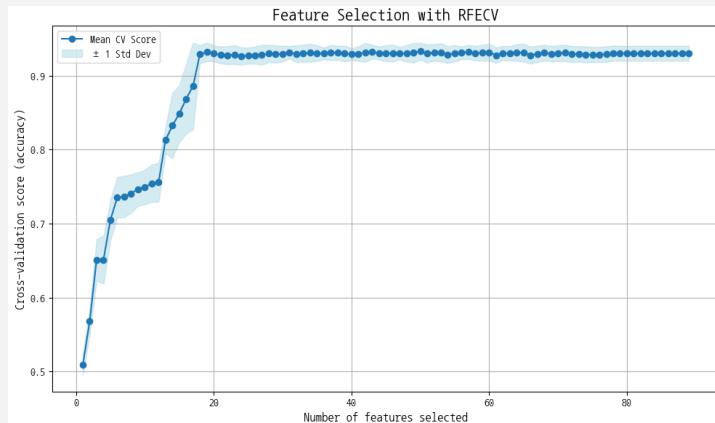
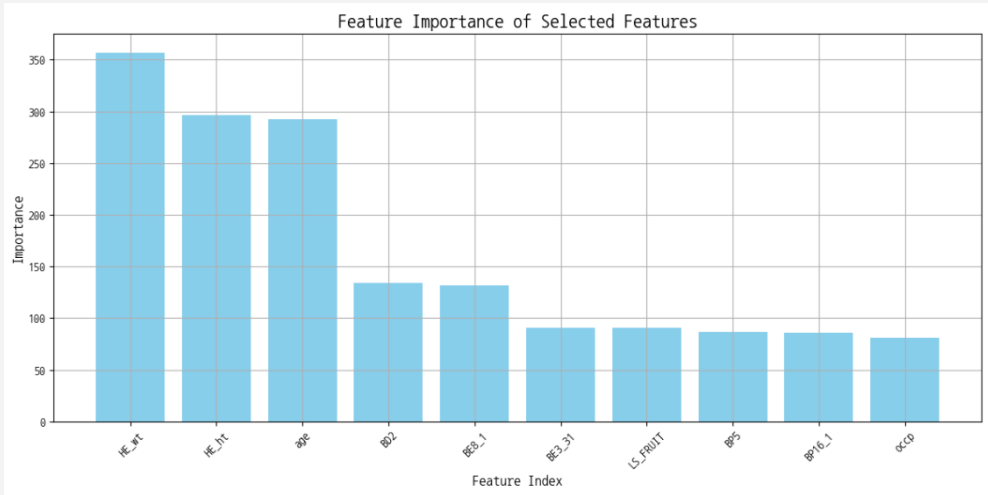
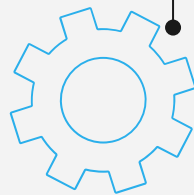
의심군

- 안전/유병간 클래스 불균형으로 인한 분류 정확도 저하 문제를 해결하기 위해 2개의 모델을 사용

교차 검증 결과(고혈압, 건강검진 결과, Strong 모델 기준)



교차 검증 결과(고혈압, 설문조사 결과, Strong 모델 기준)



전체 질병에 대한 모델 테스트 결과 및 실제 설문 결과

	label	strength	model	accuracy	0_precision	0_recall	0_f1	1_precision	1_recall	1_f1	features
0	HE_obese	strong	lgbm	0.99	0.99	0.99	0.99	0.99	0.99	0.99	['HE_ht', 'HE_wt', 'BP16_1']
1	HE_obese	weak	lgbm	0.99	0.99	1.00	0.99	1.00	0.99	0.99	['HE_ht', 'HE_wt']
2	HE_HCHOL	strong	lgbm	0.68	0.73	0.60	0.66	0.64	0.77	0.70	['age', 'HE_ht', 'HE_wt', 'BE8_1']
3	HE_HCHOL	weak	lgbm	0.68	0.73	0.60	0.66	0.64	0.77	0.70	['age', 'HE_ht', 'HE_wt', 'BE8_1']
4	HE_HTG	strong	lgbm	0.63	0.66	0.61	0.64	0.61	0.66	0.63	['age', 'HE_ht', 'HE_wt']
5	HE_HTG	weak	lgbm	0.63	0.66	0.61	0.64	0.61	0.66	0.63	['age', 'HE_ht', 'HE_wt']
6	HE_anem	strong	lgbm	0.67	0.68	0.66	0.67	0.66	0.69	0.68	['age', 'HE_ht', 'HE_wt', 'BE8_1']
7	HE_anem	weak	lgbm	0.67	0.68	0.66	0.67	0.66	0.69	0.68	['age', 'HE_ht', 'HE_wt', 'BE8_1']
8	HE_HP	strong	lgbm	0.85	0.92	0.78	0.84	0.79	0.92	0.85	['HE_wt', 'HE_ht', 'age', 'BD2', 'BE8_1', 'BE3...']
9	HE_HP	weak	lgbm	0.72	0.72	0.71	0.72	0.72	0.73	0.73	['age', 'HE_ht', 'HE_wt']
10	HE_DM_HbA1c	strong	lgbm	0.70	0.76	0.61	0.68	0.65	0.80	0.72	['age', 'HE_ht', 'HE_wt', 'BE8_1']
11	HE_DM_HbA1c	weak	lgbm	0.72	0.75	0.65	0.70	0.69	0.79	0.74	['age', 'HE_ht', 'HE_wt', 'BE8_1']

비만 - 위험도: 안전, 확률: 0.0, 평균 대비 발병 위험: 0.0배
 고콜레스테롤 혈증 - 위험도: 안전, 확률: 0.09, 평균 대비 발병 위험: 0.86배
 고지혈증 - 위험도: 안전, 확률: 0.04, 평균 대비 발병 위험: 0.13배
 빈혈 - 위험도: 안전, 확률: 0.13, 평균 대비 발병 위험: 0.83배
 고혈압 - 위험도: 안전, 확률: 0.15, 평균 대비 발병 위험: 0.77배
 당뇨 - 위험도: 안전, 확률: 0.05, 평균 대비 발병 위험: 0.5배



User Input

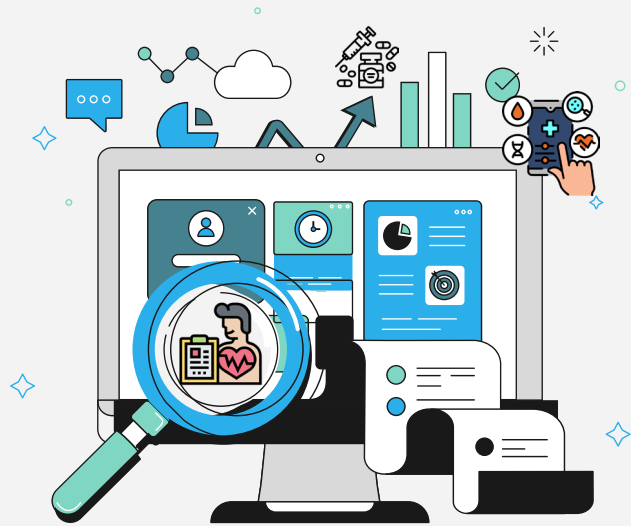


○ 건강검진정보를 알고 있을 때

- 성별
- 나이
- 체중
- 신장
- 허리둘레
- LDL-콜레스테롤 수치
- 혈중 크레아티닌 수치
- HDL-콜레스테롤 수치
- AST/ALT 수치
- 1년간 음주 빈도

○ 건강검진정보를 모를 때

- 성별
- 나이
- 체중
- 신장
- 허리둘레
- 하루 앉아서 보내는 시간
- 주중(또는 일하는 날) 하루 평균 수면 시간
- 음주 시작 연령
- 과일류 섭취 빈도
- 2주 이상 연속 우울감 여부
- 직종
- 1주일간 걷기 일수



04

Recommendation Model

Technology Infographics

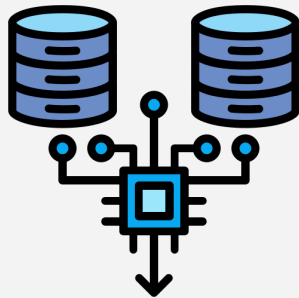


JSON 데이터
로드 및 전처리

건강기능식품 품목
분류 정보를 JSON
데이터에서 로드

Open AI 임베딩 생성

텍스트 데이터를 분할한 후
Open AI의 **text embedding**
-3-small 모델을
사용해 **벡터** 임베딩 생성



FAISS 벡터 DB 구축

생성된 임베딩을
바탕으로 FAISS를 활용해
벡터 데이터베이스 구축

LLM을 통한 검색 Query 생성

예측된 질병 정보를
LLM에 전달해 증상
정보 생성

검색 Query : 증상
정보 + 사용자 입력
쿼리 + 질병 정보



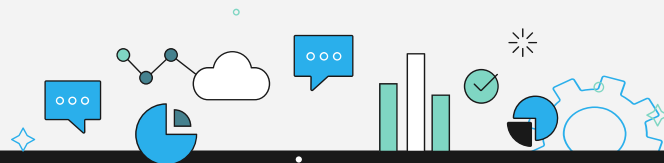
건강기능식품 추천

FAISS 벡터 DB를 활용해
코사인 유사도로 원료
효능과 이름 검색

네이버 쇼핑 API를 통해
최종 건강기능식품 추천
제공

Database

- 식품안전의약처 건강기능식품 품목분류정보 DB

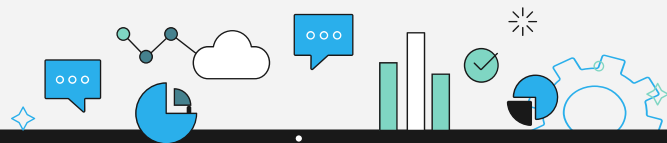


- 건강기능식품
원료별 기능성,
성분명 등에
대한 속성 정보
- 461개 원료
성분에 대한
섭취 효능 및
주의사항

```
"row": [
  {
    "PRIMARY_FCNCTY": "배변활동 원활에 도움을 줄 수 있음",
    "SKLL_IX_IRDNT_RAWMTRL": "아라비아검(아카시아검)식이첨유",
    "LAST_UPDT_DTM": "20200820",
    "CRET_DTM": "20080922",
    "PRDCT_NM": "아라비아검(아카시아검)",
    "IFTKN_ATNT_MATR_CN": "반드시 충분한 물과 함께 섭취할 것 (역상제외)",
    "INTK_MEMO": "",
    "DAY_INTK_LOWLIMIT": "19.99",
    "DAY_INTK_HIGHLIMIT": "20.01",
    "INTK_UNIT": "g"
  },
  {
    "PRIMARY_FCNCTY": "(1) 어두운 곳에서 시각 적응을 위해 필요\n(2) 피부와 점막을 형성하고 기능을 유지하는데 필요\n(3) 상피세포의 성장과 발달에 필요",
    "SKLL_IX_IRDNT_RAWMTRL": "",
    "LAST_UPDT_DTM": "20210924",
    "CRET_DTM": "20080626",
    "PRDCT_NM": "베타카로틴",
    "IFTKN_ATNT_MATR_CN": "(1) 흡연자는 섭취 시 전문가와 상담할 것\n(2) 이상사례 발생 시 섭취를 중단하고 전문가와 상담할 것",
    "INTK_MEMO": "식물원료를 사용하여 제조가공한 경우는 1.26mg 이상(상한섭취량 없음)",
    "DAY_INTK_LOWLIMIT": "0.42",
    "DAY_INTK_HIGHLIMIT": "7",
    "INTK_UNIT": "mg"
  }
],
```

Database (cont'd)

• 네이버 검색 쇼핑 API



- API를 통해
사용자 질병을
완화할 수 있는
성분이 들어있는
건강기능식품
추천
- 상품평, 링크,
가격, 제품 이름
등을 json
형식으로 받음

[상위 3개 검색 결과]

결과 1

상품 정보: {'title': '랩온랩 비디컬 맥주효모 비오틴 10000 0.5g x 60정', 'link': 'https://search.shopping.naver.com/catalog/51261846618', 'image': 'https://shopping-phinf.pstatic.net/main_5126184/51261846618.20241106165748.jpg', 'lprice': '19500', 'hprice': '', 'mallName': '네이버', 'productId': '51261846618', 'productType': '1', 'brand': '랩온랩', 'maker': '7LAB', 'category1': '식품', 'category2': '건강식품', 'category3': '비타민제', 'category4': '비오틴'}

결과 2

상품 정보: {'title': '닥터아들 B발라민 비오틴 징크비오틴 120정 2개월', 'link': 'https://smartstore.naver.com/main/products/4609303560', 'image': 'https://shopping-phinf.pstatic.net/main_8215382/82153823855.4.jpg', 'lprice': '38000', 'hprice': '', 'mallName': '닥터아들', 'productId': '82153823855', 'productType': '2', 'brand': '닥터아들', 'maker': '', 'category1': '식품', 'category2': '건강식품', 'category3': '비타민제', 'category4': '비오틴'}

결과 3

상품 정보: {'title': '포뉴 유기농 비오틴 600mg x 60캡슐', 'link': 'https://search.shopping.naver.com/catalog/20812559737', 'image': 'https://shopping-phinf.pstatic.net/main_2081255/20812559737.202406141114114.jpg', 'lprice': '33000', 'hprice': '', 'mallName': '네이버', 'productId': '20812559737', 'productType': '1', 'brand': '포뉴', 'maker': '한국씨엔에스팜', 'category1': '식품', 'category2': '건강식품', 'category3': '비타민제', 'category4': '비오틴'}

05

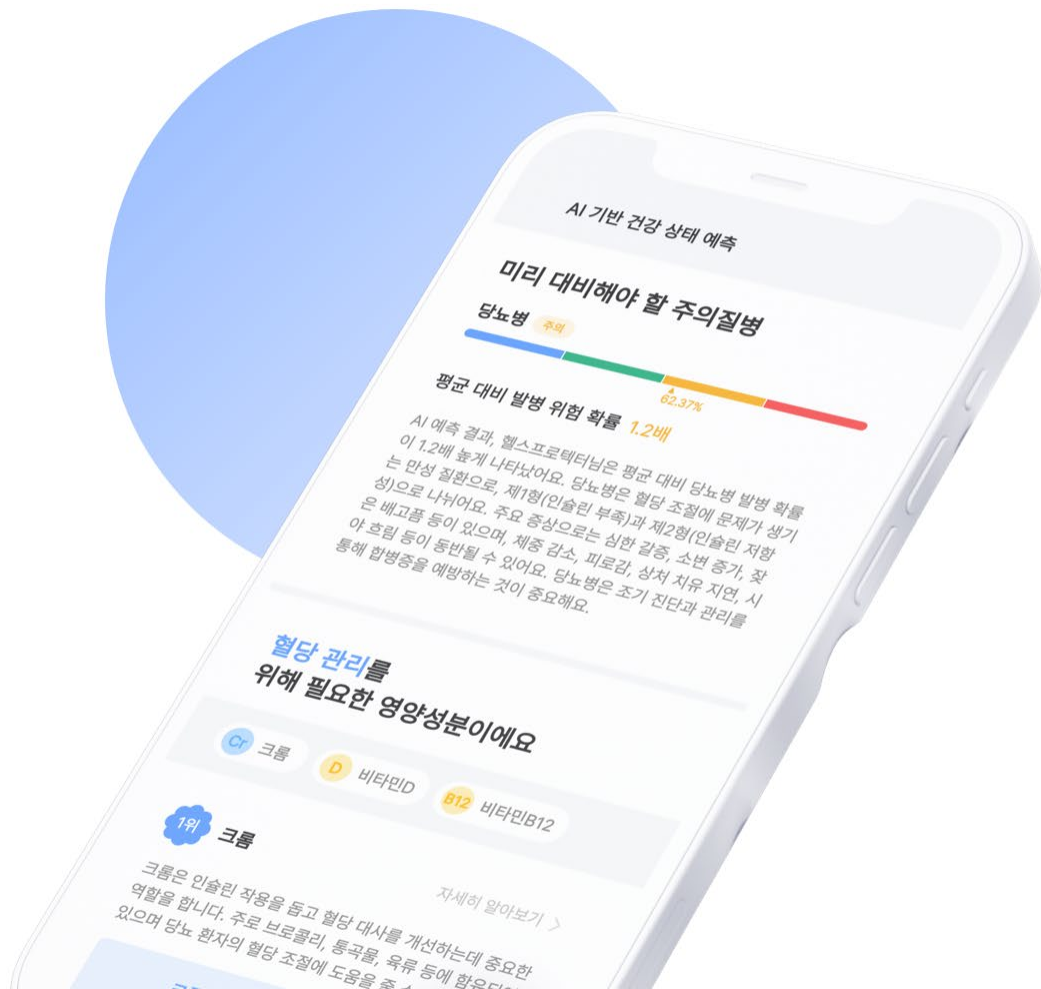
Practical Applications





AI를 활용한 건강 예측 및 개인 맞춤형 영양제 추천 서비스

정확한 건강검진 데이터 또는 사용자의 설문조사
결과를 기반으로 건강 상태를 예측하여
개인 맞춤형 영양제를 추천해주는 서비스입니다.



5:13 5G

< 생활 습관으로 건강 진단하기

20%

Q. 신장과 체중을 입력해주세요.

신장	체중
175 cm	64 kg

다음

1	2 ABC	3 DEF
4 GHI	5 JKL	6 MNO
7 PQRS	8 TUV	9 WXYZ
.	0	✕

5:13

< 생활 습관으로 건강 진단하기

Q. 하루 평균 수면 시간은 얼마나 되나요?

- ☐ 11시간 이상
- ☐ 9 ~ 10시간
- ☐ 7 ~ 8시간
- ☒ 5 ~ 6시간
- ☐ 4시간 이하

< 생활 습관으로 건강 진단하기

40%

Q. 하루 중 앉아서 보내는 시간은 몇 시간 인가요?

- ☐ 10시간 이상
- ☒ 7 ~ 9시간
- ☐ 4 ~ 6시간
- ☐ 2 ~ 3시간
- ☐ 1시간 이하



평균적인 걷기 활동 빈도



평균 과일/채소류 섭취 빈도

하루 중 앉아서 보내는 시간은?

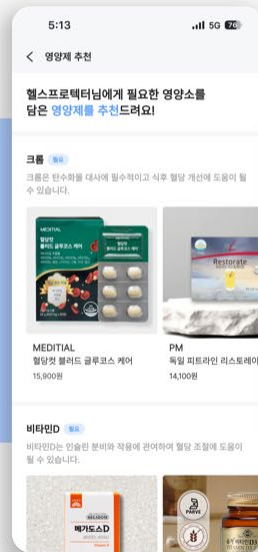
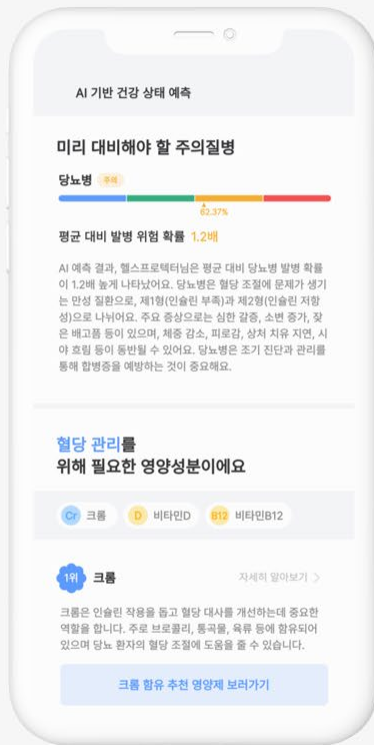
하루 평균 수면 시간은?

정확한 건강 검진 결과를 모르시나요?

건강 검진 결과를 모르더라도, 설문조사를 기반으로 건강 상태를 예측할 수 있어요

건강 데이터 기반 건강 상태 예측 및 맞춤형 영양제 추천

건강 검진 데이터 또는 설문조사 결과를
기반으로 대비해야 할 주의 질병을 예측한 뒤,
맞춤형 영양제를 추천해드려요



상세 성분 분석

크롬

크롬은 탄수화물 대사에 필수적이고 식후 혈당 개선에 도움이 될 수 있습니다.

비타민D

비타민D는 인슐린 분비와 작용에 관여하여 혈당 조절에 도움이 될 수 있습니다.

나이아신

나이아신은 체내 에너지 생성에 도움을 줄 수 있습니다.





Thanks!

Do you have any questions?

